University of Groningen

Trajectories of Health Status in Older People

Feenstra, Marlies

DOI:
[10.33612/diss.160814885](https://doi.org/10.33612/diss.160814885)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Feenstra, M. (2021). *Trajectories of Health Status in Older People*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.160814885>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Trajectories of Health Status in Older People

Colofon

Trajectories of Health Status in Older People.
Dissertation, University of Groningen, The Netherlands

Author: Marlies Feenstra
Cover design: Marlies Feenstra (painting), Maaïke van der Post (layout)
Lay out: Maaïke van der Post
Photography: Unsplash Inc, Montreal (licensed)
Printing: Ridderprint BV, Ridderkerk

Copyright © M. Feenstra, 2021

All rights reserved. No part of this dissertation may be reproduced or transmitted in any form by any means without explicit written permission of the author.

The printing of this dissertation was financially supported by the University Medical Center Groningen, University of Groningen, and the Research Institute SHARE, Graduate School of Medical Sciences, University of Groningen, University Medical Center Groningen.

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 31 maart 2021 om 11.00 uur

door

Marlies Feenstra

geboren op 21 maart 1985
te Roden

Promotores

Prof. dr. S.E.J.A. de Rooij

Prof. dr. B.C. van Munster

Dr. N. Smidt

Beoordelingscommissie

Prof. dr. T. Hortobágyi

Prof. dr. J.A.M. van der Palen

Prof. dr. M.H. Emmelot-Vonk

Paranimfen

Kiki Buijs

Fiona van El

TABLE OF CONTENTS

	PAGE
Chapter 1	General Introduction 11
Chapter 2	Discrepancies in Self-Perceived and Measured Change in Activities of Daily Living after Hospital Admission among Older Adults 29
Chapter 3	Reproducibility and Responsiveness of the Frailty Index and Frailty Phenotype in Older Hospitalized Patients 59
Chapter 4	Translation and validation of the Dutch Pittsburgh Fatigability Scale for older adults 93
Chapter 5	Trajectories of self-rated health in an older general population and their determinants: The Lifelines Cohort Study 133
Chapter 6	Determinants of trajectories of fatigability and mobility among older medical patients during and after hospitalization, an explorative study 167
Chapter 7	General Discussion 195
Chapter 8	Summary 223 Nederlandse samenvatting About the author Dankwoord (Acknowledgements) Research Institute SHARE

Aanpassingsvermogen in één zin:

*Al ben ik ook op leeftijd nog een wankelmoedig figuur
en trek ik het me heel erg aan dat ik zo afhankelijk ben
van van alles en nog wat, toch voel ik me ook best vaak
een nuttig lid van de samenleving, bijvoorbeeld als ik
bedenk dat mijn fiets zonder mij zijn evenwicht verliest.*

(Hans Dorrestijn, 2018¹)

¹ Uit balans. In H. Dorrestijn, *Het Rimpelperspectief* (p.37). Amsterdam: Nijgh & Van Ditmar. Opgenomen in dit proefschrift met toestemming van de auteur.

VOORWOORD

Dag ouderen.

Ik zie u staan. Gezamenlijk, ogenschijnlijk onwrikbaar, gevormd en geleefd, levend, sterk. Mag ik u bekijken? Mijn perspectief veranderen?

Ik kijk naar u op. Een leven lang ervaring, kennis, wijsheid. Zoveel om me nietig bij te voelen.

Wanneer ik op u neerkijk zie ik patronen die ik probeer te vangen. Trachtend u zodoende te reduceren tot slechts enkele groepen. U lijkt op elkaar maar toch ook niet. Elk uw eigen verhaal.

Dan, plots, een vonk, een lichtpunt, uw kloppende hart. Als een kwetsbaar, maar uitbundig bloeiende vrucht. Dat is waar u voor leeft.

Verder inzoomend zie ik uw huid: gebruind, gerimpeld, gebutst. De krassen, littekens, zij spreken voor zich. Hoe dikker hoe weerbaarder?

Ik kijk omlaag en zie dat u steeds meer heeft moeten afstoten. Als uitgevallen dennennaalden.

Bent u zelf gevallen? Nee, ik zie het al. U ondersteunt uw buur. Zij aan zij. Als lotgenoten.

Dan opent u mijn ogen. U verandert mijn perspectief en laat me door de bomen het prachtige bos weer zien. U opent mijn ogen door ze, zonder weg te kijken, te sluiten en me te laten luisteren.

CHAPTER 1

General introduction

Enabling older people to live as independently and healthy as long as possible is one of the challenges societies face as a result of global population aging.^{1,2} In order to design interventions to maintain or regain optimal health status in older people, even after serious health events, adequate tools and instruments are required to assess the course and deviations of health status over time. The overall aim of this dissertation is to evaluate health status trajectories and their determinants in older people. The first chapters focus on psychometric issues of existing instruments used in practice for assessment of the health status of older people. The second part includes actual trajectories of health status and their determinants among older people living in the community and in hospitalized older patients.

Population aging

One of the main successes of health care is the continuing increase in life expectancy.^{3,4} Due to the longer life expectancy, the population of people aged 65 years and older will rapidly expand in the coming decades. More specifically, in the Netherlands, the proportion of citizens aged 65 years and older relative to the total population is expected to rise from 20% now to 25% by 2040.⁵ Within this ageing population the proportion of people aged 80 years and older will increase even more rapidly as well from one in four now to one in three by 2040.⁵ Enabling older people to live as independently and healthy as long as possible serves a twofold purpose: on the one hand, this meets the needs of the majority of older people themselves², on the other hand, it ensures that the pressure on facility-based long-term care remains manageable (in terms of workforce) and affordable (in terms of costs).⁶ However, the increase in lifespan will also lead to a rise in the incidence of age-related chronic conditions, which are known to be associated with greater health-service use such as hospitalization.⁷⁻⁹ Hospitalization, in turn, often results in a loss of functioning¹⁰, especially among older people with multiple chronic conditions (multimorbidity) and pre-existing limitations in activities of daily living (ADL), thus posing a serious threat to independence. It is therefore important to prevent hospitalizations or delay hospitalization¹¹, but also, when hospital admission is inevitable, to enable older people to return to their pre-hospital level of functioning as quickly as possible during and after hospitalisation.¹²⁻¹⁴

Consequences of hospitalization in older people

In 2018, around 25% of the people aged 65 years and older had a clinical hospital admission in the Netherlands.¹⁵ Hospitalization is major risk factor for mortality among older people. The in-hospital and 30-day post-hospital discharge mortality rates of older acutely admitted people are 19%¹⁶ and 5% respectively.^{16,17} In addition, 30% of the acutely hospitalized older patients experience hospital-associated loss of function.¹⁸ Recovery of hospital-associated loss of function occurs in 30% to 70% of the cases, predominantly within the first three months after discharge.^{19–21} There is a long history of scientific interest in the causal relationships between loss of function and recovery from hospitalization in older people.^{22–26} Older age, problems in performing ADL, cognitive impairment, depression, impaired social functioning, geriatric conditions and length of hospital stay are known predictors of loss of function after hospital admission.^{10,27} Remarkably, fatigue has not been considered as a predictor of persistent functional decline and lack of recovery, though fatigue is one of the most common symptoms in older hospitalized people, with a prevalence rate of around 70%.^{28,29} After hospital discharge, many older people still experience fatigue, often leading to disruption of activities of daily living.²⁹ This high prevalence of fatigue may be followed by a high prevalence of fatigability as well, which is defined as perception of fatigue while performing activities of a fixed intensity and duration.³⁰ Relating perception of fatigue to activities of a fixed intensity and duration allows better comparisons between people with comparable activity levels, and accounts for so called self-pacing bias.^{30,31} Fatigability might be considered as a more sensitive measure of someone's susceptibility to fatigue, making this construct more suitable as a determinant and a potentially reversible research outcome.^{30,32,33} Over the past few years, fatigability has emerged as an important marker of functional decline in mobility-intact community-dwelling older people^{34,35}, yet its relationship with loss of function after hospitalization is unclear.

Defining health

There is much debate going on about the definition of health. The main point of criticism of the classic definition of health as 'a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity' as defined by the World Health Organisation³⁶, is the central position of illness by using the term 'complete well-being'.^{37,38}

Huber considers illness as a challenge and incorporates the dynamic nature of health by defining health as 'the ability to adapt and self-manage in the face of social, physical, and emotional challenges'.³⁸ In this dissertation, we chose for the latter definition because it is closer to what we all are dealing with in aging: As we get older, it is inevitable that we will be faced with changing circumstances of all kinds (physical, mental, social or environmental) that invite us to adapt to our environment and circumstances, which may in turn affect our health status.^{38–40} Consequently, when we want to understand the course of health, we should consider multiple assessments of health status over time to catch its fluctuations.^{39,40}

Trajectories of health status

Mrs. White

Mrs. White (1936) has been widowed for 15 years and lives independently. She has a past medical history of Type 2 Diabetes Mellitus, hypertension, osteoporosis and osteoarthritis. Together with her daughter, she has been participating in the Lifelines Cohort Study since 2012. Every eighteen months, she has to complete a questionnaire and once every five years she is invited for a physical examination at one of the Lifelines research sites. To the question 'In general, how would you rate your health?' she answers invariably 'good', because within the constraints of having multiple chronic conditions, she feels good. In 2019 she was admitted to our hospital after a fall due to sudden loss of strength in her left leg caused by a Transient Ischemic Attack (TIA). During hospitalization, I approached Mrs. White for the Seniorlines 2.0 study. Despite the pain caused by the fall, she decided to participate. Her replies to the questionnaires showed that she was able to function quite independently before hospitalization. She did use the supermarket delivery service, had help with housekeeping once a week, and used the medication dispensing service of her local pharmacy. Every week, she took the bike to visit her former neighbor. She reported to be fatigued after performing physical and social activities sometimes; "But,"

she said, “that was all part of getting older.” The physical tests required a lot of effort during both the first assessment and the second assessment around discharge. However, she was very motivated and insisted to walk through the hospital corridors while she endured the pain.

Observational studies are usually longitudinal, which means they assess an outcome using two or more time points. Study designs in which at least three time-points are measured can be used to answer questions like ‘how does people’s health status develop over time?’ and ‘what characteristics predict that people are more likely to recover from functional decline after hospitalization?’.⁴¹ Of the methods available to analyze such developmental trajectories over time, group-based trajectory modeling (GBTM) is characterized by letting go the assumption of predefined subpopulations within a certain population.^{42,43} GBTM should rather be considered as a statistical method to identify individuals within a population that follow the same trajectory concerning shape and level. Once distinct trajectories within a population are modeled, questions like ‘Are characteristics of individuals that follow a certain trajectory different from individuals belonging to another trajectory?’ and ‘How is the trajectory of outcome x related to trajectories of outcome y?’ could be subsequently investigated.⁴³ GBTM will not only provide us a more accurate view of reality, it gives us also the opportunity to look more specifically at characteristics of individuals that follow a certain trajectory which is important, for example, when targeting interventions.

Measuring health status

Mrs. White, three months after hospital discharge

Three months after discharge from the hospital, I visited Mrs. White’s home for the first follow-up measurement of the Seniorlines 2.0 study. We drank tea while I administered the questionnaires. She told me she was afraid of falling and needed help with bathing and dressing. Walking was more painful than

cycling, but due to the fear of falling, she did not use her bike. Twice every week she went to a senior fitness class and practiced on the exercise bike with her physiotherapist. The fatigability remained unchanged. The physical tests took her less effort than during the hospitalization. “If someone walks with me, I even dare to walk without a walker”, she said proudly.

Measuring health status in older people is not straightforward. Traditional concepts to assess health status such as ‘disease-free years’ and survival are seldom a priority to older people with multimorbidity.⁴⁴ In contrast, independence and quality of life are more important health status outcomes among older people.^{45,46} But, there is a paradox here: even with increasing physical limitations and declining objective health status outcomes, the self-perceived health status of older people remains almost stable.^{47–51} The subjective nature of health status makes this construct difficult to measure as illustrated by the clinical case description. However, measuring health status and comparing health status differences between populations is important because it can support improving the quality of care. Roughly, three types of instruments to measure health status are distinguished: disease-specific, generic, and domain-specific instruments.⁵² Disease specific instruments focus on disease-related symptoms and problems. Often these instruments are extensively tested for validity and reliability within the intended patient groups, but they do not allow comparisons between conditions.^{52,53} Generic instruments are non-disease specific by definition and cover a wide spectrum of self-perceived physical function, disability and emotional functioning that affects health status.⁵³ These instruments are especially useful to compare general health status of multiple populations or conditions, but disadvantages are the lack of focus on specific domains of health status and they often lack responsiveness.⁵³ Domain specific instrument cover only one domain of health status which allows to focus on one specific aspect of health status, without losing the ability to compare this domain of health status between different populations or conditions. The instruments used in the studies of this dissertation are predominantly domain-specific covering domains that are known to be important to older people’s health status, such as ADL,

mobility, fatigability, and frailty. Exception is the study of trajectories of self-rated health (Chapter 5) which is considered a generic instrument.

The importance of sound measurement properties

Within the multitude of available instruments, measurement properties may guide the decision which instrument to select to answer clinical or research questions.⁵⁴ In general, three types of measurement properties can be distinguished⁵⁵: 1. Validity, which is defined as the degree to which the instrument measures the construct intended to measure⁵⁶; 2. Reliability, defined as the extent to which the instrument is free of measurement error⁵⁶, and; 3. Responsiveness, defined as the extent to which the instrument is able to measure change in the construct to be measured.⁵⁶ The purpose with which an instrument has been developed has consequences for the measurement properties.^{57,58} For instance, for instruments that have been developed for risk-assessment, such as instruments used for predicting adverse health outcomes, information about the validity is most important. Instruments developed for discriminative purpose should be valid and reliable⁵⁷, because the ability to distinguish scores of, for example, patients with high risk frailty from low risk frailty becomes important. When instruments are developed for the evaluation of interventions, also responsiveness becomes an important measurement property because it distinguishes real (clinical relevant) change in the construct from normal variability of the change scores.⁵³ Consequently, measurements properties of instruments have implications for their application as well.⁵⁷

Rationale and data sources

Mrs. White, six months after hospital discharge

Three months later, six months after her hospital discharge, I met Mrs. White at the agreed time in front of the entrance to her apartment on a three-wheeled bicycle. Together we went inside and I noticed she was using a cane. I administered the questionnaires and physical tests for the fourth and last time. "She was doing well," she said. She still had pain in her hip from the fall, but her new bicycle enabled her to visit her former neighbor again. In addition, she made friends at the senior fitness class

whom she also meets outside the classes. She was able to dress and shower independently again, but her fatigability remained unchanged. When I asked her, "How would you rate your health in general?" She answered, "Good." Without hesitation, after which she explained to be happy that she has been able to adapt and that she can manage independently again.

As previously reported, this dissertation is about trajectories of health status in older people, measured in different relevant populations. From the introduction it can be concluded that to be able to measure trajectories of health status over time, we first need to know whether our instruments are responsive and suitable for longitudinal data analyses. We also introduced fatigability as a promising marker of functional decline in older people. Perceived fatigability is expected to be a useful outcome in populations not capable to perform physical activities, such as hospitalized older patients, but it has not yet been validated in this population. Part 1 (chapters 2, 3 and 4) of this dissertation therefore addresses multiple psychometric properties of several instruments used to measure health status in older people. Part 2 (chapters 5 and 6) is devoted to the core question of this dissertation:

What are determinants of health status trajectories in community-dwelling and hospitalized older people?

Two cohorts are used to answer the research questions: the Seniorlines Cohort Study, including hospitalized patient aged 70 years and older; and the Lifelines Cohort study, of which we used a subsample including all participants aged 65 years or older at baseline. A short description of both cohorts is provided below.

The Seniorlines Cohort Study

Four chapters of this dissertation used data of the Seniorlines Cohort Study (Chapter 2, 3, 4, and 6). The Seniorlines Cohort Study was started in 2014 in the University Medical Centre Groningen (UMCG) to investigate the added value of hospital treatment for hospitalized patients aged 70 years and older. Data collection continued to 2020 using qualitative and quantitative research methods. The Seniorlines Cohort Study investigated the goals, wishes and expectations of older patients admitted to the

hospital and it also examines the extent to which the goals appointed by the patients were achieved after discharge from the hospital. The course of physical, mental and social well-being from hospital admission up to three and twelve months after hospital discharge was also examined by using extensive questionnaires. More than 2500 patients gave consent to participate in the Seniorlines cohort study, including patients who were admitted to cardiology, oncology, vascular and hepatobiliary, trauma, and internal medicine wards.

Seniorlines 2.0

Chapter 6 of this dissertation is based on an add-on study of the Seniorlines Cohort Study called Seniorlines 2.0. Seniorlines 2.0 was designed to investigate the relation between fatigability and physical functioning in older medical patients. Next to the questionnaires which largely overlapped the questionnaires used in the Seniorlines cohort study, physical tests including mobility, spirometry, bio impedance, hand grip strength were administered. Also activity monitoring and the collection of blood samples from routine blood sampling stay were part of the Seniorlines 2.0 study protocol. Baseline and discharge assessments took place within the hospital. The other follow-up assessments were done by home visits after three and six months after hospital discharge.

The Lifelines Cohort Study

Chapter 5 used data of the Lifelines Cohort Study. The Lifelines Cohort Study is a large cohort study with the objective to facilitate scientific research and the development of policy focused on healthy ageing. The data collection of the Lifelines Cohort Study included 1. Self-assessment questionnaires, including a broad range of domains covering lifestyle, health, personality, work, and living environment; 2. Physical assessment including blood pressure, BMI, electrocardiography, spirometry, and cognitive function, and; 3. Biological samples, including blood, faeces, urine, and hair. A detailed description of the complete Lifelines cohort profile is described elsewhere.⁵⁹ Recruitment of participants took place from 2006 to 2013 with an intended follow-up period of at least 30 years. Baseline assessment data included questionnaire administration, physical assessment and biological sampling. Every eighteen months questionnaire data was collected, and once every five years also physical assessment and biological samples were collected. For chapter 5, the so-

called elderly cohort of the Lifelines Cohort Study was used, including more than 12.000 participants aged 65 years or older at baseline, obtained from the Lifelines data and biobank.

Structure of the dissertation

The first three chapters are devoted to examining relevant psychometric properties of various instruments representing aspects of health status in older people. In Chapter 2 we addressed validity by comparing older people's own perceptions of change in ADL and measured change in ADL assessed with the modified Katz ADL index. In Chapter 3 we compared the reproducibility, a form of reliability, and responsiveness of two widely used frailty instruments: the Frailty Index and the Frailty Phenotype. Chapter 4 concerns the results of the translation and validation process of the Dutch version of the Pittsburgh Fatigability Scale. After the translation procedure and pretesting phase, we investigated the content validity, construct validity and test-retest reliability of the Dutch version of the Pittsburgh Fatigability Scale in older hospitalized patients.

The second part of this dissertation concerns the actual health status trajectories of both community-dwelling older people and hospitalized older patients. In Chapter 5 we first identified trajectories of general self-rated health over five years. Second, we assessed whether there were demographic, clinical or behavioural determinants that were associated with an increased likelihood of belonging to one of the identified self-rated health trajectories in community-dwelling older people. Lastly, in Chapter 6 we identified trajectories of fatigability and mobility over the course from hospital admission up to six months after hospital discharge and explored the association between these trajectories and demographic and clinical characteristics in older medical patients. The general discussion, in which a summary of the main findings is presented, the methodological strengths and weaknesses are discussed, and recommendations for future research are provided, is presented in Chapter 7.

REFERENCES

1. OECD. Promoting Healthy Ageing. Background Report for the 2019 Japanese G20 Presidency. Paris; 2019.
2. Doekhie KD, de Veer AJ, Rademakers JJ, Schellevis FG, Francke AL. Ouderen van de Toekomst. Verschillen in de Wensen En Mogelijkheden Voor Wonen, Welzijn En Zorg. Utrecht; 2014.
3. Oeppen J, Vaupel JW. Broken Limits to Life Expectancy. *J Palliat Med*. 2002;296(5570):1029-1031.
4. Mackenbach JP, Slobbe L, Looman CWN, Van Der Heide A, Polder J, Garssen J. Sharp upturn of life expectancy in the Netherlands: Effect of more health care for the elderly? *Eur J Epidemiol*. 2011;26(12):903-914. doi:10.1007/s10654-011-9633-y
5. Stoeldraijer L, Duin C Van, Huisman C. Bevolkingsprognose 2017-2060. Den Haag; 2017.
6. Actiz. Arbeidsmarktbeleid in de Zorg. Utrecht; 2020.
7. Gijzen R, Hoeymans N, Schellevis FG, Ruwaard D, Satariano W a., van den Bos G a. M. Causes and consequences of comorbidity. *J Clin Epidemiol*. 2001;54(7):661-674. doi:10.1016/S0895-4356(00)00363-2
8. McPhail SM. Multimorbidity in chronic disease: Impact on health care resources and costs. *Risk Manag Healthc Policy*. 2016;9:143-156. doi:10.2147/RMHP.S97248
9. Galenkamp H, Deeg DJH, de Jongh RT, Kardaun JWPF, Huisman M. Trend study on the association between hospital admissions and the health of Dutch older adults (1995-2009). *BMJ Open*. 2016;6(8):e011967. doi:10.1136/bmjopen-2016-011967
10. Covinsky KE, Pierluissi E, Johnston CB. Hospitalization-Associated Disability " She Was Probably Able to Ambulate, but I'm Not Sure ". *JAMA*. 2011;306(16):1782-1793. www.jama.com.
11. Huntley AL, Chalder M, Shaw ARG, et al. A systematic review to identify and assess the effectiveness of alternatives for people over the age of 65 who are at risk of potentially avoidable hospital admission. *BMJ Open*. 2017;7(7):1-7. doi:10.1136/bmjopen-2017-016236
12. Baztán JJ, Gálvez CP, Socorro A. Recovery of functional impairment after acute illness and mortality: One-year follow-up study. *Gerontology*. 2009;55(3):269-274. doi:10.1159/000193068
13. Bachmann S, Finger C, Huss A, Egger M, Stuck AE, Clough-Gorr KM. Inpatient rehabilitation specifically designed for geriatric patients: Systematic review and meta-analysis of randomised controlled trials. *BMJ*. 2010;340(7758):1230. doi:10.1136/bmj.c1718
14. Verweij L, van de Korput E, Daams JG, et al. Effects of Postacute Multidisciplinary Rehabilitation Including Exercise in Out-of-Hospital Settings in the Aged: Systematic Review and Meta-analysis. *Arch Phys Med Rehabil*. 2019;100(3):530-550. doi:10.1016/j.apmr.2018.05.010
15. CBS Statline. Ziekenhuisopnamen en -patiënten; diagnose-indeling ICD-10.
16. Van Rijn M, Buurman BM, Macneil Vroomen JL, et al. Changes in the in-hospital mortality and 30-day post-discharge mortality in acutely admitted older patients: Retrospective observational study. *Age Ageing*. 2016;45(1):41-47. doi:10.1093/ageing/afv165
17. Boyd CM, Landefeld CS, Counsell SR, Robert M, Fortinsky RH, Kresevic D. Recovery in Activities of Daily Living Among Older Adults Following Hospitalization for Acute Medical Illness. *J Am Geriatr Soc*. 2008;56(12):2171-2179. doi:10.1111/j.1532-5415.2008.02023.x.Recovery
18. Loyd C, Do ADM, Zhang Y, et al. Prevalence of Hospital-Associated Disability in Older Adults : A Meta-analysis. *J Am Med Dir Assoc*. 2019. doi:10.1016/j.jamda.2019.09.015
19. Boyd C, Ricks M, Fried LP, Guralnik JM, Xue Q, Bandeen-Roche K. Functional decline and recovery of activities of daily living in hospitalized, disabled older women: the Women's Health and Aging Study I. *J Am Geriatr Soc*. 2009;57(10):1757-1766. doi:10.1111/j.1532-5415.2009.02455.x.Functional
20. Huang HT, Chang CM, Liu LF, Lin HS, Chen CH. Trajectories and predictors of functional decline of hospitalised older patients. *J Clin Nurs*. 2013;22(9-10):1322-1331. doi:10.1111/jocn.12055
21. Chen CC-H, Wang C, Huang G-H. Functional Trajectory 6 Months Posthospitalization. *Nurs Res*. 2008;57(2):93-100. doi:10.1097/01.nnr.0000313485.18670.e2
22. Inouye SK, Wagner DR, Acampora D, et al. A predictive index for functional decline in hospitalized elderly medical patients. *J Gen Intern Med*. 1993;8(12):645-652. doi:10.1007/BF02598279

23. Wu AW, Damiano AM, Lynn J, et al. Predicting future functional status for seriously ill hospitalized adults. The SUPPORT prognostic model. *Ann Intern Med*. 1995;122(5):342-350. doi:10.7326/0003-4819-122-5-199503010-00004
24. Sager M, Franke T, Inouye SK, et al. Functional Outcomes of Acute Medical Illness and Hospitalization in Older Persons. *arch intern med*. 1996;156(Mar 25):645-652.
25. Covinsky KE, Justice AC, Rosenthal GE, Palmer RM, Landefeld CS. Measuring prognosis and case mix in hospitalized elders: The importance of functional status. *J Gen Intern Med*. 1997;12(4):203-208. doi:10.1046/j.1525-1497.1997.012004203.x
26. McCusker J, Bellavance F, Cardin S, Trépanier S, Verdon J, Ardman O. Detection of older people at increased risk of adverse health outcomes after an emergency visit: The ISAR screening tool. *J Am Geriatr Soc*. 1999;47(10):1229-1237. doi:10.1111/j.1532-5415.1999.tb05204.x
27. Hoogerduijn JG, Schuurmans MJ, Duijnstee MSH, De Rooij SE, Grypdonck MFH. A systematic review of predictors and screening instruments to identify older hospitalized patients at risk for functional decline. *J Clin Nurs*. 2007;16(1):46-57. doi:10.1111/j.1365-2702.2006.01579.x
28. Henoch I, Sawatzky R, Falk H, et al. Symptom distress profiles in hospitalized patients in sweden: A cross-sectional study. *Res Nurs Heal*. 2014;37(6):512-523. doi:10.1002/nur.21624
29. van Seben R, Reichardt LA, Aarden JJ, et al. The Course of Geriatric Syndromes in Acutely Hospitalized Older Adults: The Hospital-ADL Study. *J Am Med Dir Assoc*. 2019;20(2):152-158.e2. doi:10.1016/j.jamda.2018.08.003
30. Eldadah BA. Fatigue and Fatigability in Older Adults. *PM R*. 2010;2(5):406-413. doi:10.1016/j.pmrj.2010.03.022
31. Alexander NB, Taffet GE, McFarland Horne F, et al. Bedside-to-Bench Conference: Research Agenda for Idiopathic Fatigue and Aging. *J Am Geriatr Soc*. 2010;58(5):967-975. doi:10.1111/j.1532-5415.2010.02811.x.
32. LaSorda KR, Gmelin T, Kuipers AL, et al. Epidemiology of Perceived Physical Fatigability in Older Adults: The Long Life Family Study. *J Gerontol A Biol Sci Med Sci*. 2019. doi:10.1093/gerona/glx288.
33. Murphy SL, Smith DM. Ecological measurement of fatigue and fatigability in older adults with osteoarthritis. *J Gerontol A Biol Sci Med Sci*. 2010;65(2):184-189. doi:10.1093/gerona/glp137
34. Simonsick EM, Glynn NW, Jerome GJ, Shardell M, Schrack JA, Ferrucci L. Fatigued, but Not Frail: Perceived Fatigability as a Marker of Impending Decline in Mobility-Intact Older Adults. *J Am Geriatr Soc*. 2016;64(6):1287-1292. doi:10.1111/jgs.14138
35. Wanigatunga AA, Simonsick EM, Zipunnikov V, et al. Perceived Fatigability and Objective Physical Activity in Mid- to Late-Life. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2018;73(5):630-635. doi:10.1093/gerona/glx181
36. WHO. Preamble to the Constitution of the World Health Organization as Adopted by the International Health Conference, 19-22 June 1946; Signed on 22 July 1946 by the Representatives of 61 States (Official Records of the World Health Organization, No. 2, p. 100). New York; 1946.
37. Larson JS. The World Health Organization 's Definition of Health : Social versus Spiritual Health. *Soc Indic Res*. 1996;38(2):181-192.
38. Huber M, Knottnerus JA, Green L, et al. How should we define health? *Bmj*. 2011;343(jul26 2):d4163-d4163. doi:10.1136/bmj.d4163
39. Breslow L. A Quantitative Approach to the World Health Organisation Definition of Health: Physical Mental and Social Well-being. *Int J Epidemiol*. 1972;1:347-355.
40. Kuh D, Karunanathan S, Bergman H, Cooper R. A life-course approach to healthy ageing: maintaining physical capability. *Proc Nutr Soc*. 2014;73(2):237-248. doi:10.1017/S0029665113003923
41. Frankfurt S, Frazier P, Syed M, Jung KR. Using Group-Based Trajectory and Growth Mixture Modeling to Identify Classes of Change Trajectories. *Couns Psychol*. 2016;44(5):622-660. doi:10.1177/0011000016658097
42. Nagin D. Group-Based Modeling of Development. Cambridge, Massachusetts: Harvard University Press; 2005.
43. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol*. 2010;6:109-138. doi:10.1146/annurev.clinpsy.121208.131413

44. Hems M, Harkes M, Moret-Hartman M, Melis RJF, Schoon Y. Eerste ervaringen met patiënt gerapporteerde uitkomstmaten in de geriatrie. *Tijdschr Gerontol Geriatr*. 2017;48(6):287-296. doi:10.1007/s12439-017-0237-1
45. Fried TR, Tinetti M, Agostini J, Iannone L, Towle V. Health outcome prioritization to elicit preferences of older persons with multiple health conditions. *Patient Educ Couns*. 2011;83(2):278-282. doi:10.1016/j.pec.2010.04.032
46. Ramer SJ, McCall NN, Robinson-Cohen C, et al. Health outcome priorities of older adults with advanced CKD and concordance with their nephrology providers' perceptions. *J Am Soc Nephrol*. 2018;29(12):2870-2878. doi:10.1681/ASN.2018060657
47. Spuling SM, Wolff JK, Wurm S. Response shift in self-rated health after serious health events in old age. *Soc Sci Med*. 2017;192:85-93. doi:10.1016/j.socscimed.2017.09.026
48. Leinonen R, Heikkinen E, Jylhä M. Self-rated health and self-assessed change in health in elderly men and women - A five-year longitudinal study. *Soc Sci Med*. 1998;46(4-5):591-597. doi:10.1016/S0277-9536(97)00205-0
49. de Ridder D, Geenen R, Kuijer R, van Middendorp H. Psychological adjustment to chronic disease. *Lancet*. 2008;372(9634):246-255. doi:10.1016/S0140-6736(08)61078-8
50. Hoeymans N, Feskens EJM, Kromhout D, van den Bos GAM. Aging And The Relationship Between Functional Status And Self-Rated Health In Elderly Men. *Soc Sci Med*. 1997;45(10):1527-1536.
51. Henchoz K, Cavalli S, Girardin M. Health perception and health status in advanced old age: A paradox of association. *J Aging Stud*. 2008;22(3):282-290. doi:10.1016/j.jaging.2007.03.002
52. Essink-Bot ML (Marie-L. Health status as a measure of outcome of disease and treatment. 1995.
53. Guyatt GH, Veldhuyzen Van Zanten SJO, Feeny DH, Patrick DL. Measuring quality of life in clinical trials: A taxonomy and review. *Cmaj*. 1989;140(12):1441-1448.
54. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42. doi:10.1016/j.jclinepi.2006.03.012
55. Mokkink LB, Terwee CB, Patrick 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. *Qual Life Res*. 2010;19(4):539-549. doi:10.1007/s11136-010-9606-8
56. De Vet, Henrica C. W. Terwee CB, Knol DL, Mokkink LB. *Measurement in Medicine*. first. Cambridge: Cambridge University Press; 2011.
57. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: A clarification. *J Clin Epidemiol*. 1989;42(5):403-408. doi:10.1016/0895-4356(89)90128-5
58. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis*. 1985;38(1):27-36. doi:10.1016/0021-9681(85)90005-0
59. Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2015;44(4):1172-1180. doi:10.1093/ije/dyu229

CHAPTER 2

Discrepancies in Self-Perceived and
Measured Change in Activities of Daily
Living after Hospital Admission among
Older Adults

Marlies Feenstra
Barbara C van Munster
Nynke Smidt
Sophia E de Rooij

ABSTRACT

Background: Patient-reported outcome measures support shared decision making among hospitalized older patients. It is unclear to what extent change in activities of daily living (ADL) measured with established ADL questionnaires corresponds to the patients' perception of change in ADL among hospitalized older patients. Therefore, the objective was to investigate, first, the agreement between self-perceived and formally measured change in ADL by the 15-item Katz ADL Index from pre-hospital admission to three and twelve months after discharge and second, the association between patient characteristics and congruent versus incongruent outcomes of hospitalized medical and surgical patients aged ≥ 70 years.

Methods: The agreement between change in self-perceived ADL and change in measured ADL was assessed by the linear-weighted kappa. Subsequently, three groups were created: 1. Self-perceived corresponds to measured change in ADL; 2. Self-perceived is worse than measured; 3. Self-perceived is better than measured. Multinomial logistic regression analyses were used to investigate the association between age, sex, education, baseline ADL, frailty, depressive symptoms, comorbidity, the number of readmissions, health locus of control, and the outcome groups.

Results: Among 355 consecutive older hospitalized patients, poor agreement existed between change in self-perceived and measured ADL (linear-weighted kappa: 0.16; 95%CI: 0.07; 0.24). Among the patients with ≥ 1 ADL disability at baseline, older patients less often felt 'worse than measured' (OR: 0.92; 95%CI: 0.86; 0.99), increasing ADL disabilities increases the odds for feeling 'worse than measured' (OR: 1.24; 95%CI: 1.05; 1.46), and higher educated patients less often felt 'better than measured' after three months (OR: 0.35, 95%CI: 0.14; 0.87). After twelve months, a higher baseline comorbidity score was associated with incongruent responses.

Conclusions: Clinicians and researchers should be aware of any discrepancy between self-perceived and measured change in ADL and should avoid making decisions based on the differences in Katz ADL Index score scores only.

BACKGROUND

Maintaining independence is what older adults prioritize as most important health outcome above keeping alive and symptom reduction.^{1,2} Hospital admission of older adults jeopardizes independence, because it increases the risk for functional decline in activities of daily living (ADL) both during and after hospitalization.³ Older patients who develop hospital-associated functional decline in ADL are at increased risk for institutionalization^{4,5} and premature mortality.^{6,7} Overall prevalence rates of functional decline in ADL among hospitalized older patients is 30%.⁸

Recovery of functional decline in ADL is often operationalized through established questionnaires measuring the change in self-reported disabilities in ADL before and after hospitalization.^{9–12} From these studies it remained unclear what the patients' subjective perceptions were concerning deterioration or improvements in ADL. Other studies investigating the relation between longitudinal assessment of disabilities in ADL and older adults' perceptions of disability demonstrated that older adults overcome perceptions of disability by using effective coping or adaptation strategies.^{13,14} Activities, such as driving cessation and receiving home care, seem to play a key role in older adults' perceptions of disability.¹⁵ Moreover, female sex, lower income, more chronic conditions, cognitive impairment, more health anxiety, poor self-rated health, higher number of baseline ADL limitations, and an increase in ADL limitations over the past years were associated with an accelerated perception of disability among community-dwelling older adults.¹⁵ Whether these associations also exist for self-perceived in relation to formally measured change in ADL among hospitalized patients has not been investigated yet.

Therefore, our objective is to investigate the agreement of self-perceived and measured change in ADL status from pre-hospital admission to three and twelve months after discharge. Second, potential associations with demographic and clinical characteristics in congruent (self-perceived corresponds to measured change in ADL) and incongruent (when self-perceived change was better or worse than measured change in ADL) outcomes were investigated.

METHODS

Setting and patient selection

For this monocentric study, all medical and surgical patients who were admitted to the hospital from February 2017 to November 2018 were subject for the study. Inclusion criteria for participation were age of 70 years and older and a hospital stay for at least two consecutive days. Exclusion criteria were no understanding of the Dutch spoken language, any (temporary) cognitive condition that influence decision making (e.g. delirium, and diagnosed dementia), proxy interview, and no written informed consent. Baseline assessment took place within the first four days of hospital admission by face-to-face interview, administered by a trained research nurse. Follow-up assessments took place three and twelve months after discharge by telephone. The local research ethics committee has decided this study was not subject to the Dutch Medical Research Human Subjects Act (registration number: 201600268).

Outcome measures

The primary outcomes included self-perceived and measured change in ADL from pre-hospital admission to three and twelve months after discharge.

Change in self-perceived ADL was assessed at follow-up assessments by the question: 'In general, how is your daily functioning now compared to three months ago (one year ago) before hospital admission?' Answer options were much worse, slightly worse, more or less the same, slightly better, and much better. Slightly and much worse were considered deteriorated, more or less the same was considered unchanged, and much and slightly better were considered improved.

Measured change in ADL status was calculated by subtracting the outcomes of the modified Katz ADL Index score (Katz-15)^{16,17} at follow-up assessments from baseline assessment (i.e. the situation two weeks before hospital admission). The total score of the Katz-15 represents the number of self-reported ADL disabilities, ranging from 0 to 15 points, with higher scores indicating more ADL disabilities. Negative change scores were considered deteriorated, change scores of zero were considered unchanged, and positive change scores were considered improved.

With the outcomes of self-perceived and measured change in ADL three groups were created:

1. Patients whose change in self-perceived ADL status (P) corresponds to the measured change (M) in ADL ($P=M$).
2. Patients whose self-perceived change in ADL status was better than the measured change in ADL ($P>M$).
3. Patients whose self-perceived change in ADL status was worse than measured change in ADL ($P<M$).

Co-variables

A number of co-variables (demographic, clinical, and personal characteristics) were hypothesized to be associated with congruent and incongruent outcomes of self-perceived and measured change in ADL.

Demographic characteristics included age, sex (male, female), and level of education ($\leq$ high school (which equals ≤ 12 years of education), $>$ high school (>12 years of education)).

Clinical characteristics included:

- Disabilities in ADL up to two weeks before hospital admission (functional independent, at least one disability), retrospectively assessed with the Katz-15.
- Pre-hospital frailty status (frail, non-frail), assessed with the self-reported Frailty Phenotype.^{18,19} The items of the Frailty Phenotype included unintended weight loss, exhaustion, slowness, weakness, and low physical activity up to two weeks before hospital admission. Unintended weight loss was considered when people lost weight for three or more kilograms in the past month. Exhaustion was assumed when people scored 'yes' to both questions 'everything I did was an effort', and 'I could not get going'. Slowness was assumed when people were not able to walk outside for five minutes. Weakness was assumed when people

reported to have difficulties raising a chair. Low physical activity was assigned when people report not being physically active for at least 30 minutes per week. Total scores ranged from 0 to 5. We used a dichotomous scoring considering a score of 3 or more as frail.¹⁸

- Comorbidity and the severity of comorbidities were assessed by the Charlson Comorbidity Index (CCI) with higher scores indicating higher number or more severe chronic conditions (range 0 – 37 points).²⁰
- Depressive symptoms (present, absent) were assessed with the two-item Geriatric Depression Scale (GDS-2) using a cut-off of one or more.²¹
- The question 'Have you been admitted to the hospital since our last interview?' was asked to determine dichotomously whether there have been hospital readmissions between baseline and follow-up (yes, no).

Personal characteristics included:

- Health locus of control assessed by the Dutch version of the Multidimensional Health Locus of Control (MHLC).²² This scale is used to identify the origin of motivations underlying health-related behavior and exists of three dimensions: 1. Internal; 2. External by powerful others; and, 3. External by chance.²³ Each dimension has six items with a six point Likert scale. Scores ranged from 0 to 36 for each subscale, with higher scores indicating a higher belief in that dimension of health locus of control.

Statistical analysis

Baseline descriptive statistics of the total sample and per subgroup were presented as means with standard deviations, medians with inter quartile range, and counts with percentages, as appropriate based on data distribution.

Agreement between self-perceived and measured change in ADL was investigated by calculating both absolute agreement and the linear-weighted kappa statistic.²⁴ Next multinomial regression analyses were performed using the three created groups as dependent variable. The congruent group (P=M) was used as the reference category. First,

Figure 1. Flowchart of study participants.

univariable regression analyses for each co-variable were performed. Second, a multivariable model was built by sequentially adding demographic co-variables (Model 1), clinical characteristics (Model 2), and personal characteristics (Model 3). Model fit was evaluated with the Likelihood Ratio Test, and goodness-of-fit was based on difference in -2 Log Likelihood using the Chi2 distribution and a p-value of 0.05.²⁵ Data were presented as odds ratios (OR) and corresponding 95% confidence intervals (CI). The analyses were repeated for the data collected twelve months after discharge.

Subjects with missing values for the main outcome or data derived by proxy-interviews were excluded from the analysis, as imputation was not desirable for the subjective outcomes. Missing data for items of the CCI (2%), frailty phenotype (0–4%), GDS-2 (4%), and MHLC (7–9%) were assumed to be missing at random and therefore imputed using multiple imputation (10 imputations, fully conditional specification, predictive mean matching). Rubin's Rules were applied for pooling the results.^{26,27} Predefined subgroup analyses were performed using patients without disabilities at baseline and patients with one or more disabilities at baseline. All analyses were performed using IBM Social Package of Statistical Software version 23.

RESULTS

A total of 478 out of 2961 eligible patients had a baseline assessment. In total, 123 patients were excluded from the analyses, of which 64 were deceased (52%) and 59 withdrew consent (48%). The analytic samples included 355 and 333 patients three and twelve months after discharge, respectively (Figure 1). Excluded patients were more often frail and had more ADL disabilities at baseline compared to the included patients (Supplementary Table S1). Baseline demographic, clinical, and personal characteristics of the total analytic sample and the created subgroups are presented in Table 1a and Table 1b.

Absolute agreement between change in self-perceived and measured ADL from pre-hospital admission to three and twelve months after

discharge was 46% (162/355) and 44% (146/333), respectively. Corresponding linear-weighted kappa's were 0.16 (95%CI: 0.07; 0.24) and 0.13 (95%CI: 0.04; 0.21) indicating poor agreement between self-perceived and measured change in ADL over both the three and twelve months periods (Table 2).

Model 2, including demographics and clinical characteristics, was selected as best fitting multivariable model. The results of the univariable multinomial regression analyses and the outcomes of the unselected models after three, and after twelve months are presented in Supplementary Tables S2, S3, and S4, respectively. Results of the multivariable multinomial regression analyses showed that, compared to the congruent group ($P=M$), patients with existing ADL disabilities before hospital admission rated their self-perceived change in daily functioning after hospitalization more often as worse than measured ($P<M$: OR: 2.97, 95%CI: 1.52; 5.80 and OR: 4.69, 95%CI: 2.43; 9.05) and less often better than measured ($P>M$: OR: 0.56 95%CI: 0.32; 0.98 and OR: 0.49; 95%CI: 0.26; 0.92) both three and twelve months after discharge, respectively (Table 3). Three months after discharge, patients with a higher educational level less often rated their self-perceived change in daily functioning as better than measured ($P>M$) (OR: 0.44, 95%CI: 0.26; 0.77) (Table 3).

Of all cases, 46% reported no disabilities at baseline. As these patients could not improve anymore (floor effect), we decided to perform a subgroup analysis including patients with at least one disability. In this subgroup, the effect of educational level remained ($P>M$: OR: 0.35; 95%CI: 0.14; 0.87), and an additional effect of age was observed: younger patients less often rated their self-perceived change in daily functioning as worse than measured three months after discharge ($P<M$: OR: 0.92, 95%CI: 0.86; 0.99) (Table 4). In addition, we found that patients with more baseline comorbidities more often reported incongruent outcomes after twelve months ($P<M$: OR: 1.31, 95%CI: 1.05; 1.62; $P>M$: OR: 1.41, 95%CI: 1.10; 1.82) (Table 4), which we did not find among the subgroup of patients without disabilities at baseline (Supplementary Table S6). Contingency tables of the subgroups are presented in Supplementary Tables S5 and S7.

Table 1a. Characteristics of the total, congruent (P=M), 'perceived is better than measured' (P>M), and 'perceived is worse than measured' (P<M) groups three months after discharge.

	Total sample	3 months after discharge		
	n = 355	P=M (n = 162)	P>M (n = 117)	P<M (n = 76)
Sociodemographic characteristics				
Age in years, median (IQR)	76 (72; 80)	76 (72; 80)	76 (72; 81)	75 (71; 79)
Sex, male	233 (66)	111 (48)	75 (32)	47 (20)
Low educated (≤ high school)	240 (68)	97 (40)	90 (38)*	53 (22)
Clinical characteristics				
Functional independent (Katz-15 = 0)	163 (46)	76 (47)	68 (42)**	19 (12)**
Frail (Frailty Phenotype ≥ 3)	45 (13)	17 (38)**	14 (31)**	14 (31)**
Comorbidity, median (IQR)	2 (1; 3)	2 (1; 3)	2 (1; 3)	2 (1; 4)
Depressive symptoms (GDS-2 ≥ 1)	115 (32)	59 (51)**	31 (27)**	25 (22)**
Hospital readmission	79 (22)	36 (46)**	21 (27)**	22 (28)**
Personal characteristics				
Health locus of control				
Internal, mean (SD)	21.5 (5.3)	21.3 (5.5)	21.8 (4.9)	21.4 (5.6)
Others, mean (SD)	20.7 (5.3)	20.8 (5.1)	20.9 (5.2)	20.0 (5.9)
Chance, mean (SD)	20.2 (5.4)	20.5 (5.3)	20.0 (5.4)	20.1 (5.9)

All numbers are presented in n (%) unless indicated otherwise.
Inequalities of medians and proportions of the co-variable within each subgroup relative to the total group are indicated by * (p<0.05) and ** (p<0.01).

DISCUSSION

This study investigated how older hospitalized patients' perceptions of change in daily functioning correspond to measured change in ADL comparing the pre-hospital admission situation to three and twelve months after discharge. We found poor agreement between self-perceived and measured change in ADL, both three and twelve months after discharge. Patients who rated themselves as deteriorated in daily functioning even though the Katz ADL Index score indicated no change (P=M) or even improvement (P<M) was observed in 21% and 28% of the cases after three and twelve months, respectively. These results are

Table 1b. Characteristics of the total, congruent (P=M), 'perceived is better than measured' (P>M), and 'perceived is worse than measured' (P<M) groups twelve months after discharge.

	Total sample	12 months after discharge		
	n = 333	P=M (n = 146)	P>M (n = 95)	P<M (n = 92)
Sociodemographic characteristics				
Age in years, median (IQR)	75 (72; 80)	75 (72; 80)	76 (72; 81)	75 (72; 81)
Sex, male	221 (66)	97 (44)	61 (28)	63 (29)
Low educated (≤ high school)	220 (66)	98 (45)	62 (28)	60 (27)
Clinical characteristics				
Functional independent (Katz-15 = 0)	162 (49)	79 (49)	62 (38)*	21 (13)**
Frail (Frailty Phenotype ≥ 3)	39 (12)	15 (38)**	10 (26)**	14 (36)**
Comorbidity, median (IQR)	2 (1; 3)	2 (1; 3)	2 (1; 4)	2 (1; 4)
Depressive symptoms (GDS-2 ≥ 1)	103 (31)	42 (41)**	29 (28)**	32 (31)**
Hospital readmission	84 (36)	37 (44)**	20 (24)**	27 (32)**
Personal characteristics				
Health locus of control				
Internal, mean (SD)	21.3 (5.3)	21.4 (5.4)	22.2 (4.9)	20.0 (5.5)
Others, mean (SD)	20.8 (5.3)	20.8 (5.4)	21.6 (4.5)	20.0 (5.9)
Chance, mean (SD)	20.4 (5.3)	20.2 (5.2)	20.5 (5.1)	20.4 (5.7)

All numbers are presented in n (%) unless indicated otherwise.
Inequalities of medians and proportions of the co-variable within each subgroup relative to the total group are indicated by * (p<0.05) and ** (p<0.01).

consistent with previous research among community-dwelling older people that demonstrated an overestimation of self-reported ADL instruments compared to the actual performance of daily activities.^{28,29} Instead of actual ADL performance, in this study we compared the change in self-reported ADL relative to self-perceived daily functioning by using an anchor question. Around 20% to 25% of the patients with disabilities at baseline rated their daily functioning as more improved compared to the Katz ADL Index score (P>M), both three and twelve months after discharge. Misclassification in this direction has been previously reported,

Table 2. Contingency tables presenting the frequency distributions (n) of the agreement between change in self-perceived and measured ADL three (a) and twelve (b) months after discharge.

a. Three months after discharge		Self-perceived change in ADL			Total
Measured change in ADL (Katz-15)	Deteriorated	Deteriorated	Unchanged	Improved	
	Unchanged	44	45	23	112
	Improved	36	92	49	177
		16	24	26	66
Total		96	161	98	355

b. Twelve months after discharge		Self-perceived change in ADL			Total
Measured change in ADL (Katz-15)	Deteriorated	Deteriorated	Unchanged	Improved	
	Unchanged	45	43	13	101
	Improved	41	84	39	164
		19	32	17	68
Total		105	159	69	333

albeit a smaller proportion.³⁰ A possible explanation may be the effect of recalibration or reprioritization of the construct ‘ADL’. Such adaptation mechanisms are known as ‘response shift’³¹, which has been previously demonstrated in other studies concerning older community-dwelling adults²⁸ and stroke survivors.³²

In line with the study of Kelley-Moore and Schumacher (2007)¹⁵, we found that pre-admission ADL disabilities were an important predictor of patients to rate their daily functioning as deteriorated compared to formally measured ADL. They also demonstrated that more prevalent medical conditions were associated with increased perception of disability¹⁵, which we identified only twelve months after discharge in the subgroup analysis of patients with one or more disability at baseline. Next to the different target population, the differences in study design may explain the differences: our study compared congruent and incongruent outcomes of self-perceived and measured change in ADL, while Kelley-Moore & Schumacher (2007)¹⁵ investigated predictors associated with perceptions of disability. In addition, we found that higher educated patients less often rated their change in daily functioning as better than measured three months after discharge. A possible explanation

for this association is that higher levels of education are associated with better acceptance of disease and disability, and more effective coping strategies.^{33,34}

Of course, this study has its limitations. First, to our knowledge, there are no studies that investigated the association between patient characteristics of congruent versus incongruent outcomes of self-perceived and measured change in ADL among older hospitalized patients. Consequently, an existing theory or model was not available for the selection of co-variables and other potential variables that contribute to the variance underlying the concept studied are probably missing. This is also indicated by the low pseudo R-squared found in the current study. Second, only baseline clinical and patient characteristics were considered as co-variables whereas time-varying co-variables assessing the change in frailty status, depressive symptoms, and number of comorbidities would have contributed to the model as well. Third, due to our selection criteria (i.e. excluding cognitive vulnerable patients), our study population concerned a select patient group of relatively fit hospitalized older adults. The study population was less frail and had less baseline ADL disabilities than the excluded subjects. Even though our subgroup analysis included a group of patients with baseline disabilities, the results cannot be generalized to specific vulnerable hospitalized older patients.

Despite the limitations, our results have important clinical implications. Clinicians and researchers should take notion on the high proportions of misclassification between change scores of the Katz ADL Index and patients’ perceptions of change in ADL in both directions, even after correcting for the floor effect. In other words, detecting change in ADL status based on calculation of disabilities according to the Katz ADL Index score does not necessarily mean that a patient perceived himself deteriorated or improved as well. Without taking this phenomenon into account, the effectiveness of interventions may be over- or underestimated.^{35,36} For instance, when the objective of an intervention trial is to increase mobility, this might be best assessed using objective and responsive instruments.³⁷ Conversely, when the objective is to investigate a patient’s perceptions of (in)dependency after hospitalization or their perceptions of adaptation to loss in ADL, goal attainment or regular patient-doctor dialogues would suffice as informative outcomes.³⁸

Table 3. Results of the multinomial regression analyses three and twelve months after discharge (Model 2): odds ratio's (OR) and 95% confidence intervals (CI) of co-variables associated with incongruent groups (P>M and P<M) relative to the congruent group (P=M).

	3 months AD [†] n = 355 OR (95% CI)	12 months AD [‡] n = 333 OR (95% CI)
Group: self-perceived change in ADL is worse than measured change in ADL (P<M)		
Age, per year	0.96 (0.91; 1.01)	0.98 (0.94; 1.04)
Male	1.28 (0.68; 2.40)	0.80 (0.43; 1.46)
High educated (>high school)	0.66 (0.35; 1.23)	1.13 (0.63; 2.03)
Baseline Disability, per disability	2.97 (1.52; 5.80)**	4.69 (2.43; 9.05)**
Frail (Frailty Phenotype ≥3)	1.56 (0.65; 3.73)	1.04 (0.45; 2.43)
Comorbidity, per comorbidity	1.13 (0.99; 1.29)	1.07 (0.93; 1.24)
Depressive symptoms (GDS-2 ≥1)	0.57 (0.29; 1.12)	0.97 (0.51; 1.83)
Hospital readmission	1.31 (0.67; 2.54)	0.89 (0.47; 1.68)
Group: self-perceived change in ADL is better than measured change in ADL (P>M)		
Age, per year	1.02 (0.97; 1.07)	1.04 (0.99; 1.09)
Male	1.34 (0.78; 2.30)	1.39 (0.78; 2.48)
High educated (>high school)	0.44 (0.26; 0.77)**	1.06 (0.61; 1.86)
Disability (Katz-15 ≥1)	0.56 (0.32; 0.98)*	0.49 (0.26; 0.92)*
Frail (Frailty Phenotype ≥3)	1.70 (0.73; 3.96)	1.29 (0.50; 3.32)
Comorbidity, per comorbidity	1.04 (0.92; 1.17)	1.10 (0.96; 1.25)
Depressive symptoms (GDS-2 ≥1)	0.61 (0.34; 1.07)	1.12 (0.61; 2.06)
Hospital readmission	0.77 (0.41; 1.45)	0.84 (0.44; 1.62)

Reference group is self-perceived and measured change in ADL are congruent.
Abbreviations: AD, After Discharge; ADL, Activities of Daily Living; GDS, Geriatric Depression Scale.
* p<0.05; ** p<0.01.
[†] Model Chi²(16): 45.69, p <0.01. Pseudo R²: 0.06 (McFadden). -2LL: 660.636.
[‡] Model Chi²(16): 51.06, p <0.01. Pseudo R²: 0.07 (McFadden). -2LL: 631.400.

Table 4. Results of the multinomial regression subgroup analyses including patient with ≥1 disability at baseline: odds ratio's (OR) and 95% confidence intervals (CI) of co-variables associated with incongruent groups (P>M and P<M) relative to the congruent group (P=M).

	3 months AD [†] n = 192 OR (95% CI)	12 months AD [‡] n = 171 OR (95% CI)
Group: self-perceived change in ADL is worse than measured change in ADL (P<M)		
Age, per year	0.92 (0.86; 0.99)*	0.98 (0.93; 1.05)
Male	1.06 (0.47; 2.41)	0.59 (0.28; 1.24)
High educated (>high school)	0.57 (0.26; 1.28)	1.00 (0.46; 2.16)
Disability, per disability	1.24 (1.05; 1.46)*	1.06 (0.90; 1.24)
Frail (Frailty Phenotype ≥3)	0.83 (0.30; 2.32)	1.11 (0.41; 3.03)
Comorbidity, per comorbidity	1.14 (0.96; 1.37)	1.31 (1.05; 1.62)*
Depressive symptoms (GDS-2 ≥1)	0.66 (0.30; 1.46)	0.87 (0.39; 1.95)
Hospital readmission	0.76 (0.32; 1.80)	0.70 (0.33; 1.49)
Group: self-perceived change in ADL is better than measured change in ADL (P>M)		
Age, per year	0.99 (0.93; 1.06)	1.03 (0.95; 1.10)
Male	1.99 (0.87; 4.53)	1.04 (0.41; 2.67)
High educated (>high school)	0.35 (0.14; 0.87)*	1.30 (0.50; 3.35)
Disability (Katz-15 ≥1)	1.06 (0.88; 1.27)	0.85 (0.66; 1.09)
Frail (Frailty Phenotype ≥3)	1.41 (0.48; 4.11)	1.80 (0.49; 6.59)
Comorbidity, per comorbidity	1.09 (0.89; 1.34)	1.41 (1.10; 1.82)**
Depressive symptoms (GDS-2 ≥1)	0.45 (0.19; 1.05)	1.09 (0.41; 2.92)
Hospital readmission	0.54 (0.20; 1.47)	0.47 (0.17; 1.31)

Reference group is self-perceived and measured change in ADL are congruent.
Abbreviations: AD, After Discharge; ADL, Activities of Daily Living; GDS, Geriatric Depression Scale.
* p<0.05; ** p<0.01.
[†] Model Chi²(16): 30.71, p:0.02. Pseudo R²: 0.08 (McFadden). -2LL: 367.220.
[‡] Model Chi²(16): 21.83 p:0.15. Pseudo R²: 0.06 (McFadden). -2LL: 330.552.

CONCLUSIONS

In summary, this study demonstrated poor agreement between self-perceived and formally measured change in ADL status from pre-hospital admission compared to three and twelve months after hospital discharge. Pre-hospital ADL disabilities, a higher comorbidity index score, lower age, and a lower educational level were associated with incongruent outcomes, but future research should elucidate these associations. Clinicians and researchers should be aware of any discrepancy between self-perceived and measured change in ADL status and should avoid making decisions based on the differences in Katz ADL Index score scores only.

REFERENCES

1. Fried TR, Tinetti M, Agostini J, Iannone L, Towle V. Health outcome prioritization to elicit preferences of older persons with multiple health conditions. *Patient Educ Couns*. 2011;83(2):278-282. doi:10.1016/j.pec.2010.04.032
2. Ramer SJ, McCall NN, Robinson-Cohen C, et al. Health outcome priorities of older adults with advanced CKD and concordance with their nephrology providers' perceptions. *J Am Soc Nephrol*. 2018;29(12):2870-2878. doi:10.1681/ASN.2018060657
3. Covinsky KE, Pierluissi E, Johnston CB. Hospitalization-Associated Disability " She Was Probably Able to Ambulate, but I'm Not Sure ". *JAMA*. 2011;306(16):1782-1793.
4. Portegijs E, Buurman BM, Essink-Bot ML, Zwinderman AH, de Rooij SE. Failure to Regain Function at 3 months After Acute Hospital Admission Predicts Institutionalization Within 12 Months in Older Patients. *J Am Med Dir Assoc*. 2012;13(6):569.e1-569.e7. doi:10.1016/j.jamda.2012.04.003
5. Luppá M, Luck T, Weyerer S, König HH, Brähler E, Riedel-Heller SG. Prediction of institutionalization in the elderly. A systematic review. *Age Ageing*. 2009;39(1):31-38. doi:10.1093/ageing/afp202
6. Walter LC, Brand RJ, Counsell SR, et al. Development and validation of a prognostic index for 4-year mortality in older adults. *J Am Med Assoc*. 2001;285(23):2987-2994. doi:10.1001/jama.295.7.801
7. Torisson G, Stavenow L, Minthon L, Londos E. Importance and added value of functional impairment to predict mortality: A cohort study in Swedish medical inpatients. *BMJ Open*. 2017;7(5):1-8. doi:10.1136/bmjopen-2016-014464
8. Loyd C, Do ADM, Zhang Y, et al. Prevalence of Hospital-Associated Disability in Older Adults : A Meta-analysis. *J Am Med Dir Assoc*. 2019. doi:10.1016/j.jamda.2019.09.015
9. Dasgupta M, Brymer C. Poor functional recovery after delirium is associated with other geriatric syndromes and additional illnesses. *Int Psychogeriatrics*. 2015;27(5):793-802. doi:10.1017/S1041610214002658
10. Huang HT, Chang CM, Liu LF, Lin HS, Chen CH. Trajectories and predictors of functional decline of hospitalised older patients. *J Clin Nurs*. 2013;22(9-10):1322-1331. doi:10.1111/jocn.12055
11. Barnes DE, Mehta KM, Boscardin WJ, et al. Prediction of recovery, dependence or death in elders who become disabled during hospitalization. *J Gen Intern Med*. 2013;28(2):261-268. doi:10.1007/s11606-012-2226-y
12. Boyd CM, Landefeld CS, Counsell SR, Robert M, Fortinsky RH, Kresevic D. Recovery in Activities of Daily Living Among Older Adults Following Hospitalization for Acute Medical Illness. *J Am Geriatr Soc*. 2008;56(12):2171-2179. doi:10.1111/j.1532-5415.2008.02023.x.Recovery
13. Gignac MAM, Cott C, Badley EM. Adaptation To Disability and Its Relationship To Perceived Independence and Dependence. *Gerontologist*. 1999;39(6):125.
14. Lorenz RA. Coping with preclinical disability: Older women's experiences of everyday activities. *J Nurs Scholarsh*. 2010;42(4):439-447. doi:10.1111/j.1547-5069.2010.01339.x
15. Kelley-Moore JA, Schumacher JG. When do older adults become disabled? 2007;47(2):126-141.
16. Weinberger M, Samsa GP, Schmader K, Greenberg SM, Carr DB, Wildman DS. Comparing Proxy and Patients' Perceptions of Patients' Functional Status: Results from an Outpatient Geriatric Clinic. *J Am Geriatr Soc*. 1992;40(6):585-588. doi:10.1111/j.1532-5415.1992.tb02107.x
17. Laan W, Zuithoff NPA, Drubbel I, et al. Validity and reliability of the Katz-15 scale to measure unfavorable health outcomes in community-dwelling older people. *J Nutr Health Aging*. 2014;18(9):848-854. doi:10.1007/s12603-014-0479-3
18. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: Evidence for a phenotype. *Journals Gerontol Ser a-Biological Sci Med Sci*. 2001;56(3):M146-M156. doi:10.1093/gerona/56.3.M146
19. Theou O, Rockwood K. Comparison and Clinical Applications of the Frailty Phenotype and Frailty Index Approaches. *Interdiscip Top Gerontol Geriatr*. 2015;41:74-84. doi:10.1159/000381166
20. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of

- classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40(5):373-383. <http://www.ncbi.nlm.nih.gov/pubmed/3558716>
21. Cully JA, Graham DP, Kramer JR. A 2-item screen for depression in rehabilitation inpatients. *Arch Phys Med Rehabil.* 2005;86(3):469-472. doi:10.1016/j.apmr.2004.04.042
 22. Halfens RJG. Locus of Control, de Beheersingsorientatie in Relatie Tot Ziekte En Gezondheidsgedrag [Locus of Control and Health and Illness Behavior]. Maastricht, The Netherlands; 1985.
 23. Wallston K a, Wallston BS, Devellis R. Development of the Multidimensional Development. *Health Educ Monogr.* 1978;6(2):160-170.
 24. Warrens MJ. Conditional inequalities between Cohen's kappa and weighted kappas. *Stat Methodol.* 2013;10(1):14-22. doi:10.1016/j.stamet.2012.05.004
 25. Field A. *Discovering Statistics Using SPSS.* Third edit. London: SAGE Publications Ltd; 2000.
 26. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons, Inc.; 1987. doi:10.1002/9780470316696
 27. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30(4):377-399. doi:10.1002/sim.4067
 28. Daltroy LH, Larson MG, Eaton HM, Phillips CB, Liang MH. Discrepancies between self-reported and observed physical function in the elderly: The influence of response shift and other factors. *Soc Sci Med.* 1999;48(11):1549-1561. doi:10.1016/S0277-9536(99)00048-9
 29. Kelly-Hayes M, Jette AM, Wolf PA, D'Agostino RB, Odell PM. Functional limitations and disability among elders in the Framingham study. *Am J Public Health.* 1992;82(6):841-845. doi:10.2105/AJPH.82.6.841
 30. Suijker JJ, Rijn M van, Riet G ter, Charante EPM van, Rooij SE de, Buurman BM. Minimal Important Change and Minimal Detectable Change in Activities of Daily Living in Community-Living Older People. *J Nutr Health Aging.* 2017;21(2):165-172.
 31. Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med.* 1999;48(1):1507-1515.
 32. Barclay-Goddard R, Lix LM, Tate R, Weinberg L, Mayo NE. Health-related quality of life after stroke: Does response shift occur in self-perceived physical function? *Arch Phys Med Rehabil.* 2011;92(11):1762-1769. doi:10.1016/j.apmr.2011.06.013
 33. Richardson A, Adner N, Nordström G. Persons with insulin-dependent diabetes mellitus: Acceptance and coping ability. *J Adv Nurs.* 2001;33(6):758-763. doi:10.1046/j.1365-2648.2001.01717.x
 34. Cano A, Mayo A, Ventimiglia M. Coping, Pain Severity, Interference, and Disability: The Potential Mediating and Moderating Roles of Race and Education. *J Pain.* 2006;7(7):459-468. doi:10.1016/j.jpain.2006.01.445
 35. Mcphail BS, Haines T. "Can You Handle the Truth?" The Response Shift Phenomenon in Clinical Trials. *J Clin Res Best Pract.* 2010;6(2):1-8.
 36. Schwartz CE, Sprangers MAG. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med.* 1999;48(11):1531-1548. doi:10.1016/S0277-9536(99)00047-7
 37. Soares Menezes KVR, Auger C, de Souza Menezes WR, Guerra RO. Instruments to evaluate mobility capacity of older adults during hospitalization: A systematic review. *Arch Gerontol Geriatr.* 2017;72(May):67-79. doi:10.1016/j.archger.2017.05.009
 38. van Seben R, Reichardt L, Smorenburg S, Buurman BM. Goal-Setting Instruments in Geriatric Rehabilitation: A Systematic Review. *J Frailty Aging.* 2017;6(1):37-45. doi:10.14283/jfa.2016.103.

SUPPLEMENTARY FILES – CHAPTER 2

Table S1. Comparison of baseline sociodemographic and patient characteristics patients included and excluded for the analyses.

	Included ¹ (n = 355)	Excluded (n = 123)	P value ²
Sociodemographic characteristics			
Age in years	75 (72; 80) [70–95]	77 (74; 84) [70–98]	.28
Sex, male	232 (65)	85 (69)	.45
Education ≤ high school	239 (67)	89 (72)	.34
Clinical characteristics			
Functional independent (Katz-15 = 0)	163 (46)	39 (32)	.01
Frail (Frailty Phenotype ≥ 3)	45 (13)	22 (18)	.05
Comorbidity, median (IQR)	2 (1; 3) [0–10]	2 (1; 4) [0–11]	.17
Depressive symptoms (GDS-2 ≥ 1)	115 (32)	38 (31)	.50
Hospital readmission in past 3 months	70 (20)	n.a.	-
Personal characteristics			
Health locus of control			
Internal, mean (SD)	21.4 (5.4)	21.8 (5.1)	.52
Others, mean (SD)	20.6 (5.4)	20.6 (4.9)	.96
Chance, mean (SD)	20.2 (5.5)	19.0 (5.5)	.06

All numbers are presented in n (%) or median (IQR) [range], unless indicated otherwise.
1. Patients who had three months follow-up data
2. Differences between included and excluded samples were tested by Kruskal Wallis and Chi² tests.
Abbreviations: GDS, Geriatric Depression Scale; IQR, interquartile range;

Table S2. Univariate multinomial regression results: odds ratio's and 95%CI of pooled estimates.

	3 months AD n = 355	12 months AD n = 333
Group: self-perceived is worse than measured change in ADL (P<M)		
Age, per year	0.97 (0.92; 1.02)	1.02 (0.97; 1.07)
Male (ref: female)	1.34 (0.76; 2.37)	0.91 (0.52; 1.59)
High educated (ref: ≤ high school)	0.65 (0.36; 1.16)	1.09 (0.63; 1.89)
Disability (ref: independent)	2.65 (1.45; 4.85)**	3.99 (2.22; 7.16)**
Frail (ref: not frail)	1.92 (0.89; 4.13)	1.40 (0.65; 3.01)
Comorbidity, per comorbidity	1.10 (0.98; 1.24)	1.09 (0.96; 1.24)
Depression (ref: no depression)	0.90 (0.50; 1.62)	1.30 (0.74; 2.30)
Readmission (ref: no readmission)	1.45 (0.78; 2.69)	1.23 (0.69; 2.21)
Health locus of control, per point		
Internal	1.01 (0.96; 1.07)	0.95 (0.90; 1.00)
Others	0.98 (0.93; 1.04)	0.98 (0.93; 1.03)
Chance	0.99 (0.94; 1.05)	1.01 (0.96; 1.07)
Group: self-perceived is better than measured change in ADL (P>M)		
Age, per year	1.00 (0.96; 1.04)	1.02 (0.97; 1.07)
Male (ref: female)	1.22 (0.74; 2.02)	1.10 (0.64; 1.90)
High educated (ref: ≤ high school)	0.45 (0.26; 0.76)**	1.09 (0.63; 1.88)
Disability (ref: independent)	0.64 (0.39; 1.03)	0.63 (0.37; 1.07)
Frail (ref: not frail)	1.15 (0.54; 2.46)	0.90 (0.39; 2.11)
Comorbidity, per comorbidity	0.98 (0.88; 1.10)	1.08 (0.95; 1.22)
Depression (ref: no depression)	0.63 (0.38; 1.06)	1.02 (0.58; 1.78)
Readmission (ref: no readmission)	0.75 (0.41; 1.36)	0.79 (0.43; 1.47)
Health locus of control, per point		
Internal	1.02 (0.97; 1.07)	1.02 (0.97; 1.08)
Others	1.01 (0.96; 1.06)	1.02 (0.97; 1.08)
Chance	0.99 (0.94; 1.04)	1.00 (0.95; 1.05)

Abbreviations: AD, after discharge

Table S3. Results of the multinomial regression analyses three months after discharge: odds ratio's and 95% confidence intervals of incongruent outcomes between self-perceived and measured change in ADL relative to congruent outcomes.

	Model 1 ^a n = 355	Model 2 n = 355	Model 3 ^b n = 355
Group: self-perceived is worse than measured change in ADL (P<M)			
Age, per year	0.97 (0.92; 1.03)		0.96 (0.91; 1.01)
Male	1.25 (0.70; 2.23)		1.30 (0.67; 2.53)
High educated (>high school)	0.67 (0.37; 1.20)		0.68 (0.36; 1.28)
Disability (Katz-15 ≥1)			2.99 (1.47; 6.10)**
Frail (Frailty Phenotype ≥3)		<i>Results of Model</i>	1.57 (0.66; 3.77)
Comorbidity, per comorbidity		<i>2 are presented in</i>	1.13 (0.98; 1.29)
Depressive symptoms (GDS-2 ≥1)		<i>the original paper</i>	0.57 (0.29; 1.11)
Hospital readmission		<i>(Table 3)</i>	1.30 (0.67; 2.54)
Health locus of control, per point			
Internal			1.01 (0.95; 1.07)
Others			0.98 (0.92; 1.05)
Change			1.00 (0.94; 1.06)
Group: self-perceived is better than measured change in ADL (P>M)			
Age, per year	1.00 (0.96; 1.04)		1.02 (0.97; 1.07)
Male	1.10 (0.66; 1.84)		1.25 (0.71; 2.20)
High educated (>high school)	0.45 (0.27; 0.78)**		0.44 (0.25; 0.77)**
Disability (Katz-15 ≥1)			0.55 (0.32; 0.97)*
Frail (Frailty Phenotype ≥3)		<i>Results of Model</i>	1.67 (0.71; 3.91)
Comorbidity, per comorbidity		<i>2 are presented in</i>	1.03 (0.91; 1.17)
Depressive symptoms (GDS-2 ≥1)		<i>the original paper</i>	0.60 (0.34; 1.07)
Hospital readmission		<i>(Table 3)</i>	0.76 (0.40; 1.43)
Health locus of control, per point			
Internal			1.03 (0.98; 1.08)
Others			1.00 (0.95; 1.06)
Change			0.99 (0.94; 1.04)

Reference group is self-perceived and measured change in ADL are congruent.

* significant at p:0.05 level; ** significant at p:0.01 level.

^a Model Chi²(6): 11.21, p: 0.082. Pseudo R²: 0.02 (McFadden). -2LL: 338.610.

^b Model Chi²(22): 49.23, p< 0.01. Pseudo R²: 0.07 (McFadden). -2LL: 688.065.

Table S4. Results of the multinomial regression analyses twelve months after discharge: odds ratio's and 95% confidence intervals of incongruent outcomes between self-perceived and measured change in ADL relative to congruent outcomes.

	Model 1 ^a n = 333	Model 2 n = 333	Model 3 ^b n = 333
Group: self-perceived is worse than measured change in ADL (P<M)			
Age, per year	1.02 (0.97; 1.07)		0.99 (0.93; 1.04)
Male	0.93 (0.53; 1.62)		0.94 (0.50; 1.77)
High educated (>high school)	1.08 (0.62; 1.88)		1.14 (0.62; 2.08)
Disability (Katz-15 ≥1)			4.67 (2.40; 9.09)**
Frail (Frailty Phenotype ≥3)		<i>Results of Model</i>	1.14 (0.48; 2.73)
Comorbidity, per comorbidity		<i>2 are presented in</i>	1.08 (0.94; 1.25)
Depressive symptoms (GDS-2 ≥1)		<i>the original paper</i>	1.05 (0.55; 2.01)
Hospital readmission		<i>(Table 3)</i>	0.93 (0.49; 1.78)
Health locus of control, per point			
Internal			0.94 (0.88; 0.99)*
Others			0.99 (0.93; 1.05)
Change			1.01 (0.96; 1.08)
Group: self-perceived is better than measured change in ADL (P>M)			
Age, per year	1.02 (0.97; 1.07)		1.04 (0.99; 1.10)
Male	0.12 (0.65; 1.94)		1.29 (0.71; 2.34)
High educated (>high school)	1.10 (0.63; 1.90)		1.02 (0.58; 1.81)
Disability (Katz-15 ≥1)			0.50 (0.26; 0.94)*
Frail (Frailty Phenotype ≥3)		<i>Results of Model</i>	1.24 (0.48; 3.23)
Comorbidity, per comorbidity		<i>2 are presented in</i>	1.10 (0.97; 1.26)
Depressive symptoms (GDS-2 ≥1)		<i>the original paper</i>	1.08 (0.58; 1.99)
Hospital readmission		<i>(Table 3)</i>	0.80 (0.41; 1.55)
Health locus of control, per point			
Internal			1.02 (0.97; 1.08)
Others			1.02 (0.96; 1.08)
Change			1.01 (0.95; 1.07)

Reference group is self-perceived and measured change in ADL are congruent.

* significant at p:0.05 level; ** significant at p:0.01 level.

^a Model Chi²(6): 1.37, p: 0.968. Pseudo R²: 0.02 (McFadden). -2LL: 343.828.

^b Model Chi²(22): 61.89, p <0.01. Pseudo R²: 0.09 (McFadden). -2LL: 647.148.

Table S5. Results subgroup analysis including patients without disabilities at baseline (Katz ADL Index score =0): Contingency tables of the agreement between change in self-perceived and measured ADL three (a) and twelve (b) months after discharge.

a. Three months after discharge ^s		Self-perceived change in ADL		
Measured change in ADL (Katz-15)		Deteriorated	Unchanged + Improved	Total
	Deteriorated	14	32	46
	Unchanged	19	98	117
Total		33	130	163

b. Twelve months after discharge [#]		Perceived change in ADL		
Measured change in ADL (Katz-15)		Deteriorated	Unchanged + Improved	Total
	Deteriorated	13	28	41
	Unchanged	21	100	121
Total		34	128	162

^s Absolute agreement after 3 months: 69 %. Kappa: 0.16 (95%CI: 0.00; 0.32).
[#] Absolute agreement after 12 months: 70 %. Kappa: 0.15 (95%CI: 0.01; 0.29).

Table S6. Results of the multinomial regression subgroup analysis including patients without disabilities at baseline (Katz ADL Index score =0): odds ratio's and 95% confidence intervals of incongruent outcomes between self-perceived and measured change in ADL relative to congruent outcomes.

	3 months AD ^a n = 163	12 months AD ^b n = 162
Group: self-perceived change in ADL is worse than measured change in ADL (P<M)		
Age, per year	1.03 (0.913; 1.18)	0.95 (0.83; 1.08)
Male	1.19 (0.38; 3.78)	1.13 (0.38; 3.35)
High educated (>high school)	0.66 (0.22; 2.02)	1.81 (0.66; 4.92)
Disability (Katz-15 ≥1)	-	-
Frail (Frailty Phenotype ≥3)	3.70 (0.19; 73.97)	n.a. ^c
Comorbidity, per comorbidity	1.16 (0.93; 1.46)	0.84 (0.63; 1.12)
Depressive symptoms (GDS-2 ≥1)	0.14 (0.01; 1.62)	0.91 (0.27; 3.11)
Hospital readmission	2.76 (0.91; 8.34)	0.46 (0.40; 5.30)
Group: self-perceived change in ADL is better than measured change in ADL (P>M)		
Age, per year	1.12 (1.02; 1.23)*	1.14 (1.03; 1.25)*
Male	0.62 (0.22; 1.74)	1.32 (0.47; 3.71)
High educated (>high school)	0.39 (0.15; 0.99)*	1.32 (0.54; 3.21)
Disability (Katz-15 ≥1)	-	-
Frail (Frailty Phenotype ≥3)	1.10 (0.10; 12.13)	n.a. ^c
Comorbidity, per comorbidity	1.17 (0.97; 1.40)	1.03 (0.84; 1.36)
Depressive symptoms (GDS-2 ≥1)	0.67 (0.23; 1.95)	0.63 (0.19; 2.05)
Hospital readmission	1.53 (0.55; 4.24)	1.22 (0.37; 4.01)

Reference group is self-perceived and measured change in ADL are congruent.
Abbreviations: AD, After Discharge; ADL, Activities of Daily Living; GDS, Geriatric Depression Scale.
* significant at p:0.05 level; ** significant at p:0.01 level.
^a Model Chi²(16): 24.07, p<0.05. Pseudo R²: 0.09 (McFadden). -2LL: 221.333.
^b Model Chi²(14): 14.50, p:0.41. Pseudo R²: 0.06 (McFadden). -2LL: 230.834.
^c Not available due to too less frail cases

Table S7. Results subgroup analysis including patient with ≥ 1 disability at baseline: Contingency tables presenting the frequency distribution (n) of the agreement between change in self-perceived and measured ADL three (a) and twelve (b) months after discharge.

a. Three months after discharge [§]		Self-perceived change in ADL			
		Deteriorated	Unchanged	Improved	Total
Measured	Deteriorated	30	26	10	66
change in ADL	Unchanged	17	30	13	60
(Katz-15)	Improved	16	24	26	66
Total		63	80	49	192

b. Twelve months after discharge [#]		Perceived change in ADL			
		Deteriorated	Unchanged	Improved	Total
Measured	Deteriorated	32	22	6	60
change in ADL	Unchanged	20	18	5	43
(Katz-15)	Improved	19	32	17	68
Total		71	72	28	171

[§]Absolute agreement after three months: 45%. Linear weighted kappa: 0.21 (95%CI: 0.10;0.32).
[#] Absolute agreement after twelve months:39%. Linear weighted kappa: 0.17 (95%CI: 0.06; 0.28).

CHAPTER 3

Reproducibility and Responsiveness of
the Frailty Index and Frailty Phenotype
in Older Hospitalized Patients

Marlies Feenstra
Frederike M M Oud
Carolien J Jansen
Nynke Smidt
Barbara C van Munster
Sophia E de Rooij

ABSTRACT

Background: There is growing interest for interventions aiming at preventing frailty progression or even to reverse frailty in older people, yet it is still unclear which frailty instrument is most appropriate for measuring change scores over time to determine the effectiveness of interventions. The aim of this prospective cohort study was to determine reproducibility and responsiveness properties of the Frailty Index (FI) and Frailty Phenotype (FP) in acutely hospitalized medical patients aged 70 years and older.

Methods: Reproducibility was assessed by Intra-Class Correlation Coefficients (ICC), standard error of measurement (SEM) and smallest detectable change (SDC); Responsiveness was assessed by the standardized response mean (SRM), and area under the receiver operating characteristic curve (AUC).

Results: 243 patients were included with a median age of 76 years (range 70–98). ICC of the FI were 0.85 (95% confidence interval [CI]: 0.76; 0.91) and 0.88 (95%CI: 0.82; 0.92), and 0.65 (95%CI: 0.49; 0.77) and 0.76 (95%CI: 0.65; 0.84) for the FP. SEM ranged from 5% to 15%; SDC from 13% to 37%. SRMs were good in patients with unchanged frailty status (<0.50), and doubtful to good for deteriorated and improved patients (0.42–1.04). AUC's over three months were 0.78 (95%CI: 0.70; 0.86) and 0.70 (95%CI: 0.62; 0.79) for the FI, and 0.67 (95%CI: 0.58; 0.77) and 0.64 (95%CI: 0.54; 0.73) for the FP. Over twelve months, AUCs were 0.76 (95%CI: 0.68; 0.84) and 0.82 (95%CI: 0.74; 0.90) for the FI, and 0.77 (95%CI: 0.69; 0.86) and 0.76 (95%CI: 0.67; 0.84) for the FP.

Conclusion: The Frailty Index showed good reproducibility and was able to discriminate deteriorated from stable patients three and twelve months after hospitalization. The Frailty Phenotype showed poor to moderate reproducibility and was able to discriminate deteriorated from stable patients at twelve months after hospitalization only.

BACKGROUND

Frailty is a medical condition of increased vulnerability due to a reduced ability to maintain homeostasis after a stressor event as a consequence of cumulative decline in multiple physiological systems during a lifetime.¹ Around 40% of the hospitalized older patients are frail which is associated with poor health outcomes, such as functional decline, hospital re-admission, institutionalization, and mortality.^{2,3}

Identifying (pre) frail older adults, and those at risk for progression of frailty is important. Some older adults may benefit from interventions targeted at prevention of frailty progression to lower the risk of poor health outcomes like functional decline.^{4,5} Reliable and valid assessment of frailty and how to measure relevant changes in frailty over time is therefore crucial.

Several frailty instruments exist for the purpose of diagnosing, risk stratification, and evaluating frailty over time.⁶ Comprehensive geriatric assessment is currently the gold standard for diagnosing the frailty status in clinical practice¹, but the cumulative deficits model or Frailty Index (FI) and the Frailty Phenotype (FP) are the most widely used instruments used to establish frailty status in research.^{7,8} Construct validity and predictive validity of negative health outcomes of the FI and FP have been extensively evaluated and are proven to be satisfactory in both community-dwelling and hospitalized older adults.^{9–11} Reproducibility and responsiveness of change scores of frailty instruments are poorly studied especially after hospitalization and it is still unclear which frailty instrument is most appropriate for measuring change scores over time or the effectiveness of interventions.^{9,11,12}

Therefore, the aim of this study is to determine the reproducibility and responsiveness of the FI and FP in acutely admitted hospitalized older medical patients.

METHODS

Study population

During this monocentric prospective cohort study, patients aged ≥ 70 years were recruited between February 2017 and May 2018. Twelve months' follow-up measurements continued to the end of April 2019. During weekdays patients admitted to surgical, cardiology, pulmonology, medical oncology, nephrology, and general internal medicine wards were checked for eligibility to participate. Inclusion criteria were age ≥ 70 years and an expected hospital stay of at least two days. Exclusion criteria were no understanding of the Dutch language, any (temporary) cognitive condition that influenced decision making capacity, and no written informed consent. The Research Ethics Committee of the University Medical Center Groningen ruled that no formal ethics approval was required (file number: 201600268). All participants provided written informed consent before participation.

Data collection

Baseline assessment took place within four days after admission. Telephonic assessments were performed at three and twelve months' post-discharge, in which baseline questions were repeated and two anchor questions were added.¹³ Data were collected by trained research staff.

Questionnaires

Frailty instruments

The FI score was calculated using 34 deficits associated with health status.¹⁴ The FP was assessed by five self-reporting criteria including strength, walking ability, weight loss, physical activity, and exhaustion.^{15,16} At baseline assessment, patients were asked to retrospectively refer to the situation two weeks before hospital admission for all items, except for the cognitive domain of the FI. A detailed description of the included items of the FI and FP are presented in the Supplementary files, Tables S1 and S2.

Patient-reported anchor questions

Anchor questions were used as an external criterion for measuring responsiveness.¹⁷ Two different anchor questions were used: 1. 'In general, how is your health state now, compared to three months/twelve

months ago before hospitalization?' 2. 'In general how is your daily functioning now, compared to three months/twelve months ago before hospitalization?' Response options were a five point Likert scale.

Sociodemographic and patient characteristics

For all subjects, baseline sociodemographic characteristics were collected including age, sex, living situation (independent living vs not independent living), and educational level ($\leq$ highschool vs $>$ highschool). After discharge, medical charts were consulted to assess baseline comorbidity (Charlson Comorbidity Index)¹⁸ and mortality during follow-up time.

Statistical analysis

Descriptive statistics

For all baseline sociodemographic and patient characteristics, descriptive statistics were calculated. The distribution of the scores on frailty instruments at baseline were inspected for possible floor and ceiling effects. Thresholds for floor and ceiling effects were if $\geq 15\%$ of the patients achieved the lowest or highest possible score, respectively.¹⁹

Reproducibility

Test-retest reproducibility was assessed among patients who reported to be unchanged according to the anchor question three months post discharge. The following parameters were calculated:

1. The intraclass correlation coefficient (ICC) using a two-way mixed effects model for absolute agreement was used for the baseline and three months follow-up measurements of the FI and the FP. Cut-off values for interpretation of the ICC including the 95% confidence interval were <0.5 poor, ≥ 0.5 and <0.75 moderate, ≥ 0.75 and <0.9 good, ≥ 0.9 excellent reliability.²⁰
2. Cohen's kappa and absolute agreement were calculated to assess the reproducibility of the FP using the categorized outcome (robust, prefrail, frail). Cut-off values were <0.40 poor, ≥ 0.40 and <0.75 fair to good, ≥ 0.75 excellent agreement.¹⁹
3. Measurement error of the FI and FP was assessed by calculating Bland-Altman plots and the standard error of measurement (SEM). Bland-Altman plots were calculated by the mean change scores of baseline and three months post discharge assessments

plotted against the difference on both scores. SEM was calculated using the following formula: $SEM = SD(TO) \times \sqrt{(1-r)}$. SD is the standard deviation of the baseline measurement of the Frailty Index of the unchanged group; r refers to the ICC. To interpret the SEM, scores are converted to percentages of the scale range. Cut-off values were: $\leq 5\%$ very good, $\leq 10\%$ good, $>10\%$ and $<20\%$ doubtful, $\geq 20\%$ poor.²¹

4. To be able to interpret change scores, the smallest detectable change (SDC) was calculated for the FI and FP using continuous scores. SDC reflects the variance of the distribution of change scores among stable patients. Patients who reported no change according to the anchor question were assumed to be stable. SDC was calculated by the following formula: $SDC = 1.96 \times \sqrt{2} \times SEM$.²¹ Both the absolute SDC value as well as the SDC as a percentage of the scale range were calculated.

Responsiveness

Two types of responsiveness were determined over the timeframes from pre-hospital admission to three and twelve months' post discharge: Internal responsiveness, defined as the magnitude of change related to the variance in change scores, was determined by the standardized response mean (SRM). SRM is calculated by dividing the mean change score by the standard deviation of the mean change score.²² SRMs were separately calculated for improved, unchanged, and deteriorated patients according to the anchor question for both the FI and FP using continuous scores. Cut-off values were: ≤ 0.2 small, >0.2 and ≤ 0.5 doubtful, >0.5 and ≤ 0.8 good, >0.8 very good internal responsiveness for the improved and deteriorated patients; <0.50 good, ≥ 0.50 small for the unchanged patients.²²

External responsiveness, defined as the ability to detect change over time in the construct to be measured, was assessed by investigating the ability of the instruments to discriminate between relevant changes (improved and deteriorated patients) and irrelevant changes (unchanged patients).¹⁹ This is reflected by the area under the receiver operating characteristic curve (AUC). For these analyses, the anchor question was considered as the gold standard for change, and the change scores on the FI and FP using continuous scores were considered as the 'diagnostic test' for

Table 1 Baseline characteristics of total sample, completers, and non-completers.

Baseline characteristics	Total sample (n=243)	Completers (n=148)	Non-completers (n=95)	P value ^a
Age, median (IQR 25;75)	76 (72;81)	75 (72;80)	78 (73;84)	0.053
range (years)	70 – 98	70 – 96	70 – 98	
Sex, male	165 (68)	101 (68)	64 (67)	0.887
Housing situation				
independent	225 (93)	141 (95)	84 (88)	0.047
not independent	18 (8)	7 (5)	11 (11)	
Education				
$\leq$ high school	173 (71)	101 (68)	72 (76)	0.205
$>$ high school	70 (29)	47 (32)	23 (24)	
CCI, median (IQR 25;75)	2 (1; 4)	2 (1; 4)	2 (1; 4)	0.083
Frailty Index				
median (IQR 25;75)	0.18 (0.08; 0.31)	0.15 (0.07; 0.27)	0.26 (0.10; 0.40)	0.014
lowest possible score	0 (0)	0 (0)	0 (0)	
highest possible score	0 (0)	0 (0)	0 (0)	
Frailty Phenotype				
median (IQR 25;75)	1.00 (0; 2.00)	0.50 (0; 2.00)	2.00 (0; 3.00)	0.011
robust (0 criteria)	97 (40)	72 (49)	25 (26)	
prefrail (1 or 2 criteria)	88 (36)	49 (33)	39 (41)	0.004
frail (≥ 3 criteria)	47 (19)	23 (16)	24 (25)	
lowest possible score	66 (27)	72 (49)	25 (26)	
highest possible score	4 (2)	1 (1)	3 (3)	

Notes: Values are presented as numbers and percentage (%) unless indicated otherwise. Percentages may not equal 100% due to rounding and missing. CCI, Charlson Comorbidity Index; IQR, inter quartile range.

a. Differences between completers and non-completers were tested using Mann-Whitney U test in case of continuous variables, and Chi Square test in case of categorical variables.

measuring change. An AUC of ≥ 0.70 was considered to be adequate. In addition, for each instrument the optimal cut-off point was calculated for which the sensitivity and specificity together revealed the least error in classifying patients as improved versus unchanged and deteriorated versus unchanged. To carry out these analyses, the correlation (Spearman's rho) between the change score on the frailty instrument and the anchor question should be at least 0.40.²³

Missing data

Patients who died during follow-up were included in the analytic sample by imputing the highest prevalent frailty category for each frailty instrument and the worst outcome for the anchor questions. For all analyses IBM SPSS Statistics, version 23 was used.

RESULTS

Descriptive statistics

Baseline characteristics were presented in Table 1. Of the 243 participants included in the study, 51 had no follow-up data three months post discharge and 35 had no follow-up data twelve months post discharge (Supplementary files, Figure S1). In total, 39 patients (16%) died during the study. Patients who dropped out during the study (n = 95) were more often living with support or were institutionalized (11%) and had higher frailty scores at baseline (Table 1).

The FP showed a floor effect at baseline assessment, hampering detection of improvement in 49% of these patients (Table 1; Supplementary files, Figures S2 and S3). Mean scores of the frailty instruments at baseline and follow-up measurements for all and within collapsed categories as used in the analyses are presented in Table 2 and in the Supplementary files, Table S3.

Table 2. Mean (SD) frailty scores at baseline (T0) and 3 months post discharge (T1) .

	n	Frailty Index		Frailty Phenotype	
		T0	T1	T0	T1
		mean (SD)	mean (SD)	mean (SD)	mean (SD)
Health anchor					
All categories					
much better	24	0.19 (0.12)	0.13 (0.09)	1.38 (1.28)	0.35 (0.65)
slightly better	24	0.23 (0.19)	0.22 (0.20)	1.57 (1.27)	1.00 (1.18)
more or less the same	74	0.16 (0.12)	0.13 (0.11)	0.73 (1.12)	0.67 (0.98)
slightly worse	34	0.19 (0.15)	0.22 (0.16)	0.97 (1.36)	1.09 (1.36)
much worse	35	0.32 (0.17)	0.60 (0.25)	2.12 (1.41)	3.60 (1.79)
Collapsed categories ^a .					
improved	48	0.21 (0.16)	0.17 (0.16)	1.48 (1.27)	0.68 (1.00)
unchanged	74	0.16 (0.12)	0.13 (0.11)	0.73 (1.12)	0.67 (0.98)
deteriorated	69	0.26 (0.17)	0.41 (0.28)	1.55 (1.49)	2.38 (2.02)
Functioning anchor					
All categories					
much better	17	0.20 (0.14)	0.13 (0.09)	1.56 (1.26)	0.29 (0.47)
slightly better	21	0.19 (0.18)	0.15 (0.17)	1.29 (1.42)	0.70 (1.03)
more or less the same	82	0.17 (0.14)	0.15 (0.13)	0.78 (1.15)	0.71 (0.98)
slightly worse	40	0.21 (0.15)	0.22 (0.16)	1.21 (1.32)	1.08 (1.37)
much worse	34	0.32 (0.15)	0.61 (0.24)	2.06 (1.41)	3.82 (1.59)
Collapsed categories ^a .					
improved	38	0.20 (0.16)	0.14 (0.14)	1.41 (1.34)	0.51 (0.84)
unchanged	82	0.17 (0.14)	0.15 (0.13)	0.78 (1.15)	0.71 (0.98)
deteriorated	74	0.26 (0.16)	0.40 (0.28)	1.60 (1.42)	2.32 (2.01)

a. The collapsed categories were used in the analyses of the healthstate and functioning anchor questions.

Reproducibility

ICCs, SEMs, and SDCs were presented in Table 3. ICCs of the FI were 0.85 (95% confidence interval [CI]: 0.76; 0.91) and 0.88 (95% CI: 0.82; 0.92), and 0.65 (95% CI: 0.49; 0.77) and 0.76 (95% CI: 0.65; 0.84) for the FP. Kappa statistics of the FP categories were 0.37 (absolute agreement: 0.66) and 0.43 (absolute agreement: 0.70) indicating poor to fair agreement (Supplementary files, Tables S4 – S7). SEM of the FI were good to very good (5 and 6%) and doubtful for the FP (11 and 13%). Bland Altman plots are presented in the Additional File 1, Figures A4 and A5. Good agreement was observed for the FI using the functioning anchor and for the FP using the health and functioning anchors ($p>0.05$). A systematic mean difference was observed between baseline and three-months follow-up tests of the FI using the health anchor (mean difference 0.02; 95% CI: 0.01; 0.04).

Table 3. Reproducibility properties of unchanged patients at three months follow-up measurement.

Health anchor (n = 74)					
Instrument	ICC (95% CI)	SEM	SEM% ^a	SDC	SDC% ^a
FI	0.85 (0.76; 0.91)	0.05	5%	0.13	13%
FP	0.65 (0.49; 0.77)	0.66	13%	1.84	37%
Functioning anchor (n = 82)					
Instrument	ICC (95% CI)	SEM	SEM% ^a	SDC	SDC% ^a
FI	0.88 (0.82; 0.92)	0.06	6%	0.17	17%
FP	0.76 (0.65; 0.84)	0.57	11%	1.58	32%

Notes: Intraclass correlation coefficient for agreement using a 2 way mixed effect model. ICC, Intraclass Correlation Coefficient; FI, Frailty Index; FP, Frailty Phenotype; SDC, smallest detectable change; SEM, standard error of measurement.

a. SEM% and SDC% are SEM and SDC expressed in percentages of the continuous score of the instrument.

Responsiveness

All SRMs in the improved and deteriorated groups were higher than the SRMs in the stable groups, meaning that the measured change in frailty outcomes was lower among patients in the stable groups compared to

patients in both the improved and deteriorated groups (Table 4). Largest SRMs were found for deterioration over twelve months (Supplementary files, Table S8), with SRMs ranging from 0.77 to 1.04, indicating good to very good internal responsiveness.

Only the FI showed sufficient responsiveness to detect deterioration in frailty over three months (Figure 1). Highest AUC of the FI was found

Table 4. Mean change scores and internal responsiveness for improved, unchanged, and deteriorated patients.

	n	Mean change score (SD)	Spearman Correlation	SRM
Health anchor				
Frailty Index				
improved	48	-0.05 (0.11)	0.49	-0.45
unchanged	74	-0.02 (0.06)		-0.33
deteriorated	69	0.15 (0.23)		0.65
Frailty Phenotype				
improved	48	-0.86 (1.21)	0.46	-0.71
unchanged	74	-0.07 (0.88)		-0.08
deteriorated	69	0.80 (1.68)		0.48
Functioning anchor				
Frailty Index				
improved	38	-0.06 (0.11)	0.43	-0.55
unchanged	82	-0.01 (0.06)		-0.16
deteriorated	74	0.13 (0.23)		0.57
Frailty Phenotype				
improved	38	-0.91 (1.30)	0.45	-0.70
unchanged	82	-0.09 (0.73)		-0.12
deteriorated	74	0.71 (1.75)		0.41

Notes: Spearman correlation coefficients were calculated using mean change scores of the total group and the anchor question. SD, standard deviation; SRM, standardized response mean.

for the health anchor question (AUC: 0.78; 95% CI: 0.70; 0.86, see Figure 1b). The optimal cut-off of the change scores from baseline to three months post discharge was 0.02 with corresponding sensitivity and specificity of 69% and 84% (Supplementary files, Table S9). Over the timeframe of twelve month, both the FI and FP were responsive for deterioration in frailty status. Highest AUC's were found for the health anchor (AUC: 0.82; 95% CI: 0.74; 0.90 and AUC: 0.77; 95% CI: 0.69; 0.86) with corresponding optimal change score of 0.04 (Sensitivity: 75%; Specificity: 77%) and 0.5 (Sensitivity: 67%; Specificity: 84%) for the FI and FP, respectively (Supplementary files, Table S9). ROC curves at twelve months follow-up measurement are presented in the Supplementary files, Figure S6.

DISCUSSION

In this study on acutely hospitalized older medical patients, we found that the FI showed good reproducibility and the FP showed moderate reproducibility. In addition, the FI was responsive in detecting deterioration in frailty between pre-acute hospital admission and three and twelve months post discharge. The FP was only responsive in detecting patients with deteriorating frailty status between pre-hospital admission and twelve months post-discharge.

Like cholesterol is a measurable marker for cardiovascular disease risk and HbA1c for risk of complications of diabetes type 2, frailty can be considered as the measurable marker for risk of adverse outcomes. Responsive frailty measures can therefore act as an important intermediate outcome for the evaluation of interventions aiming at preventing adverse health outcomes in older patients. Despite that the FP is frequently used as an outcome instrument to evaluate interventions we failed to establish satisfactory reproducibility and responsiveness of the FP over a three months timeframe.^{24–26} As previously suggested²⁷, we found that the FI was more sensitive to detecting deterioration in frailty status than the FP, making it a more promising tool to evaluating frailty status over time. A change score of five items over three months and six items over twelve months corresponds with the smallest detectable change of the FI.

Figure 1. Receiver Operating Characteristic (ROC) curve comparisons and corresponding Area Under the ROC curves (AUC) between the Frailty Index (solid line) and the Frailty Phenotype (dashed line) for measuring improvement (a and c) and deterioration (b and d) in frailty status according to the health and functioning anchor questions. The diagonal lines (gray) represent the reference lines of no-discrimination.

A first explanation as to why the FI performs better than the FP could be the presence of a floor effect of the FP. At baseline, 40% scored the lowest possible FP score, preventing participants from further improvement in frailty. A second explanation could be the scale width. A broader scale range improves the ability to differentiate between changed and unchanged patients.²⁸ Lastly, the differences could be explained by the

theoretical models underlying the FI and FP. Items that are included in the FI cover multiple domains associated with health status.¹⁴ The FP included only items related to the physical domain, complicating the detection of changes in other domains.

To our knowledge, this is one of the first studies that investigated the reproducibility and responsiveness of the FI and FP in acutely hospitalized older patients. Strengths of this study are the use of various methods for assessing reproducibility and responsiveness and the large sample size. More than twice the recommended number of patients for evaluating psychometric properties were included.¹⁹ There are, however, limitations. First, instead of using a comprehensive geriatric assessment, the gold standard for frailty status, we used two relevant patient-based anchor questions to determine change in frailty. By using two anchor questions referring to change in function and health in a population that is prone to frailty, we intended to come close to evaluating real frailty status. The moderate positive correlations of these anchor questions with the change scores on the frailty instruments indicate that the chosen anchor questions may not fully capture the whole concept of frailty. Directly asking for change in frailty was, however, not an option because older adults themselves are not familiar with this concept.²⁹ Second, we cannot rule out that the results are biased by response shift, as response shift evaluation was not incorporated in the study design.³⁰ Third, the modified FP was used instead of the original performance-based measures because in our population of older medical patients, performance-based assessment is often too challenging and would have led to the inclusion of the fittest frail patients, resulting in an undesirable selection bias. Fourth, due to ethical considerations, no patients were included with cognitive impairment due to dementia or delirium, although frailty in these patients is often present and their frailty status is expected to decline after hospitalization. However, the broad inclusion criteria still resulted in the inclusion of a representative sample of the geriatric population consisting of a heterogeneous group of mentally competent older medical patients. Future research should compare the responsiveness of existing frailty instruments and their relation to the course of functional impairment in multiple studies and other patient groups. These studies should also incorporate measures such as a then-test in study designs to identify and adjust for response shift.

CONCLUSION

In this study, the Frailty Index showed better reproducibility and responsiveness properties compared to the Frailty Phenotype. Although the Frailty Index looks promising, based on this single study we cannot yet formulate concrete recommendations about the best instrument to evaluating frailty status over time.

REFERENCES

1. Clegg A, Young J, Iliffe S, Olde Rikkert MGM, Rockwood K. Frailty in older people summary. *Lancet*. 2013;381(9868):752-762. doi:10.1016/S0140-6736(12)62167-9.Frailty
2. Joosten E, Demuynck M, Detroyer E, Milisen K. Prevalence of frailty and its ability to predict in hospital delirium , falls , and 6-month mortality in hospitalized older patients. *BMC Geriatr*. 2014;14(1):1-9. doi:10.1186/1471-2318-14-1
3. Vu HT, Nguyen TX, Nguyen TN, et al. Prevalence of frailty and its associated factors in older hospitalised patients in. *BMC Geriatr*. 2017;17(216):1-7. doi:10.1186/s12877-017-0609-y
4. Apóstolo J, Cooke R, Bobrowicz-campos E, et al. Effectiveness of interventions to prevent pre-frailty and frailty progression in older adults : a systematic review. *JBISIR*. 2018;16(1):140-232. doi:10.11124/JBISIR-2017-003382
5. Dedeyne L, Deschodt M, Verschueren S, Tournoy J, Gielen E. Effects of multi-domain interventions in (pre)frail elderly on frailty, functional, and cognitive status: A systematic review. *Clin Interv Aging*. 2017;12:873-896. doi:10.2147/CIA.S130794
6. Rockwood K, Theou O, Mitnitski A. What are frailty instruments for? *Age Ageing*. 2015;44(4):545-547. doi:10.1093/ageing/afv043
7. Buta BJ, Walston JD, Godino JG, et al. Frailty assessment instruments: Systematic characterization of the uses and contexts of highly-cited instruments. *Ageing Res Rev*. 2017;March(26):53-61. doi:10.1016/j.arr.2015.12.003.Frailty
8. Morley JE, Vellas B, Kan GA Van, et al. Frailty Consensus : A Call to Action. *J Am Med Dir Assoc*. 2013;14(6):392-397. doi:10.1016/j.jamda.2013.03.022
9. Drubbel I, Numans ME, Kranenburg G, Bleijenberg N, Wit NJ De. Screening for frailty in primary care : a systematic review of the psychometric properties of the frailty index in community-dwelling older people. *BMC Geriatr*. 2014;14(1):1-13. doi:10.1186/1471-2318-14-27
10. Sutton JL, Gould RL, Daley S, et al. Psychometric properties of multicomponent tools designed to assess frailty in older adults: A systematic review. *BMC Geriatr*. 2016;16(1). doi:10.1186/s12877-016-0225-2
11. De Vries NM, Staal JB, Van Ravensberg CD, Hobbelen JSM, Olde Rikkert MGM, Nijhuis-van der Sanden MWG. Outcome instruments to measure frailty: a systematic review. *Ageing Res Rev*. 2011;10:104-114. doi:10.1016/j.arr.2010.09.001
12. Apostolo J, Cooke R, Bobowicz-Campos E, et al. Predicting risk and outcomes for frail older adults: an umbrella review of frailty screening tools ´. *JBISIR*. 2017;15(4):1154-1208. doi:10.11124/JBISIR-2016-003018
13. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: A review of strengths and weaknesses and considerations for design. *J Man Manip Ther*. 2009;17(3):163-170. doi:10.1179/jmt.2009.17.3.163
14. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8(1):24. doi:10.1186/1471-2318-8-24
15. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: Evidence for a phenotype. *Journals Gerontol Ser a-Biological Sci Med Sci*. 2001;56(3):M146-M156. doi:10.1093/gerona/56.3.M146
16. Theou O, Rockwood K. Comparison and Clinical Applications of the Frailty Phenotype and Frailty Index Approaches. *Interdiscip Top Gerontol Geriatr*. 2015;41:74-84. doi:10.1159/000381166
17. De Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care*. 2001;17(4):479-487.
18. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8
19. De Vet, Henrica C. W. Terwee CB, Knol DL, Mokkink LB. Measurement in Medicine. first. Cambridge: Cambridge University Press; 2011.
20. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

21. Ostelo RWJG, De Vet HCW, Knol DL, Van Den Brandt PA. 24-Item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *J Clin Epidemiol*. 2004;57(3):268-276. doi:10.1016/j.jclinepi.2003.09.005
22. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press; 1977.
23. Akoglu H. User's guide to correlation coefficients. *Turkish J Emerg Med*. 2018;18(3):91-93. doi:10.1016/j.tjem.2018.08.001
24. Hars M, Herrmann FR, Fielding RA, Reid KF, Rizzoli R, Trombetti A. Long-Term Exercise in Older Adults: 4-Year Outcomes of Music-Based Multitask Training. *Calcif Tissue Int*. 2014;95(5):393-404. doi:10.1007/s00223-014-9907-y
25. Tarazona-Santabalbina FJ, Gómez-Cabrera MC, Pérez-Ros P, et al. A Multicomponent Exercise Intervention that Reverses Frailty and Improves Cognition, Emotion, and Social Networking in the Community-Dwelling Frail Elderly: A Randomized Clinical Trial. *J Am Med Dir Assoc*. 2016;17(5):426-433. doi:10.1016/j.jamda.2016.01.019
26. Ng TP, Feng L, Nyunt MSZ, et al. Nutritional, Physical, Cognitive, and Combination Interventions and Frailty Reversal among Older Adults: A Randomized Controlled Trial. *Am J Med*. 2015;128(11):1225-1236.e1. doi:10.1016/j.amjmed.2015.06.017
27. Cesari M, Gambassi GIG, KAn GA van, Vellas B. The frailty phenotype and the frailty index : different instruments for different purposes. *Age Ageing*. 2014;43(October 2013):10-12. doi:10.1093/ageing/aft160
28. Sanders JL, Boudreau RM, Fried LP, Walston JD, Harris TB, Newman AB. Measurement of organ structure and function enhances understanding of the physiological basis of frailty: The Cardiovascular Health Study. *J Am Geriatr Soc*. 2011;59(9):1581-1588. doi:10.1111/j.1532-5415.2011.03557.x
29. Durepos P, Sciences HH, Alsbury K, Hewston P, Borges J. Older Adults ' Perceptions of Frailty Language : A Scoping Review. *ResearchGate*. 2020;(January). doi:10.13140/RG.2.2.27686.06722
30. Barclay-Goddard R, Epstein JD, Mayo NE. Response shift: A brief overview and proposed research priorities. *Qual Life Res*. 2009;18(3):335-346. doi:10.1007/s11136-009-9450-x

SUPPLEMENTARY FILES – CHAPTER 3

Table S1. List of 34 items used for the Frailty Index.

Deficits were rescaled into values between 0 and 1, representing the presence of the deficits measured. The total score represents present deficits as a proportion of all deficits measured, resulting in a score ranging from 0 to 1.

Variable	Scoring
ADL / iADL domain	
1. Help bathing	Yes = 1
2. Help dressing	No = 0
3. Help grooming	
4. Help using toilet	
5. Difficulty getting out chair	
6. Help getting out chair	
7. Help walking	
8. Help eating	
9. Help using phone	
10. Help transport	
11. Help shopping	
12. Help meal preparations	
13. Help housework	
14. Help taking medications	
15. Help finances	
Geriatric syndrome	
16. Urinary incontinence	Yes = 1
17. Visual impairment	No = 0
18. Hearing impairment	
19. Fall in past month	
20. Unintentional weight loss	
Psychosocial domain	
21. Feel everything is an effort	

22. Difficulty get going	Yes = 1 No = 0
23. Self-rated health	Poor = 1 Fair = 0.75 Good = 0.5 Very Good = 0.25 Excellent = 0
24. Did you feel downhearted and blue?	Both yes = 1
25. Have you dropped many of your interests and activities?	One yes = 0.5 Both no = 0

Symptoms	
26. Pain / discomfort	No pain / discomfort = 0 Moderate = 0.5 Extreme = 1
27. Mood	Not anxious / depressed = 0 Moderately = 0.5 Extreme = 1
28. Loss of appetite ^c	Yes = 1 No = 0

Physical domain	
29. Housing situation	Independent / conventional = 0 Transitional = 0.5 Institutionalized = 1
30. Living situation	Living alone = 1 Cohabiting/married = 0
31. Walk outside for 5 minutes	Impossible = 1 Only with help of somebody else = 0.75 Much effort = 0.5 Some effort = 0.25 No effort = 0
32. Physical exercise for at least 30 minutes	Never = 1 Monthly = 0.75 Weekly = 0.5 3 days/week = 0.25 Daily = 0

33. Move as intended	Yes = 0 No = 1
Cognitive domain	
34. MBT [†]	Completes test without error <60 sec. = 0 Completes test with error or >60s = 0.33 Engages, but cannot complete test = 0.66 Cannot or will not engage = 1

Notes: All baseline assessments referred to the situation two weeks before hospital admission unless indicated otherwise. (i)ADL, (instrumental) Activities of Daily Living; MBT, Months Backwards Test.
* Baseline assessment referred to a month before hospital admission.
[†] Cognitive functioning was assessed within the first four days of hospital admission.

Table S2. List of the criteria and items used for the Frailty Phenotype.

The sum score of positive criteria were calculated using the international agreed cut-offs: zero criteria indicated robust, one or two criteria indicated prefrail, 3 or more criteria indicated frail phenotype. For some calculations, the FP was considered as a continuous scale (range 0 to 5).

Criteria	Variable	Scoring
Weight loss	1. Report of weight loss of >3 kg in past month	Yes = 1 No = 0
Exhaustion	2. a. Everything I did was an effort b. I could not get going	Yes = 1 No = 0 Positive criterium if both are yes
Slowness	3. Not able to walk outside	Yes = 1 No = 0
Weakness	4. Having difficulties rising up a chair	Yes = 1 No = 0
Low physical activity	5. Being physically active for less than once a week	Yes = 1 No = 0

Notes: Baseline assessment referred to the situation two weeks before hospital admission unless indicated otherwise. Cut-offs used: frailty ≥3 positive criteria; prefrail 1 or 2 positive criteria; robust 0 positive criteria.

Figure S1. Flowchart and reasons for missing data of subjects.

Figure S2. Distribution and frequency of scores of the Frailty Index at baseline (n=228).

Figure S3. Distribution and frequency of scores of the Frailty Phenotype at baseline (n=232).

Table S3. Mean scores (SD) of the Frailty Index and Frailty Phenotype at baseline (T0) and twelve months post discharge (T2) for all and collapsed categories used in the analyses of the healthstate and functioning anchor questions.

		Frailty Index		Frailty Phenotype	
		T0 mean (SD)	T2 mean (SD)	T0 mean (SD)	T2 mean (SD)
Health anchor					
All categories					
much better	24	0.19 (0.12)	0.08 (0.07)	1.38 (1.28)	0.25 (0.68)
slightly better	19	0.23 (0.19)	0.14 (0.13)	1.57 (1.27)	0.24 (0.56)
more or less the same	74	0.16 (0.12)	0.15 (0.12)	0.73 (1.12)	0.56 (0.91)
slightly worse	21	0.19 (0.15)	0.24 (0.14)	0.97 (1.36)	1.32 (1.34)
much worse	50	0.32 (0.17)	0.73 (0.19)	2.12 (1.41)	4.44 (1.40)
Collapsed categories					
improved	43	0.21 (0.16)	0.10 (0.10)	1.48 (1.27)	0.24 (0.62)
unchanged	74	0.16 (0.12)	0.15 (0.12)	0.73 (1.12)	0.56 (0.91)
deteriorated	71	0.26 (0.17)	0.58 (0.28)	1.55 (1.49)	3.55 (1.97)
Functioning anchor					
All categories					
much better	15	0.20 (0.14)	1.56 (1.26)	1.56 (1.26)	0.07 (0.26)
slightly better	11	0.19 (0.18)	1.29 (1.42)	1.29 (1.42)	0.40 (0.70)
more or less the same	82	0.17 (0.14)	0.78 (1.15)	0.78 (1.15)	0.52 (0.92)
slightly worse	25	0.21 (0.15)	1.21 (1.32)	1.21 (1.32)	0.76 (1.10)
much worse	57	0.32 (0.15)	2.06 (1.41)	2.06 (1.41)	4.13 (1.58)
Collapsed categories					
improved	26	0.20 (0.16)	0.10 (0.12)	1.41 (1.34)	0.20 (0.50)
unchanged	82	0.17 (0.14)	0.14 (0.12)	0.78 (1.15)	0.52 (0.92)
deteriorated	82	0.26 (0.16)	0.53 (0.31)	1.60 (1.42)	3.20 (2.10)

Figure S4. Bland Altman plots of the Frailty Index (a) and Frailty Phenotype (b) for the health anchor.

Figure S5. Bland Altman plots of the Frailty Index (a) and Frailty Phenotype (b) for the functioning anchor.

Table S4. Cross table using the categorical scoring of the Frailty Phenotype and the health anchor question.

			Baseline			
Frailty phenotype - Health anchor			0	2	3	Total
FU 3 months	0	Count	32	9	1	42
		Expected Count	24.8	13.0	4.1	
	2	Count	10	12	3	25
		Expected Count	14.8	7.7	2.5	
	3	Count	0	1	3	4
		Expected Count	2.4	1.2	0.4	
Total		Count	42	22	7	71

Notes: 0 is robust, 2 is prefrail, 3 is frail phenotype. FU, follow-up measurement.

Table S5. Kappa statistic for the categorical scoring of the Frailty Phenotype in unchanged patients according to the health anchor.

	Value	Standard Error ^a	p-value
Kappa	0.369	0.102	<.001

Table S6. Cross tables using the categorical scoring of the Frailty Phenotype and the functioning anchor question.

			Baseline			
Frailty Phenotype - Functioning anchor			0	2	3	Total
FU 3 months	0	Count	35	10	0	45
		Expected Count	26.6	14.2	4.1	
	2	Count	10	14	3	27
		Expected Count	16.0	8.5	2.5	
	3	Count	0	0	4	4
		Expected Count	2.4	1.3	0.4	
Total		Count	45	24	7	76

Notes: 0 is robust, 2 is prefrail, 3 is frail phenotype. FU, follow-up measurement.

Table S7. Kappa statistic for the categorical scoring of the Frailty Phenotype in unchanged patients according to the functioning anchor.

	Value	Standard Error ^a	p-value
Kappa	.432	.099	<.001

Table S8. Mean change scores and internal responsiveness for improved, unchanged, and deteriorated patients between baseline and twelve months follow-up measurement.

	n	Mean change score (SD)	Spearman Correlation	SRM
Health anchor				
Frailty Index				
improved	43	-0.07 (0.12)	0.60	-0.58
unchanged	74	-0.01 (0.08)		-0.13
deteriorated	71	0.27 (0.26)		1.04
Frailty Phenotype				
improved	43	-0.93 (1.38)	0.57	-0.67
unchanged	74	-0.18 (0.86)		-0.21
deteriorated	71	1.62 (1.99)		0.81
Functioning anchor				
Frailty Index				
improved	26	-0.08 (0.13)	0.56	-0.62
unchanged	82	-0.01 (0.09)		-0.11
deteriorated	82	0.22 (0.27)		0.81
Frailty Phenotype				
improved	26	-1.17 (1.46)	0.58	-0.80
unchanged	82	-0.23 (1.01)		-0.23
deteriorated	82	1.39 (1.95)		0.71

Table S9. External responsiveness of the Frailty Index and the Frailty Phenotype over three and twelve months timeframes.

		3 months post discharge			12 months post discharge		
Health anchor		AUC (95%CI)	Cut off	Se	Sp	AUC (95%CI)	Sp
FI	improvement	0.57 (0.45;0.69)	-0.04	0.68	0.56	0.64 (0.53;0.76)	0.59
	deterioration	0.78 (0.70;0.86)	0.02	0.69	0.84	0.82 (0.74;0.90)	0.77
FP	improvement	0.68 (0.57;0.78)	-0.50	0.78	0.56	0.65 (0.54;0.76)	0.49
	deterioration	0.67 (0.58;0.77)	0.50	0.52	0.85	0.77 (0.69;0.86)	0.84
Functioning anchor							
FI	improvement	0.61 (0.49;0.74)	-0.03	0.62	0.57	0.67 (0.53;0.80)	0.61
	deterioration	0.70 (0.62;0.79)	0.02	0.60	0.78	0.76 (0.68;0.84)	0.73
FP	improvement	0.68 (0.56;0.80)	-0.5	0.781	0.57	0.68 (0.54;0.81)	0.57
	deterioration	0.64 (0.54;0.73)	0.5	0.471	0.85	0.76 (0.67;0.84)	0.86

Notes: Cut-offs are based on highest sensitivity and specificity. Numbers in italic are above the threshold for responsiveness (≥ 0.70). AUC, Area Under the Curve; FI, Frailty Index; FP, Frailty Phenotype; Se, sensitivity; Sp, specificity.

Figure S6. Receiver Operating Characteristic (ROC) curve comparisons between the Frailty Index and the Frailty Phenotype for measuring improvement (a and c) and deterioration (b and d) in frailty status according to the health and functioning anchor questions at twelve months follow-up measurement. The diagonal lines represent the reference lines of no discrimination.

CHAPTER 4

4.

Translation and validation of the Dutch
Pittsburgh Fatigability Scale for older
adults

Marlies Feenstra
Nynke Smidt
Barbara C van Munster
Nancy W Glynn
Sophia E de Rooij

BMC Geriatrics 2020; 20(1): 234–244

ABSTRACT

Background: The original Pittsburgh Fatigability Scale (PFS) was developed to assess perceived fatigability in older adults. The objective of this study was to translate the PFS into Dutch and investigate its validity and reliability among hospitalized older adults aged ≥ 70 years.

Methods: The PFS was translated into Dutch and pretested for comprehensibility by the Three-Step Test Interview method. The factor structure underlying the final version was evaluated by confirmatory factor analysis (CFA) and exploratory factor analyses (EFA). Internal consistency of the identified subscales was evaluated by Cronbach's alpha. Construct validity was evaluated by hypothesis testing. Test-retest reliability was evaluated using intraclass correlation coefficients (ICC) and Bland Altman plots.

Results: The validation sample included 233 patients. CFA of the original factor structure resulted in poor model fit in our Dutch sample. EFA of PFS physical and mental subscales resulted in a two-factor solution underlying the data with good internal consistency of the identified subscales (Cronbach's alpha: 0.80 – 0.92). Five out of six hypotheses were confirmed, indicating good construct validity. Retest assessments were performed among 50 patients and showed good reliability for both the physical (ICC: 0.80, 95%CI: 0.68; 0.88) and mental subscale (ICC: 0.81, 95%CI: 0.68; 0.89).

Conclusion: The Dutch PFS is a valid and reliable instrument to assess fatigability in older hospitalized patients.

BACKGROUND

Fatigue is a common burdening symptom among older adults that affects well-being.¹ Older adults often perceive fatigue without a clear physiological cause, which is suggested to be a consequence of aging in general.^{2,3} Prevalence rates of fatigue vary from 28% to 55% in adults aged 65 years and older and from 68% to 87% among the 85-year olds depending on the measurement tool.^{2,4} Fatigue in older adults is associated with persistent functional decline^{5,6} and is suggested to be a clinical marker for identifying persons at risk for negative health outcomes, such as frailty, disability, and hospitalizations.^{2,3,7}

There are many well-developed and validated uni- or multidimensional instruments to assess fatigue, such as the Brief Fatigue Inventory, Multidimensional Fatigue Symptom Inventory or the Fatigue Severity Scale.^{8,9} However, a key limitation of these tools is that they do not take the activity level of the respondent into account, potentially leading to information bias. For instance, sedentary people may perceive lower fatigue levels than their more active counterparts, making their outcomes hard to compare and interpret. Fatigability, defined as the measurement of fatigue in relation to an activity of fixed intensity and duration, accounts for this so called self-pacing bias.^{7,10} Thus, fatigability is a more sensitive measure of someone's susceptibility to fatigue, making this construct more suitable as a research outcome.^{7,11,12}

Fatigability can be assessed using performance-based or self-report tools.^{10,13} Performance fatigability is measured by the change in performance (e.g. slowing down during a physical performance test such as a 400 meter walk) without asking for the perception of fatigue.^{10,14} Self-reported or perceived fatigability can be measured in two ways: 1) self-reported fatigue after performing an activity, and 2) asking for the expected fatigue after performing activities of a certain intensity and duration, without actually performing the activity.^{7,14,15} The latter is especially suitable for populations in which performing activities is too challenging, such as hospitalized, bedridden older adults, or for large scale studies where equipment or space is not available.¹⁵ A few self-administered instruments are available that attempt to measure fatigability: Mobility Tiredness Scale¹⁶, Dutch Exertion Fatigue Scale¹⁷,

Situational Fatigue Scale¹⁸, and the Pittsburgh Fatigability Scale (PFS).¹⁵ Both the Mobility Tiredness Scale and the Dutch Exertion Fatigue Scale were developed for the assessment of fatigue during activities of daily living.^{16,17} The first was developed for older people, the latter for patients with heart failure. However, the items of both instruments do not consistently refer to the intensity and duration of every activity, and making them prone to self-pacing bias. The Situational Fatigue Scale included items referring to both physical and mental fatigability using items referring to activities with fixed intensity and duration, but was developed for and validated in younger age groups.¹⁸ The PFS was specifically developed for the assessment of both physical and mental fatigability in older people and consists of multiple items that describe activities of varying duration and intensity.^{13,15} The PFS showed good convergent validity against measures of performance fatigability in older American adults¹⁵, and its predictive validity for mobility decline has recently been demonstrated.¹⁹ Considering these aspects, together with the fact that, in Dutch, an instrument to measure perceived fatigability that is suitable for older adults is still lacking, the purpose of this study was to translate the PFS into Dutch and subsequently investigate its validity and reliability among Dutch hospitalized older adults.

METHODS

Study design and patient selection

A prospective cohort of patients aged 70 years and older admitted to the University Medical Center of Groningen (UMCG) were recruited from April 2018 to April 2019. Both acute and elective patients admitted to cardiology, oncology, vascular and hepatobiliary, trauma, and internal medicine units were screened for participation every weekday. Inclusion criteria were age 70 years and older and an expected hospital stay of at least two consecutive days. Exclusion criteria were: no understanding of the Dutch spoken language, any (temporary) cognitive condition that influence decision making capacity, and no written informed consent. Enrollment took place within the first four days of hospital admission. Baseline questionnaire data included demographics, activities of daily living, cognitive functioning, the Frailty Phenotype, and the PFS. To assess

reliability, a random selection of the included patients were invited for a second assessment of the PFS one day after their baseline assessment. No surgical procedures were conducted between test and retest assessments. After hospital discharge, medical charts were consulted to measure comorbidity status and type of hospital admission and treatment (acute, elective, non-surgical, surgical). Data were collected by trained research staff of the geriatric medicine department at the UMCG.

Translation procedure of the PFS

The PFS was originally developed in the United States to measure perceived physical and mental fatigability for adults aged 60 years and older. The PFS consists of ten items describing an activity with a certain intensity and duration. For each item, three questions are formulated: first the expected or imagined physical fatigue immediately after completing the listed activity; second the expected or imagined mental fatigue immediately after completing the listed activity; and third, whether or not the activity was done in the past month. For each item, physical and mental fatigue are rated on a scale from 0 to 5, where 0 represents no fatigue and 5 represents extreme fatigue. Scoring results in a physical fatigability score and a mental fatigability score, each ranging from 0 to 50 points, with higher scores indicating greater fatigability. The original questionnaire was validated for the PFS physical subscale, which yielded four factors: 1. moderate to high intensity activities, 2. social activities, 3. sedentary activities, and 4. lifestyle or light intensity activities.¹⁵

The PFS was translated into Dutch according to a formal forward/backward translation protocol.²⁰ Briefly, the following stages were completed: (1) forward translation by two independent bilingual, native Dutch speaking translators; (2) synthesis of discrepancies of the two translated versions into one translated version by the translators and an observant; (3) back translation by two independent bilingual, native English speaking translators, who were blinded for the original version of the PFS; (4) review and syntheses of the two translated versions by an expert panel consisting of the original developer, a linguistic, the translators, and a geriatrician; (5) field testing of the pre-final version among the target population. The field test included conducting interviews using the Three-Step Test Interview method in a small sample of the target population to assess comprehensibility of the Dutch PFS.^{21,22}

Questionnaires

Concurrent with the PFS, patient characteristic and questionnaire data were ascertained for the purpose of descriptive statistics and construct validity. Demographic characteristics included age, sex, and level of education (low (<high school), moderate (high school), high (college / university))²³, all measured via self-report questionnaires. Clinical characteristics included the Charlson Comorbidity Index (CCI)²⁴ and length of hospital stay (LoS), both obtained from medical records.

Pre-hospital existing problems with performing instrumental activities of daily living (iADL) were assessed using the seven original Lawton and Brody items including the use of a walking aid.²⁵ Patients were asked to refer to the situation two weeks before hospital admission. Total scores ranged from 0 to 8 with higher scores indicating more limitations in iADL. Pre-hospital frailty status was assessed using the Frailty Phenotype based on self-report.^{26,27} Unintended weight loss was considered when patients reported losing three or more kilograms of body weight in the past month. The items exhaustion, slowness, weakness, and low physical activity referred to the situation two weeks before hospital admission. Exhaustion was assumed when people scored 'yes' to both questions, 'everything I did was an effort', and 'I could not get going'.²⁸ Slowness was assumed when people were not able to walk outside for 5 minutes. Weakness was assumed when people reported to have difficulties rising a chair. Low physical activity was assigned when people reported not being physically active for at least 30 minutes per week. Total scores ranged from 0 to 5, with higher scores indicating more frailty.

Cognitive functioning was assessed using the Short Blessed Test (SBT).²⁹ The SBT included three questions addressing orientation: two items addressing attention and one item addressing memory. A weighted scoring was used ranging from 0 to 28 points, with higher scores indicating more cognitive impairment.

Statistical analysis

Descriptive statistics

For all baseline demographic and clinical characteristics, descriptive statistics were calculated as mean (standard deviation), median (inter-quartile range), or N (%).

Validity

Construct validity, defined as the degree to which scores of a measurement are consistent with hypotheses, was assessed by hypotheses testing and factor analyses.

1. Hypothesis testing was performed using iADL and physical frailty as related dissimilar constructs and cognitive functioning as an unrelated construct. In total six hypotheses were formulated:
 - a. The PFS physical score has a moderate positive correlation ($r = 0.3 - 0.5$) with the sum score of the iADL items.
 - b. The PFS physical score has a moderate positive correlation ($r = 0.3 - 0.5$) with the Frailty Phenotype.
 - c. The PFS mental score has a moderate positive correlation ($r = 0.3 - 0.5$) with the sum score of the iADL items.
 - d. The PFS mental score has a moderate positive correlation ($r = 0.3 - 0.5$) with the Frailty Phenotype.
 - e. The PFS physical score has a low correlation ($r \leq 0.3$) with the sum score of the SBT.
 - f. The PFS mental score has a low correlation ($r \leq 0.3$) with the sum score of the SBT.

Pearson's Correlation Coefficients or Spearman's Rank Correlation Coefficients were used based on normal or non-normal data distribution, respectively. Construct validity of the PFS was considered to be good if at least five hypotheses were confirmed.³⁰

2. Confirmatory Factor Analysis was utilized to investigate if the factor structure of the original PFS physical subscale was also found in the Dutch sample. To evaluate model fit, standardized root mean square residual ($SRMR \leq 0.08$) and comparative fit index ($CFI \geq 0.95$) were used.³¹
3. Exploratory Factor Analyses (EFA) of the PFS physical and mental subscales were performed to explore the best factor structure underlying the Dutch sample. We followed the procedure suggested by Osborne & Ostelo (2005)³², which is explained in detail in Additional File 1. A good fit was considered when the items within the factor had all loadings >0.3 , but preferably >0.5 , and no or few items existed with cross-loadings >0.32 .³² Finally, the interrelatedness of the items within each factor of the

best-fitted factor solution was examined by Cronbach's alpha. Cronbach's alpha of 0.70 or higher was considered appropriate.³³

Reliability

Reliability, defined as the extent to which the measure is free of measurement error, was assessed for both the PFS physical and mental scores of a random subsample who gave consent to a second PFS assessment one day after the baseline assessment.

1. Test-retest reliability was calculated with an intraclass correlation coefficient (ICC) using a two-way mixed effects model for absolute agreement. Cut-off values for interpretation of the ICC were <0.5 poor, ≥0.5 and <0.75 moderate, ≥0.75 and <0.9 good, ≥0.9 excellent reliability.³⁴
2. Agreement, defined as the ability of the instrument to result in the same scores when applied twice under the same conditions, was assessed by calculating Bland-Altman plots. Therefore, mean change scores of baseline and retest assessments were plotted against the difference of both scores. Paired t-tests were calculated to assess systematic measurement error.^{33,35}
3. Measurement error was calculated by the standard error of measurement (SEM) using the following formula: $SEM = SD(TO) \times \sqrt{1-r}$. SD(TO) refers to the standard deviation of the baseline measurement; r refers to the ICC calculated in step 1. To interpret the SEM, scores were converted to percentages of the scale range and cut-off values were used as suggested by Ostelo and colleagues (2004)³⁶: ≤5% very good, ≤10% good, >10% and <20% moderate, ≥20% poor.
4. Smallest detectable change (SDC) defined as the smallest unit of change that can be detected by the instrument beyond measurement error was calculated by the following formula: $SDC = 1.96 \times \sqrt{2} \times SEM$. Both the absolute SDC value as well as the SDC as a percentage of the scale range were calculated.

Subgroup and sensitivity analysis

Construct validity was determined for both the total group, and a subgroup of elective and acute patients, and the subgroup of surgical and non-surgical patients. Data were imputed by item mean imputation using sample-specific correction factors for sex and activity level where three or less items of either the PFS physical or mental subscale were missing, but the related question on whether the activity had been done in the past month was completed.³⁷ Sensitivity analyses included validity and reliability analysis of patients with complete PFS data. Imputations and CFA analysis were performed in STATA SE, version 14. All other analyses were performed using Statistical Package for Social Science (SPSS) software, version 23.

RESULTS

Translation of the PFS

During the forward translation of the PFS some language difficulties were identified, namely the translation of the answer category 'extreme fatigue' lead to discussion. Consensus was reached with assistance of a third independent person. The unit pounds (lbs.) (item g) is not used in the Dutch language and therefore needed to be converted into kilograms. The examples 'senior center', 'hiking', and 'Zumba' revealed translation problems, but consensus was reached by the translators. All solutions were approved by the members of the expert meeting. Twelve patients participated in the field test of the pre-final version. Patients sometimes reported difficulties distinguishing between physical and mental fatigue. In general, we identified two types of patients: 1. Patients who distinguished physical from mental fatigue using different definitions for both constructs; 2. Patients who did not distinguish these two concepts. However, all interviewed patients mentioned that both concepts were somehow related to each other. To clarify the construct of mental fatigue to patients that requested, we stated that mental fatigability refers to fatigue of the mind, not fatigue of the body or muscles. Patients reported to understand the questions, so no further adjustments of the Dutch PFS were done. The final version of the Dutch PFS is available on request by dr. Nancy W. Glynn.³⁸

Table 1. Baseline demographic and clinical characteristics of the study samples.

Characteristic	Validation sample (n = 233)	Reliability sample (n = 50)
Age in years, median (IQR)	76 (73; 81)	76 (72; 82)
range	70 – 95	70 – 87
Sex, male	154 (66)	27 (54)
Educational level		
low (< high school)	61 (26)	13 (27)
moderate (high school)	84 (36)	23 (48)
high (college / university)	83 (36)	12 (25)
Medical specialty associated with hospital admission		
trauma and orthopedics	17 (7)	2 (4)
vascular and hepatobiliary	37 (16)	6 (12)
medical oncology	40 (17)	11 (22)
internal medicine	20 (9)	9 (18)
cardiology	114 (50)	18 (36)
Comorbidity ^a , median (IQR)	2 (1; 3)	2 (1; 3)
range	0 – 10	0 – 9
LoS in days, median (IQR)	4 (3; 8)	5 (4; 11)
range	1 – 37	2 – 31
Fatigability score		
physical, mean (SD)	23.7 (11.5)	30.3 (9.8)
mental, mean (SD)	14.9 (13.5)	24.4 (13.6)

All numbers are presented in n (%), unless indicated otherwise.

a. Comorbidity score was assessed using the Charlson Comorbidity Index

Abbreviations: IQR, Inter Quartile Range; LoS, Length of Stay; SD, Standard Deviation.

Descriptive statistics

After the field test, the Dutch PFS was administered to 249 patients. Sixteen patients (6%) were excluded from the analysis, because they had more than three missing values for the PFS, leaving 233 consecutive patients in the analytic sample for validation of the PFS. A flowchart of the study participants is included in Additional File 2, Figure S2-1. A random sample of 50 patients were selected out of the validation sample for re-assessment of the PFS one day after the baseline assessment to evaluate reliability. The median age of patients was 76 years (IQR: 73; 81) and the majority of the patients were male (66%). Table 1 presents baseline demographic and clinical characteristics of the total study population and reliability sample.

Table 2. Expected and observed correlations of the Dutch PFS physical and mental subscales with other constructs.

Instrument (scale range)	Spearman's rank correlation			
	PFS physical score (0-50)		PFS mental score (0-50)	
	Expected	Observed	Expected	Observed
iADL (0-8)	0.3 – 0.5	0.4	0.3 – 0.5	0.3
Frailty Phenotype (0-5)	0.3 – 0.5	0.6	0.3 – 0.5	0.4
SBT (0-28)	≤ 0.3	0.1	≤ 0.3	0.1

Abbreviations: iADL, instrumental Activities of Daily Living; PFS, Pittsburgh fatigability scale; SBT, Short Blessed Test.

Construct validity

Spearman's correlation coefficients between PFS subscales and the other constructs are presented in Table 2. Five out of the six hypotheses were confirmed. Confirmatory Factor Analysis of the original four-factor model of the PFS physical subscale resulted in poor model fit (SRMR: 0.29; CFI: 0.75). Kaiser-Meyer-Olkin values were 0.84 and 0.89 for PFS physical and mental subscales, respectively. Bartlett's test of sphericity were >0.001 for both subscales, and no item correlations were <0.2 or >0.9, indicating that EFA could be applied. Data for the PFS physical and mental items were non-normally distributed [Additional File 2, Table S2-1a and 1b, respectively], therefore the principle axis factoring extraction method (PAF) was used, with oblique rotation to estimate the optimal factor structure underlying the Dutch data. Regarding the PFS physical subscale, first the preselected four-factor solution was estimated, resulting in the following eigenvalues and cumulative percentage explained variance: 4.796 (48.0%), 1.711 (65.1%), 0.799 (73.1%), 0.620 (79.3%). Four items did not load on their intended scales based on the factor structure underlying the original validation sample. The scree plot showed points of inflexion three factors, which was indicative for a two factor solution underlying the Dutch dataset. Therefore, PAF using oblique oblimin rotations for a two-factor, three-factor, and five-factor solution were performed as well. The three-factor and five-factor solutions did not result in a clean factor structure. The two-factor solution resulted in two factors of each five items with factor loadings of >0.5 with two theoretically interpretable factors. The first factor included items a, b, d, g, and j and was named 'activities requiring physical effort'. The second factor included items c, e, f, h, and i which we named 'activities requiring low physical effort'. One cross-loading of 0.33

for item c (light household activity for 1 hour) existed in the two-factor solution. Cronbach's alpha analysis showed good relatedness of the items of factor 1 (α : 0.87), and factor 2 (α : 0.83). Regarding the PFS mental subscale, the scree plot indicated a two factor solution underlying the Dutch data, with two eigenvalues of 5.960 and 1.019 explaining 59.6% and 10.2% of the variance. The two-factor solution resulted in one factor including the items a, b, c, d, g, and j referring to 'activities requiring physical effort' and a second factor including the items e, f, h, and i referring to 'activities requiring low physical effort'. All factor loadings were >0.5 , but again item c (light household activity for 1 hour) loaded on both factors. The pattern matrices of the two-factor solutions of both the physical and mental subscales relative to the factors of the original PFS physical subscale are presented in Table 3.

Reliability

Mean PFS physical and mental scores, ICC, SEM, and SDC are presented in Table 4. ICC of the PFS physical and mental subscales were 0.80 (95%CI: 0.68; 0.88) and 0.81 (95%CI: 0.68; 0.89), respectively, indicating good reliability. The Bland Altman plot showed good agreement for the PFS physical subscale (mean difference: -1.60, 95%CI: -3.27; 0.06). The PFS mental subscale had a small mean systematic difference between baseline and re-assessment (mean difference: -2.30, 95% CI: -4.54; -0.05) (Figure 1). SEM was good for the PFS physical subscale (9%), and moderate for the PFS mental subscale (12%) (Table 4). Smallest detectable change beyond measurement error was 12 (24%) and 16 (33%) points for the PFS physical and mental subscales, respectively (Table 4).

Subgroup and sensitivity analysis

We performed subgroup analyses of elective ($n = 92$) and acutely admitted ($n = 97$) patients as well as surgical ($n = 142$) and non-surgical patients ($n = 60$) [Additional file 3]. For all subgroups at least five out of six hypotheses were confirmed. EFA resulted in a two-factor solution for the PFS physical subscale for every subgroup. EFA of the mental subscale resulted in either a clean single factor solution (non-surgical patients), or a two factor solution (elective, acute, and surgical subgroups). The results of the sensitivity analysis using only data of patients with complete PFS data ($n = 201$) were generally comparable and did not lead to different conclusions [Additional file 3].

Table 3. Exploratory Factor Analyses results: Pattern matrices of the Dutch PFS physical and mental subscales.

Factor / Item description	Physical subscale		Mental subscale	
	Factor 1 (α : 0.87) ^a	Factor 2 (α : 0.83) ^a	Factor 1 (α : 0.92) ^b	Factor 2 (α : 0.80) ^b
Moderate to high intensity activities				
B Brisk walking	0.800	-	0.984	-
D Heavy gardening	0.657	-	0.699	-
G Strength training	0.781	-	0.596	-
J High intensity activity	0.832	-	0.525	-
Lifestyle or light intensity activities				
C Household activity	0.325	0.562	0.504	0.345
A Leisurely walk	0.602	-	0.913	-
Social activities				
H Participate in social activity	-	0.806	-	0.850
I Hosting social event	-	0.753	-	0.794
Sedentary activities				
E Watching TV	-	0.667	-	0.606
F Sitting quietly	-	0.596	-	0.546

Factors were extracted by Principle Axis Factor analysis and rotated by oblique oblimin rotation.

Only factor loadings >0.3 are presented.

a. Cronbach's alpha was calculated using items B, D, G, J, A (Factor 1); C, E, F, H, I (Factor 2).

b. Cronbach's alpha was calculated using items B, C, D, G, J, A (Factor 1); E, F, H, I (Factor 2).

DISCUSSION

This study showed that the Dutch version of the Pittsburgh Fatigability Scale had good content validity, construct validity, and test-retest reliability for measuring perceived physical and mental fatigability in hospitalized adults aged 70 years and older.

The CFA analysis performed on the original four-factor solution of the PFS physical subscale resulted in poor model fit. A possible explanation for this finding was that the original validation study retrieved four factors from 26 initial items. During the development process, EFA was used to reduce the number of items resulting in the ten items included in the original PFS. The guideline provided by Osborne & Costello stated that factors with fewer than three items are likely to be unstable and should advocate to reduce the number of factors.³² However, the original

Table 4. Mean baseline and retest scores and reliability properties of the Dutch PFS (n = 50).

PFS	Baseline mean (SD)	Retest mean (SD)	ICC ^a (95% CI)	SEM	SEM (%) ^b	SDC	SDC (%) ^b
Physical subscale	30 (10)	31 (10)	0.80 (0.68;0.88)	4	9%	12	24%
Mental subscale	24 (14)	26 (13)	0.81 (0.68;0.89)	6	12%	16	33%

Abbreviations: ICC, Intra-class Correlation Coefficient; PFS, Pittsburgh Fatigability Scale; SD, Standard Deviation; SDC, Smallest Detectable Change; SEM, Standard Error of Measurement; MIC, Minimal Important Change.

- a. A 2-way mixed effect model was used.
b. SEM (%) and SDC (%) are expressed in percentages of the scale range (0-50); percentages are rounded off.

sample had a high subject to item ratio of 18:1 and all items had strong loadings on their corresponding factor (≥ 0.7), which is an indication of strong factors despite the low number of items per factor.³² The two-factor solution found in the current data may have been appropriate in the original sample as well, but this was not reported. In the current data, a clean two-factor structure was found. However, one cross-loading existed for the item ‘light to moderate household activities for 1 hour’. This cross-loading existed in both the main analysis and all subgroup analyses, which could indicate an ambiguous translation. The predefined examples of light household activities included ‘cleaning, cooking, dusting, straightening up, baking, making beds, dishwashing, and watering plants’. There may have been confusion regarding the interpretation of the intensity of the example activities. It is plausible that patients may have interpreted cleaning as mopping or vacuuming, which has a higher intensity than for instance dusting. Moreover, cooking requires more mental and cognitive processes, such as planning and sequencing, than for instance dishwashing. Patients could have interpreted cooking as an activity that required more mental than physical effort, which may explain its cross-loading on the second factor referring to activities that require less physical effort. Lastly, men could have interpreted household activities differently from women, which could have led to the cross-loading found for household activities in the current dataset.

In the current study, mean fatigability scores were on average higher compared to mean fatigability scores of the original validation study.¹⁵

Figure 1. Bland Altman plots of the Dutch Pittsburgh Fatigability Scale physical (a) and mental (b) subscales. X-axis represent mean fatigability scores; Y-axis represents difference of fatigability scores (baseline – retest assessments).

This is not surprising given the sample was drawn from hospitalized older adults. Nonetheless, supplemental analyses comparing elective and acute patients showed that PFS scores were lower for elective patients. These patients are more likely to be representative of the general population as their fatigability scores were comparable to community-dwelling Spanish older adults within the same age group³⁹ and community-dwelling Americans with comparable number of comorbidities.⁴⁰ The highest fatigability scores were found in acute and non-surgical patients, which may be attributed to the illness and subsequent disability onset in the weeks before hospital admission that is commonly seen in older medical patients.⁴¹ This explanation seems plausible, because in the Dutch sample we also found lower proportions of patients who performed activities that required physical effort in the past month compared to the original validation sample.¹⁵ Exception to this was the proportion of patients that performed a brisk or fast walk in the previous month. In our study, 29% of the patients reported to have done a brisk or fast walk for one hour compared to 17% of the older US adults in the original validation study.¹⁵ This difference was unexpected, but likely explained by cross-cultural differences. Like Americans are known for their hiking culture⁴², Dutch are known for their cycling culture.⁴³ Therefore it could be that our Dutch sample had difficulty distinguishing between a brisk and leisurely walk as walking is less culturally embedded.

Test-retest reliability was assessed using a measurement interval of one day to minimize the risk that conditions of the patients changed. The target population, hospitalized older persons, may fluctuate over time concerning illness and symptoms. The risk of recall bias has been considered, however it was expected that patients did not remember their initial answers, partly due to stressful circumstances patients are exposed to during hospitalization and because the first PFS assessment was part of a larger set of questionnaires assessed at the same time. Even though we took a random sample of the validation sample to test reliability, the reliability sample had higher initial PFS scores compared to the validation sample. As a result, there were few observations at the beginning of the scale range of corresponding Bland Altman plots, especially for the PFS physical subscale. However, the Bland Altman plots showed a homogeneous spread over the scale-range, supporting good agreement between the test and retest of both the PFS physical and mental subscales.

Strengths and limitations

This is the first study that translated the PFS into Dutch. Strengths of this study include the relative large sample size as we included more than twice the recommended sample for validation studies.³³ A second strength is the exhaustive translation methodology including pretesting of the pre-final version.²⁰ There were limitations as well. First, this study was performed in one hospital, so there was risk of selection bias. Considering the fact that the UMCG provides both primary and specialist care for patients from several regions of the Netherlands, together with the broad inclusion criteria including patients from both medical and surgical wards, we have evaluated the selection bias as low. However, the robustness of the Dutch PFS should be further examined in other Dutch cohorts. Second, relative high missing values (7 – 8%) were present in the current study for the two items including high intensity activities. However, sensitivity analysis with only complete PFS data led to the same results, indicating robustness of our findings. Third, no convergent validity against performance-based fatigability assessment was performed due to lack of suitable measurement in this hospitalized population.

Future research

The current study demonstrated that perceived physical and mental fatigability in Dutch hospitalized adults aged 70 years and over can now be assessed using the valid and reliable Pittsburgh Fatigability Scale. Future research should investigate the Dutch PFS's validity and reliability in other Dutch samples like community-dwelling older adults. Further validation for the purpose of convergent validity of the PFS Dutch version physical subscale is recommended by using performance fatigability assessments, which have demonstrated to be feasible in older community-dwelling adults^{9,10,33}. This is the first study that presented the factor solution of the PFS mental subscale, which in our population of hospitalized older adults indicated two factors underlying mental fatigability as the cleanest structure. Future research should demonstrate whether this factor structure also applies to other populations and subgroups. Furthermore, when the PFS is applied in cohort studies or as intervention outcome, future research should investigate its responsiveness and interpretability to examine the extent to which the PFS is able to measure change in fatigability over time.

CONCLUSION

The translated Dutch version of the Pittsburgh Fatigability Scale is a valid and reliable instrument to assess perceived physical and mental fatigability in hospitalized adults aged 70 years and older.

REFERENCES

1. Yu DSF, Lee DTF, Man NW. Fatigue among older people: A review of the research literature. *Int J Nurs Stud*. 2010;47(2):216–228. doi:10.1016/j.ijnurstu.2009.05.009
2. Avlund K. Fatigue in older adults: an early indicator of the aging process? *Aging Clin Exp Res*. 2010;22(2):100–115. <http://www.ncbi.nlm.nih.gov/pubmed/20440097>. Accessed January 20, 2017.
3. Zengarini E, Ruggiero C, Pérez-Zepeda MU, et al. Fatigue: Relevance and implications in the aging population. *Exp Gerontol*. 2015;70:78–83. doi:10.1016/j.exger.2015.07.011
4. Moreh E, Jacobs JM, Stessman J. Fatigue, function, and mortality in older adults. *Journals Gerontol – Ser A Biol Sci Med Sci*. 2010;65 A(8):887–895. doi:10.1093/gerona/glq064
5. Hardy SE, Studenski SA. Fatigue and function over 3 years among older adults. *J Gerontol A Biol Sci Med Sci*. 2008;63(12):1389–1392. doi:http://dx.doi.org/10.1093/gerona/63.12.1389
6. Simonsick EM, Glynn NW, Jerome GJ, Shardell M, Schrack JA, Ferrucci L. Fatigued, but Not Frail: Perceived Fatigability as a Marker of Impending Decline in Mobility-Intact Older Adults. *J Am Geriatr Soc*. 2016;64(6):1287–1292. doi:10.1111/jgs.14138
7. Eldadah BA. Fatigue and Fatigability in Older Adults. *PM R*. 2010;2(5):406–413. doi:10.1016/j.pmrj.2010.03.022
8. Dittner AJ, Wessely SC, Brown RG. The assessment of fatigue: A practical guide for clinicians and researchers. *J Psychosom Res*. 2004;56(2):157–170. doi:10.1016/S0022-3999(03)00371-4
9. Whitehead L. The Measurement of Fatigue in Chronic Illness: A Systematic Review of Unidimensional and Multidimensional Fatigue Measures. *J Pain Symptom Manage*. 2009;37(1):107–128. doi:10.1016/j.jpainsymman.2007.08.019
10. Alexander NB, Taffet GE, McFarland Horne F, et al. Bedside-to-Bench Conference: Research Agenda for Idiopathic Fatigue and Aging. *J Am Geriatr Soc*. 2010;58(5):967–975. doi:10.1111/j.1532-5415.2010.02811.x.
11. LaSorda KR, Gmelin T, Kuipers AL, et al. Epidemiology of Perceived Physical Fatigability in Older Adults: The Long Life Family Study. *J Gerontol A Biol Sci Med Sci*. 2019. doi:10.1093/gerona/glz288.
12. Murphy SL, Smith DM. Ecological measurement of fatigue and fatigability in older adults with osteoarthritis. *J Gerontol A Biol Sci Med Sci*. 2010;65(2):184–189. doi:10.1093/gerona/glp137
13. Kim I, Hacker E, Ferrans CE, Horswill C, Park C, Kapella M. Evaluation of fatigability measurement: Integrative review. *Geriatr Nurs (Minneap)*. 2018;39(1):39–47. doi:10.1016/j.gerinurse.2017.05.014
14. Simonsick EM, Schrack JA, Glynn NW, Ferrucci L. Assessing fatigability in mobility-intact older adults. *J Am Geriatr Soc*. 2014;62(2):347–351. doi:10.1111/jgs.12638
15. Glynn NW, Santanasto AJ, Simonsick EM, et al. The Pittsburgh fatigability scale for older adults: Development and validation. *J Am Geriatr Soc*. 2015;63(1):130–135. doi:10.1111/jgs.13191
16. Avlund K, Kreiner S, Schultz-Larsen K. Functional ability scales for the elderly: A validation study. *Eur J Public Health*. 1996;6(1):35–42. doi:10.1093/eurpub/6.1.35
17. Tiesinga LJ, Dassen TWN, Halfens RJG. DUFs and DEFS: Development, reliability and validity of the Dutch fatigue scale and the Dutch exertion fatigue scale. *Int J Nurs Stud*. 1998;35(1-2):115–123. doi:10.1016/S0020-7489(98)00005-4
18. Yang CM, Wu CH. The Situational Fatigue Scale: A different approach to measuring fatigue. *Qual Life Res*. 2005;14(5):1357–1362. doi:10.1007/s11136-004-5680-0
19. Simonsick EM, Schrack JA, Santanasto AJ, Studenski SA, Ferrucci L, Glynn NW. Pittsburgh Fatigability Scale: One-Page Predictor of Mobility Decline in Mobility-Intact Older Adults. *J Am Geriatr Soc*. 2018;66(11):2092–2096. doi:10.1111/jgs.15531
20. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186–3191. doi:10.1097/00007632-200012150-00014
21. Hak T, Veer K Van Der, Jansen H. The Three-Step Test-Interview (TSTI): An observational instrument for pretesting self-completion questionnaires. 2004. <https://repub.eur.nl/pub/1265>. Accessed 17 Jan 2020.

22. Pool JJM, Hiralal SR, Ostelo RWJG, van der Veer K, de Vet HCW. Added value of qualitative studies in the development of health related patient reported outcomes such as the pain coping and cognition list in patients with sub-acute neck pain. *Man Ther.* 2010;15(1):43-47. doi:10.1016/j.math.2009.05.010
23. Unesco Institute for Statistics. ISCED 2011, International Standard Classification of Education. Montreal, Canada; 2012.
24. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40(5):373-383.
25. Lawton MP, Brody EM. Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist.* 1969;9(3):179-186.
26. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: Evidence for a phenotype. *Journals Gerontol Ser a-Biological Sci Med Sci.* 2001;56(3):M146-M156. doi:10.1093/gerona/56.3.M146
27. Theou O, Cann L, Blodgett J, Wallace LMK, Brothers TD, Rockwood K. Modifications to the frailty phenotype criteria: Systematic review of the current literature and investigation of 262 frailty phenotypes in the survey of health, ageing, and retirement in Europe. *Ageing Res Rev.* 2015;21:78-94. doi:10.1016/j.arr.2015.04.001
28. Radloff LS. A Self-Report Depression Scale for Research in the General Population. *Appl Psychol Meas.* 1977;1(3):385-401. doi:10.1177/014662167700100306
29. Katzman R, Brown T, Fuld P, Peck A, Schechter R, Schimmel H. Validation of a short orientation-memory-concentration test of congestive impairment. *Am J Psychiatry.* 1983;140(6):734-739. doi:10.1176/ajp.140.6.734
30. Mokkink LB, Terwee CB, Patrick 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. *Qual Life Res.* 2010;19(4):539-549. doi:10.1007/s11136-010-9606-8
31. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model a Multidisciplinary J.* 1999;6(1):1-55. doi:10.1080/10705519909540118
32. Osborne JW, Costello AB. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Pract Assess Res Eval.* 2005;10(7).
33. De Vet, Henrica C. W. Terwee CB, Knol DL, Mokkink LB. Measurement in Medicine. first. Cambridge: Cambridge University Press; 2011.
34. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
35. Berchtold A. Test-retest: Agreement or reliability? *Methodol Innov.* 2016;9:205979911667287. doi:10.1177/2059799116672875
36. Ostelo RWJG, De Vet HCW, Knol DL, Van Den Brandt PA. 24-Item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *J Clin Epidemiol.* 2004;57(3):268-276. doi:10.1016/j.jclinepi.2003.09.005
37. Cooper R, Popham M, Santanasto AJ, Hardy R, Glynn NW, Kuh D. Are BMI and inflammatory markers independently associated with physical fatigability in old age? *Int J Obes.* 2019;43(4):832-841. doi:10.1038/s41366-018-0087-0
38. Glynn NW. Pittsburgh Fatigability Scale. University of Pittsburgh Graduate School of Public Health.
39. Pérez LM, Roqué M, Glynn NW, et al. Validation of the Spanish version of the Pittsburgh Fatigability Scale for older adults. *Aging Clin Exp Res.* 2019;31(2):209-214. doi:10.1007/s40520-018-0959-0
40. Carlozzi NE, Boileau NR, Murphy SL, Braley TJ, Kratz AL. Validation of the Pittsburgh Fatigability Scale in a mixed sample of adults with and without chronic conditions. *J Health Psychol.* 2019. doi:10.1177/1359105319877448
41. Covinsky KE, Palmer RM, Fortinsky RH, et al. Loss of independence in activities of daily living in older adults hospitalized with medical illnesses: Increased vulnerability with age. *J Am Geriatr Soc.* 2003;51(4):451-458. doi:10.1046/j.1532-5415.2003.51152.x
42. Chamberlin S. On the trail: A history of American hiking. *Trail A Hist Am Hik.* 2016:1-243.
43. Dutch Cycling Embassy. Dutch Cycling Vision. 2018:30. https://www.dutchcycling.nl/images/downloads/Dutch-Cycling-Vision_EN.pdf.

SUPPLEMENTARY FILES – CHAPTER 4

ADDITIONAL FILE 1. Exploratory Factor Analysis procedure

Before starting the Exploratory Factor Analysis (EFA), suitability of the data for performing EFA was tested by the Kaiser-Meyer-Olkin test (≥ 0.8) and Bartlett's test of sphericity ($P < 0.05$).¹ All EFA were performed using the procedure suggested by Osborne and Costello (2005).² First, the extraction and rotation methods were selected based on data characteristics. Concerning the extraction method, either maximum likelihood (ML) or principle axis factoring (PAF) extraction methods were used, depending on normal (ML) or non-normal (PAF) data distribution respectively. The rotation method should be selected based on the correlation between the factors: orthogonal rotations are applied to uncorrelated factors, oblique rotations are applied when factors correlate. In the current study, for all EFA analyses, direct oblique rotation was used to allow the factors to correlate. Second, the correlations between items were inspected. Items with correlations < 0.20 with all other items were deleted immediately³; items with correlations > 0.90 might indicate multicollinearity. Finally, to explore the ideal number of factors retained in our Dutch sample we performed several factor analyses as suggested by Osborne & Costello (2005): (a) the four factor solution underlying the original PFS data was evaluated; (b) the factor solution suggested by the scree-test was evaluated; (c) if a and b resulted in different factor solutions, all factors above and below the suggested factor solutions of a and b were evaluated as well. After rotation we compared the different factor solutions to select the solution that fits best to the Dutch sample.

1. Field A. Discovering Statistics Using SPSS. Third edit. London: SAGE Publications Ltd; 2000. 821 p.
2. Osborne JW, Costello AB. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. Pract Assess Res Eval. 2005;10(7).
3. De Vet, Henrica C. W. Terwee CB, Knol DL, Mokkink LB. Measurement in Medicine. first. Cambridge: Cambridge University Press; 2011. 347 p.

ADDITIONAL FILE 2. Pittsburgh Fatigability Scale¹ (Dutch translation)

Met onderstaande vragenlijst geeft u het niveau van lichamelijke en mentale vermoeidheid aan die u verwacht te vernemen onmiddellijk na het uitvoeren van elk van de 10 opgesomde activiteiten. Omcirkel het juiste getal voor uw lichamelijke en mentale vermoeidheid bij elke activiteit (a-j) op een schaal van 0 tot 5, waarbij '0' niet vermoeid en '5' ernstig vermoeid betekent.

In de rechterkolom geeft u aan of u de betreffende activiteit hebt ondernomen in de afgelopen maand. Als uw antwoord 'nee' is, maak dan een zo goed mogelijke inschatting met betrekking tot de vragen over vermoeidheid (zie voorbeeld activiteit 2).

Vul alle drie de kolommen in voor elke activiteit, ook voor de activiteiten die u niet zelf doet. Let goed op de duur (bv. 30 minuten) en intensiteit (bv. licht / zwaar inspannend) van elke activiteit.

	Lichamelijke vermoeidheid					Mentale vermoeidheid					Heeft u deze activiteit gedaan in <u>de afgelopen maand</u> ?			
	Niet vermoeid			Ernstig vermoeid		Niet vermoeid			Ernstig vermoeid					
	0 <-----> 5					0 <-----> 5								
Voorbeelden:														
Voorbeeld activiteit 1	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
Voorbeeld activiteit 2	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee

	Lichamelijke vermoeidheid						Mentale vermoeidheid						Heeft u deze activiteit gedaan in de <u>afgelopen</u> maand?	
	Niet vermoeid			Ernstig vermoeid			Niet vermoeid			Ernstig vermoeid				
	0 ←————→ 5						0 ←————→ 5							
a. Een ontspannen wandeling voor 30 minuten	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
b. Stevig doorwandelen voor 1 uur	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
c. Licht inspannend huishoudelijk werk voor 1 uur (<i>schoonmaken, koken, stoffen, opruimen, bed opmaken, afwassen, planten water geven</i>)	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
d. Zwaar inspannend werk buiten, zoals tuinieren voor 1 uur (<i>maaien (duwmaaier), harken, onkruid wieden, planten verpotten, sneeuw schuiven</i>)	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
e. Televisie kijken voor 2 uur	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
f. Rustig zitten voor 1 uur	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
g. Matig tot zwaar inspannende krachttraining voor 30 minuten (<i>met gewichten of fitness apparaten zwaarder dan 2,5 kg, opdrukken</i>)	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
h. Deelname aan een sociale activiteit voor 1 uur (<i>feestje, diner, senioren club, bezoek van of bij familie/vrienden, kaarten, briden</i>)	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
i. Gastheer/gastvrouw zijn bij een sociale aangelegenheid voor 1 uur (<i>voorbereidingstijd niet meegerekend</i>)	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee
j. Zwaar inspannend sporten voor 30 minuten (<i>hardlopen, nordic walking, fietsen (niet elektrisch), zwemmen, tennis, fitness, dansen</i>)	0	1	2	3	4	5	0	1	2	3	4	5	Ja	Nee

Copyright 2014, University of Pittsburgh. All rights reserved. Translated 2017 by University Medical Center of Groningen (UMCG) under license

1. The Pittsburgh Fatigability Scale for Older Adults: Development and Validation. (Authors: Nancy W. Glynn, Adam J. Santanasto, Eleanor M. Simonsick, Robert M. Boudreau, Scott Beach, Richard Schulz and Anne B. Newman. © 2014, University of Pittsburgh. All Rights Reserved.)

Figure S3 – 1. Flowchart of study participants

^a. Hospitalized patients aged 70 years and older admitted to one of the participating wards
^b. A random sample (n = 50) of the validation sample participated in the reliability sample

Table S3-1a. Descriptive statistics for the items of the PFS physical subscale (n = 249)

Original Subscales	Items	Mean (SD)	Median	Skewness (SE)	Kurtosis (SE)	Activity performed [#] (n (%))	Missing (n (%))
Moderate to high intensity activity	b. Brisk or fast walk for 1 hour	3.73 (1.58)	4.00	-1.060 (0.158)	-0.044 (0.314)	70 (29)	11 (4)
	d. Heavy gardening or outdoor work for 1 hour	3.39 (1.83)	4.00	-0.783 (0.158)	-0.844 (0.314)	102 (42)	11 (4)
	g. Moderate- to high- intensity strength training for 30 min.	3.92 (1.42)	5.00	-1.357 (0.160)	0.969 (0.320)	52 (21)	19 (8)
	j. High-intensity activity for 30 min.	3.91 (1.60)	5.00	-1.347 (0.160)	0.566 (0.319)	61 (26)	18 (7)
Social activity	h. Participating in a social activity for 1 hour	1.26 (1.61)	0.00	1.002 (0.156)	-0.239 (0.311)	184 (76)	6 (2)
	i. Hosting a social event for 1 hour	1.92 (1.85)	2.00	0.390 (0.156)	-1.307 (0.312)	133 (53)	7 (3)
Sedentary activity	e. Watching TV for 2 hours	0.73 (1.25)	0.00	1.669 (0.157)	1.859 (0.312)	218 (90)	8 (3)
	f. Sitting quietly for 1 hour	0.59 (1.15)	0.00	1.963 (0.156)	2.979 (0.310)	231 (96)	5 (2)
Lifestyle or light intensity activity	a. Leisurely walk for 30 min.	2.59 (1.86)	3.00	-0.168 (0.156)	-1.339 (0.310)	141 (58)	5 (2)
	c. Light household activity for 1 hour	1.90 (1.76)	2.00	0.394 (0.155)	-1.167 (0.309)	189 (77)	2 (1)

Number and valid percentage of patients who reported to perform the activity in the past month.

Abbreviations: SE, standard error; SD, standard deviation

Table S3- 1b. Descriptive statistics for the items of the PFS mental subscale

Original Subscales	Items	Mean (SD)	Median	Skewness (SE)	Kurtosis (SE)	Missing (n (%))
Moderate to high intensity activity	b. Brisk or fast walk for 1 hour	1.86 (1.97)	1.00	0.474 (0.158)	-1.365 (0.314)	11 (4)
	d. Heavy gardening or outdoor work for 1 hour	1.78 (1.24)	1.00	0.566 (0.158)	-1.328 (0.316)	13 (5)
	g. Moderate- to high- intensity strength training for 30 minutes	2.23 (2.03)	2.00	0.171 (0.160)	-1.602 (0.319)	18 (7)
	j. High-intensity activity for 30 minutes	2.34 (2.08)	2.00	0.092 (0.161)	-1.649 (0.320)	20 (8)
Social activity	h. Participating in a social activity for 1 hour	1.12 (1.54)	0.00	1.154 (0.156)	0.127 (0.310)	4 (2)
	i. Hosting a social event for 1 hour	1.65 (1.79)	1.00	0.556 (0.156)	-1.161 (0.312)	7 (3)
Sedentary activity	e. Watching TV for 2 hours	0.73 (1.24)	0.00	1.572 (0.157)	1.404 (0.312)	8 (3)
	f. Sitting quietly for 1 hour	0.57 (1.14)	0.00	2.026 (0.156)	3.392 (0.310)	4 (2)
Lifestyle or light intensity activity	a. Leisurely walk for 30 minutes	1.28 (1.71)	0.00	0.976 (0.156)	-0.507 (0.312)	7 (3)
	c. Light household activity for 1 hour	1.15 (1.68)	0.00	1.191 (0.155)	-0.014 (0.309)	3 (1)

Abbreviations: SD, standard deviation; SE, standard error

Subgroups elective and acutely admitted patients

Table S4 – 1. Baseline sociodemographic and patient characteristics of total group and subgroups.

Characteristic	Total (n=233)	Elective (n = 92)	Acute (n = 97)	Unknown (n = 44)
Age in years, median (IQR)	76 (73; 81)	76 (72; 81)	76 (73; 82)	77 (72; 83)
range	70 – 95	70 – 89	70 – 95	70 – 87
Sex, male	154 (66)	60 (65)	69 (71)	25 (57)
Educational level				
low (< high school)	61 (26)	24 (26)	21 (22)	16 (36)
moderate (high school)	84 (36)	33 (36)	35 (36)	16 (36)
high (college / university)	83 (36)	33 (36)	40 (41)	10 (23)
Medical specialty				
trauma and orthopedics	17 (7)	6 (7)	9 (9)	2 (4)
vascular and hepatobiliary	37 (16)	19 (21)	9 (9)	9 (21)
medical oncology	40 (17)	20 (22)	11 (11)	9 (21)
internal medicine	20 (9)	4 (4)	6 (6)	10 (23)
cardiology	114 (50)	43 (47)	62 (64)	9 (21)
Comorbidity ^a , median (IQR)	2 (1; 3)	2 (1; 3)	2 (1; 3)	unknown
range	0 – 10	0 – 8	0 – 10	
LoS in days, median (IQR)	4 (3; 8)	4 (3; 6)	5 (3; 11)	unknown
range	1 – 37	1 – 30	1 – 37	
Fatigability score				
physical, mean (SD)	23.7 (11.5)	21.6 (11.2)	23.6 (11.2)	28.4 (11.4)
mental, mean (SD)	14.9 (13.5)	11.8 (12.1)	14.9 (13.2)	21.4 (14.7)

Notes: All numbers are presented in n (%), unless indicated otherwise.
Abbreviations: IQR, interquartile range; LoS, length of stay; SD, standard deviation.
a. Comorbidity score was assessed using the Charlson Comorbidity Index

Table S4 – 2. Expected and observed correlations of the PFS subscales per subgroup with other constructs.

Instrument (scale range)	Spearman's rank correlation			
	PFS physical score (0-50)		PFS mental score (0-50)	
	Expected	Observed	Expected	Observed
Elective patients				
iADL (0-9)	0.3 – 0.5	0.3	0.3 – 0.5	0.2
Frailty Phenotype (0-5)	0.3 – 0.5	0.6	0.3 – 0.5	0.3
SBT (0-28)	≤ 0.3	0.04	≤ 0.3	0.04
Acute patients				
iADL (0-9)	0.3 – 0.5	0.5	0.3 – 0.5	0.3
Frailty Phenotype (0-5)	0.3 – 0.5	0.6	0.3 – 0.5	0.5
SBT (0-28)	≤ 0.3	0.1	≤ 0.3	0.1

Abbreviations: iADL, instrumental activities of daily living; PFS, Pittsburgh fatigability scale; SBT, short blessed test.

EFA PFS physical subscale

Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity were performed for both subgroups, to check whether the data were suitable for EFA. KMO of the PFS physical subscale was 0.80 and 0.85 for elective and acute patients respectively, and Bartlett's test of sphericity was >0.001 for both subgroups, indicating that EFA could be applied.

Inspection of the scree-test and based on eigenvalues above 1 opt for a two-factor solution of both elective and acutely admitted patients with corresponding eigen values (explained variances): 4.55 (45%) plus 1.82 (18%), and 5.10 (51%) plus 1.72 (17%), respectively. Results of the two-factor solutions for both subgroups are presented in Table 3. Among elective patients the 2-factor solution resulted in a clean factor structure with all factor loadings > 0.5. The first factor included items (A, B, C, D, G, and J) that required physical effort to perform. The second factor included the four items (E, F, H, and I) that require less physical activity. For acutely admitted patients one crossloading exists for household activities (item C). Cronbach's alphas indicated good internal consistency of the two factors for both subgroups.

EFA PFS mental subscale

KMO of the PFS mental subscale was 0.82 and 0.88 for elective and acute patients respectively, and Bartlett's test of sphericity was >0.001 for both subgroups, indicating that EFA could be applied.

The scree test and eigenvalues above 1 opt for a two-factor solution underlying the PFS mental subscale as well. Eigen values (% explained variance) were 5.55 (55%) plus 1.23 (12%), and 5.79 (58%) and 1.09 (11%) for elective and acutely admitted patients respectively. Results of the two-factor solution of the PFS mental subscale are presented in Table 4. Among elective patients, a crossloadings exists for item I (hosting a social event for one hour). Among acutely admitted patients, a crossloading existed for item C (light household activity for one hour). Cronbach's alphas indicated good internal consistency of the two factors for both subgroups.

Table S4 – 3. EFA results: Two-factor solutions for the Dutch version of the PFS physical subscale of elective and acutely admitted patients

Factor / item description	Elective		Acute	
	Factor 1	Factor 2	Factor 1	Factor 2
	α =0.89 ^a	α =0.75 ^a	α = 0.90 ^a	α = 0.81 ^a
Moderate to high intensity activities				
B Brisk walking	.947	-	.870	-
D Heavy gardening	.792	-	.697	-
G Strength training	.708	-	.820	-
J High intensity activity	.709	-	.875	-
Lifestyle or light intensity activities				
C Household activity	.517	-	.377	.532
A Leisurely walk	.797	-	.681	-
Social activities				
H Participate in social activity	-	.751	-	.767
I Hosting social event	-	.655	-	.818
Sedentary activities				
E Watching TV	-	.655	-	.655
F Sitting quietly	-	.541	-	.585

Only factor loadings >0.3 are presented.

Cronbach's alpha's were calculated for items A, B, C, D, G, J (Factor 1) and items E, F, H, I (Factor 2)

Table S4 – 4. EFA results: Two-factor solutions for the Dutch version of the PFS mental subscale of elective and acutely admitted patients.

Factor / item description	Elective		Acute	
	Factor 1	Factor 2	Factor 1	Factor 2
	α =0.91 ^a	α =0.74 ^a	α = 0.92 ^a	α = 0.80 ^a
Moderate to high intensity activities				
B Brisk walking	.952	-	.943	-
D Heavy gardening	.807	-	.850	-
G Strength training	.767	-	.642	-
J High intensity activity	.649	-	.639	-
Lifestyle or light intensity activities				
C Household activity	.578	-	.499	.336
A Leisurely walk	.873	-	.927	-
Social activities				
H Participate in social activity	-	.593	-	.754
I Hosting social event	.512	.408	-	1.006
Sedentary activities				
E Watching TV	-	.519	-	.489
F Sitting quietly	-	.583	-	.460

Only factor loadings >0.3 are presented.

a. Cronbach's alpha's were calculated for items A, B, C, D, G, J (Factor 1) and items E, F, H, I (Factor 2)

Subgroups surgical and non-surgical patients

Table S4 – 5. Baseline sociodemographic and patient characteristics of total group and subgroups.

Characteristic	Total (n=233)	Surgical (n = 142)	Non-surgical (n = 60)	Unknown (n = 44)
Age in years, median (IQR) range	76 (73; 81) 70 – 95	77 (73; 81) 70 – 93	76 (72; 82) 70 – 95	77 (72; 83) 70 – 87
Sex, male	154 (66)	92 (70)	37 (65)	25 (57)
Educational level				
low (< high school)	61 (26)	36 (27)	9 (16)	16 (36)
moderate (high school)	84 (36)	46 (35)	22 (39)	16 (36)
high (college / university)	83 (36)	47 (36)	26 (46)	10 (23)
Medical specialty				
trauma and orthopedics	17 (7)	8 (6)	7 (12)	2 (4)
vascular and hepatobiliary	37 (16)	20 (15)	8 (14)	9 (21)
medical oncology	40 (17)	24 (18)	7 (12)	9 (21)
internal medicine	20 (9)	1 (1)	9 (16)	10 (23)
cardiology	114 (50)	79 (60)	26 (46)	9 (21)
Comorbidity ^a , median (IQR) range	2 (1; 3) 0 – 10	2 (1; 3) 0 – 9	2 (1; 3) 0 – 10	unknown
LoS in days, median (IQR) range	4 (3; 8) 1 – 37	4 (3; 7) 1 – 37	4 (3; 9) 1 – 31	unknown
Fatigability score				
physical, mean (SD)	23.7 (11.5)	21.8 (11.2)	24.6 (11.3)	28.4 (11.4)
mental, mean (SD)	14.9 (13.5)	12.9 (12.1)	14.5 (14.2)	21.4 (14.7)

Notes: All numbers are presented in n (%), unless indicated otherwise.
Abbreviations: IQR, interquartile range; LoS, length of stay; SD, standard deviation.
a. Comorbidity score was assessed using the Charlson Comorbidity Index

Table S4 – 6. Expected and observed correlations of the PFS subscales per subgroup with other constructs.

Instrument (scale range)	Spearman's rank correlation			
	PFS physical score (0-50)		PFS mental score (0-50)	
Non surgical patients (n = 60)	Expected	Observed	Expected	Observed
iADL (0-9)	0.3 – 0.5	0.4	0.3 – 0.5	0.1
Frailty Phenotype (0-5)	0.3 – 0.5	0.5	0.3 – 0.5	0.4
SBT (0-28)	≤ 0.3	0.1	≤ 0.3	0.04
Surgical patients (n = 142)				
iADL (0-9)	0.3 – 0.5	0.4	0.3 – 0.5	0.3
Frailty Phenotype (0-5)	0.3 – 0.5	0.6	0.3 – 0.5	0.4
SBT (0-28)	≤ 0.3	0.1	≤ 0.3	0.1

Abbreviations: iADL, instrumental activities of daily living; PFS, Pittsburgh fatigability scale; SBT, short blessed test.

EFA PFS physical subscale

Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity were performed for both subgroups, to check whether the data were suitable for EFA. KMO of the PFS physical subscale was 0.81 and 0.83 for non-surgical and surgical patients respectively, and Bartlett's test of sphericity was >0.001 for both subgroups, indicating that EFA could be applied.

Inspection of the scree-test and based on eigenvalues above 1 opt for a two-factor solution of both non-surgical and surgical patients with corresponding eigen values (explained variances): 4.72 (47%) plus 2.03 (20%), and 4.86 (49%) plus 1.63 (16%), respectively. Results of the two-factor solutions for both subgroups are presented in Table 3. Among both non-surgical and surgical patients the 2-factor solution resulted in a clear factor structure with all items loadings > 0.5 on one factor, except for the item household activities (item C), which show a crossloading. Theoretically we can distinguish the first factor that includes items that required physical effort to perform (A, B, C, D, G, and J). The second factor included items that require less physical activity (E, F, H, and I). An explanation why household activities load on both factors may be that depending on what type of household activity is thought of by answering; cooking can be considered as a predominantly mental activity, whereas vacuuming and cleaning windows requires physical effort to perform. Cronbach's alphas including household activity in the first factor indicated good internal consistency of the two factors for both subgroups.

EFA PFS mental subscale

KMO of the PFS mental subscale were 0.87 and 0.85 for non-surgical and surgical patients respectively, and Bartlett's test of sphericity was >0.001 for both subgroups, indicating that EFA could be applied.

The scree test and eigenvalues above 1 opt for a one factor solution underlying the PFS physical subscale data among non-surgical patients. Corresponding eigenvalue (6.31) explained 63% of the variance. Surgical patient data opt for a two-factor solution based on eigenvalues above 1 and scree test. Eigen values (% explained variance) were 5.31 (53%) plus 1.26 (13%) for surgical patients. Results of the factor solutions of the PFS mental subscale of non-surgical and surgical patients are presented in Table 4. All factor loadings of non-surgical patients were higher than 0.5 indicating a clean factor structure of the single factor solution. Exploring the two-factor solution for non-surgical patients resulted in an unexplainable cross-loading of item G (moderate to high intensity strength training). Among surgical patients, the two-factor solution resulted in a clean factor structure with six items (A, B, C, D, G, J) loading to the first factor and four items (E, F, H, I) loading to the second factor. All factor loadings were higher than 0.5 and no cross loadings indicating a clean factor structure. Cronbach's alphas indicated good internal consistency of the single factor solution for non-surgical patients and the two factor solution for surgical patients.

Table S4 – 7. EFA results: Two-factor solutions for the Dutch version of the PFS physical subscale of non-surgical and surgical patients (pattern matrix).

		Non-surgical		Surgical	
		Factor 1	Factor 2	Factor 1	Factor 2
		$\alpha = 0.88^a$	$\alpha = 0.83^a$	$\alpha = 0.90^a$	$\alpha = 0.75^a$
Moderate to high intensity activities					
B	Brisk walking	.810	-	.953	-
D	Heavy gardening	.707	-	.749	-
G	Strength training	.781	-	.755	-
J	High intensity activity	.869	-	.757	-
Lifestyle or light intensity activities					
C	Household activity	.443	.460	.449	.370
A	Leisurely walk	.684	-	.784	-
Social activities					
H	Participate in social activity	-	.824	-	.751
I	Hosting social event	-	.784	-	.727
Sedentary activities					
E	Watching TV	-	.649	-	.652
F	Sitting quietly	-	.660	-	.508

Only factor loading >0.3 are presented.
a. Cronbach's alpha's were calculated for items A, B, C, D, G, J (Factor 1) and items E, F, H, I (Factor 2)

Table S4 – 8. EFA results: Two-factor solutions for the Dutch version of the PFS mental subscale of non-surgical and surgical patients (pattern matrix).

		Non-surgical		Surgical	
		Factor 1		Factor 1	Factor 2
		$\alpha = 0.94^a$		$\alpha = 0.91^b$	$\alpha = 0.74^b$
Moderate to high intensity activities					
B	Brisk walking	.852		.965	-
D	Heavy gardening	.848		.705	-
G	Strength training	.819		.653	-
J	High intensity activity	.803		.594	-
Lifestyle or light intensity activities					
C	Household activity	.766		.574	-
A	Leisurely walk	.832		.903	-
Social activities					
H	Participate in social activity	.863		-	.744
I	Hosting social event	.828		-	.757
Sedentary activities					
E	Watching TV	.592		-	.457
F	Sitting quietly	.559		-	.486

Only factor loading >0.3 are presented.
a. Cronbach's alpha was calculated for all ten items.
b. Cronbach's alpha's were calculated for items A, B, C, D, G, J (Factor 1) and items E, F, H, I (Factor 2).

Complete case analysis

Table S4 – 9. Baseline characteristics of total group and subgroups.

Characteristic	Total (n=233)	Complete PFS data (n = 201)
Age in years, median (IQR) range	76 (73; 81) 70 – 95	76 (73; 82) 70 – 95
Sex, male	154 (66)	137 (68)
Educational level		
low (< high school)	61 (26)	51 (25)
moderate (high school)	84 (36)	73 (36)
high (college / university)	83 (36)	72 (36)
Medical specialism		
trauma and orthopedics	17 (7)	15 (8)
vascular and hepatobiliary	37 (16)	31 (15)
medical oncology	40 (17)	29 (14)
internal medicine	20 (9)	19 (10)
cardiology	114 (50)	103 (51)
Comorbidity ^a , median (IQR) range	2 (1; 3) 0 – 10	2 (1; 3) 0 – 10
LoS in days, median (IQR) range	4 (3; 8) 1 – 37	4 (3; 8) 1 – 37
Fatigability score		
physical, mean (SD)	23.7 (11.5)	24.0 (11.5)
mental, mean (SD)	14.9 (13.5)	15.4 (13.6)

Notes: All numbers are presented in n (%), unless indicated otherwise.
Abbreviations: IQR, interquartile range; LoS, length of stay; SD, standard deviation.
a. Comorbidity score was assessed using the Charlson Comorbidity Index

Table S4 – 10. Expected and observed correlations of the PFS subscales of patients with complete PFS data (n = 201).

Instrument (scale range)	Spearman's rank correlation			
	PFS physical score (0-50)		PFS mental score (0-50)	
	Expected	Observed	Expected	Observed
iADL (0-9)	0.3 – 0.5	0.4	0.3 – 0.5	0.3
Frailty Phenotype (0-5)	0.3 – 0.5	0.5	0.3 – 0.5	0.4
SBT (0-28)	≤ 0.3	0.05	≤ 0.3	0.05

Abbreviations: iADL, instrumental activities of daily living; PFS, Pittsburgh fatigability scale; SBT, short blessed test.

Table S4 – 11. EFA results: Pattern matrix of the 2 factor solutions of the PFS physical and mental subscale of the patients with complete PFS data (sensitivity analysis) (n=201)

Factor / Item description	Physical subscale		Mental subscale	
	Factor 1 (α: 0.89) ^a	Factor 2 (α: 0.81) ^a	Factor 1 (α: 0.93) ^a	Factor 2 (α: 0.81) ^a
Moderate to high intensity activities				
B Brisk walking	.923	-	.985	-
D Heavy gardening	.740	-	.801	-
G Strength training	.824	-	.629	-
J High intensity activity	.794	-	.591	-
Lifestyle or light intensity activities				
C Household activity	.368	.515	.520	.337
A Leisurely walk	.640	-	.900	-
Social activities				
H Participate in social activity	-	.799	-	.790
I Hosting social event	-	.730	-	.744
Sedentary activities				
E Watching TV	-	.727	-	.640
F Sitting quietly	-	.601	-	.594

Factors loaded >0.3 are presented

Principal-Axis Factor analysis was used with oblique oblimin rotation

a. Cronbach's Alpha of the PFS physical subscale was calculated using items A, B, C, D, G, J and items E, F, H, I for factors 1 and 2, respectively.

Figure S4 -1. Bland Altman plots of the PFS physical (a) and mental (b) subscales (sensitivity analysis). X-axis represent mean fatigability scores; Y-axis represents difference of fatigability scores (baseline – retest assessments).

Bland Altman plots showed good agreement for the PFS physical subscale (mean difference -1.33, 95% CI: -2.85; 0.21). The PFS mental subscale had a small mean systematic difference between baseline and re-assessment (mean difference: -2.50, 95% CI: -4.81; -0.18). Reliability outcomes are presented in table 4. ICC of the PFS physical and mental subscales were 0.84 (95% CI: 0.73; 0.91) and 0.81 (95% CI: 0.67; 0.89), respectively, indicating good reliability.

Table S4 – 12. Mean scores and standard deviations of baseline and retest assessments (n = 50).

PFS	Baseline mean (SD)	T1 mean (SD)	ICC (95% CI)	SEM	SEM (%) ^a	SDC	SDC (%) ^a
Physical subscale	30 (10)	31 (9)	0.84 (0.73; 0.91)	4	8%	11	22%
Mental subscale	24 (13)	26 (13)	0.81 (0.67; 0.89)	6	12%	17	33%

Abbreviations: PFS, Pittsburgh Fatigability Scale; ICC, intraclass correlation coefficient for agreement using a 2-way mixed effect model; SEM, standard error of measurement; SDC, smallest detectable change; MIC, minimal important change.

a. SEM (%) and SDC (%) are expressed in percentages of the scale range (0-50); percentages are rounded off.

CHAPTER 5

Trajectories of self-rated health in
an older general population and their
determinants: The Lifelines Cohort
Study

5.

Marlies Feenstra
Barbara C van Munster
Janet L MacNeil Vroomen
Sophia E de Rooij
Nynke Smidt

BMJ Open 2020;10:e035012

ABSTRACT

Objectives: Poor self-rated health (SRH) is a strong predictor of premature mortality in older adults. Trajectories of poor SRH are associated with multimorbidity and unhealthy behaviors. Whether trajectories of SRH are associated with deviating physiological markers is unclear. This study identified trajectories of self-rated health (SRH) and investigated the associations of trajectory membership with chronic diseases, health risk behaviors, and physiological markers in community-dwelling older adults.

Study design and setting: Prospective general population cohort

Participants: Trajectories of SRH over 5 years were identified using data of 11 600 participants aged 65 years and older of the Lifelines Cohort Study.

Outcome measures: Trajectories of SRH were the main outcome. Covariates included demographics (age, gender, education), chronic diseases, health-risk behavior (physical activity, smoking, drinking), and physiological markers (BMI, cardiovascular function, lung function, glucose metabolism, hematological condition, endocrine function, renal function, liver function, and cognitive function).

Results: Four stable trajectories were identified, including excellent ($n = 607$, 6%), good ($n = 2111$, 19%), moderate ($n = 7677$, 65%), and poor SRH ($n = 1205$, 10%). Being female (OR: 1.4; 95%CI: 1.0 – 1.9), low education (OR: 2.1; 95%CI: 1.5 – 3.0), one (OR: 10.4; 95%CI: 7.4 – 14.7) or multiple chronic diseases (OR: 37.8; 95%CI: 22.4 – 71.8), smoking (OR: 1.8; 95% CI: 1.0 – 3.2), physical inactivity (OR: 3.1; 95%CI: 1.8 – 5.2), alcohol abstinence (OR: 2.2; 95%CI: 1.4 – 3.2), and deviating physiological markers (OR: 1.5; 95%CI: 1.1 – 2.0) increase the odds for a higher probability of poor SRH trajectory membership compared to excellent SRH trajectory membership.

Conclusion: SRH of community-dwelling older adults is stable over time with the majority (65%) having moderate SRH. Older adults with higher probabilities of poor SRH often have unfavorable health status.

BACKGROUND

Self-rated health (SRH) is known as an inclusive measure of global health and is often used as a supplement to objective clinical measures of physical health, such as presence of disease and disability.^{1,2} In older adults, poor SRH is an independent and strong predictor of premature mortality.^{3,4} However, evidence for factors associated with poor SRH are predominantly cross-sectional and longitudinal evidence is required. Analysis of latent clusters of individuals who follow a similar pattern of SRH over time, so called trajectory analysis, can be used to explore the course of SRH in time within a certain population.⁵ Few studies have studied SRH in community-dwelling older adults by trajectory analysis revealing various numbers of identified trajectories.^{6–8} Distinct trajectories of SRH varied from persistently good^{6,7}, persistently moderate^{7,8}, persistently poor^{6,7}, declining^{6–8} to improving trajectories of SRH.⁶ People in declining SRH trajectories were differentiated at baseline by older age, lower education level, and an increased number of chronic conditions compared to people in consistently good SRH trajectories.^{6–8} However, in these studies, other measures of determinants of health status, such as abnormalities in physiological markers, like blood pressure, thyroid hormone levels, and glycated hemoglobin were not evaluated. Such markers reflect cross-sectional clinical parameters of physiological processes.⁹ Abnormal physiological processes may indicate pre-clinical prodromal phases of underlying diseases which are suggested to play a role in burden of disease expressed by poor SRH evaluations in older adults.^{1,4,10,11} We hypothesize that multi-morbidity, health risk behaviors, and deviations in physiological markers are associated with trajectories that lead to poor SRH.

The aim of this study is to identify classes of self-rated health over five years in community-dwelling older adults and to investigate whether group membership of SRH trajectories is associated with self-reported chronic diseases, health risk behaviors, and physiological markers.

METHODS

Study population

A subsample of the adult Lifelines Cohort Study was used, including participants aged 65 years or older at baseline ($n = 12\,685$). A detailed description of the complete Lifelines cohort profile is described elsewhere.¹²

Primary outcome measure

Repeated measures of self-rated health were assessed at baseline, 1.5 years, 3 years, and 5 years after baseline measurement by means of a self-reported question 'how would you rate your health in general? (excellent, very good, good, fair, poor)'.^{13,14} The single item SRH question with five response options is a valid and reliable measure of general health status in older adults.^{15–17}

Covariates

Demographics included age, sex, and education level (low, less than primary through lower secondary; intermediate, upper secondary through post-secondary, non-tertiary; high, short cycle tertiary and higher).^{18,19} Chronic diseases were categorized (none, one, two or more) based on a participant's baseline report on presence of the most burdensome chronic diseases as forecasted for the next decades, including dementia, myocardial infarction, osteoarthritis, cerebrovascular accident (CVA), diabetes, chronic obstructive pulmonary disease (COPD), cancer, anxiety, and mood disorders.

Health risk behaviors included physical activity (≥ 5 , 2–4, ≤ 2 days/week physically active for at least 30 minutes)²⁰, smoking (never, former, current smoker), alcohol consumption (abstainer, low risk, at risk)²¹. Low risk drinking is defined as no more than three drinks per day for women and men respectively, and no more than seven drinks per week.²²

Physiological markers included: body mass index (BMI) as a marker of body composition^{23,24}; Systolic and diastolic blood pressure was interpreted with total cholesterol and high density lipoprotein (HDL) ratio as a marker of cardiovascular function²³; Forced expired volume in one second (FEV1) and the forced vital capacity (FVC) ratio was used as a marker of

lung function^{25,26}; Glycated hemoglobin (HbA1c) as a marker of glucose metabolism^{23,27}; Total hemoglobin (Hb) as a marker of hematological condition²⁷, Thyroid Stimulating Hormone (TSH) and free thyroxine (fT4) were used as markers of endocrine function^{28–30}; Estimated glomerular filtration rate (eGFR) by using the Cockcroft Gault formula was used as a marker of renal function^{31–33}; Hepatic Steatosis Index (HSI) was used as a marker of liver function^{34,35}; and the mini mental state examination score (MMSE) was used as a marker of cognitive function.^{23,36} A detailed description of physiological markers used and clinical cut-offs are presented in Appendix A Table A1. Based on clinical cut-offs, both individual physiological markers (normal, abnormal values) and a sum score of abnormal physiological markers were used in the analyses (< 3 vs. ≥ 3 abnormal physiological markers).

Statistical analyses

Baseline characteristics of all participants and classified by SRH trajectory groups were expressed in median and interquartile range (IQR) for continuous variables and proportions and percentages for categorical variables. To identify distinct trajectories of self-rated health over five year, latent class analyses were performed by using Group Based Trajectory Modeling (GBTM).³⁷ The trajectory model was built by a stepwise approach:

Step 1: The basic model was build including the four repeated measures of SRH using a censored normal model. Two up to six trajectories were considered after which the optimal number of trajectories was selected using highest Bayesian Information Criterion (BIC)³⁸, and Bayes factor.³⁹ After the optimal number of trajectories was determined, optimal trajectory shape was determined by varying the growth terms. Optimal trajectory shape was evaluated based on 1. the probability of a person belonging to the selected trajectory (> 0.7), 2. the odds of correct classification (> 5.0), 3. close correspondence between the estimate of group membership probability and the proportion of individuals classified to the group, and 4. reasonable narrow confidence intervals for the estimates of group membership probability.⁴⁰ For the latter two no formal criteria for maximum deviation were available.

Step 2: Multivariable multinomial logistic regression analyses were performed to estimate associations between the probability of SRH trajectory group assignment (result of step 1) and covariates. Three theoretical models were investigated. Model 1: chronic diseases and health behaviors; Model 2: model 1 plus physiological markers; Model 3: model 1 plus the sum score of abnormal physiological markers. For all determinants, multicollinearity was checked using Pearson's correlations. Baseline age, sex, and level of education were included in all models. Model selection was based on lowest BIC, and Akaike's Information Criterion (AIC).⁴¹

Step 3: Trajectories of SRH were re-estimated by including the covariates of the selected model out of step 2. This last step allows to evaluate the influence of one covariate on the probability of belonging to each trajectory taking into account the uncertainty of posterior group membership probability that is introduced by trajectory analysis. Wald statistics were applied for testing the differences between covariates across trajectory groups.

Data of participants with missing data of SRH at all time points were excluded from all analyses (n=1085 (9%)). Participants with missing SRH data at three or less time points were handled using maximum likelihood estimation. Maximum likelihood estimation uses all available information from observed data for constructing the likely values for missing data.⁴⁰ From step 2 onwards, participants who had missing data for baseline covariates were excluded from further analyses (n=3010 (26%)). The flow of participants from the initial to the analytic sample is presented in Appendix B Figure B1.

Sensitivity analyses were performed by: 1) rerunning basic trajectory analysis accounting for non-random attrition (dual trajectory modeling), and 2) using a composite score for chronic diseases without anxiety and mood disorders. For all analyses Stata Statistical Software release 14 was used (StataCorp. 2015. College Station, Texas, USA) with the Traj plug-in.^{42,43}

Table 1. Baseline characteristics of all participants aged 65 years and older and categorized by SRH trajectory group.

Characteristic	All n	1. Excellent 602	2. Good 2111	3. Moderate 7677	4. Poor 1205
Demographics					
Age, median (IQR25-75) range (years)	69 (66, 73) 65-95	68 (66, 72) 65-90	69 (66, 72) 65-92	69 (66, 72) 65-93	70 (67, 74) 65-90
Sex, n (%) male	5484 (47)	344 (57)	1161 (55)	3523 (46)	456 (38)
Highest level of education, n (%)					
Low	6563 (57)	301 (50)	1006 (48)	4482 (58)	774 (64)
Intermediate	2037 (18)	107 (18)	407 (19)	1345 (18)	178 (15)
High	2239 (19)	168 (28)	592 (28)	1319 (17)	160 (13)
missing	761 (7)	26 (4)	106 (5)	531 (7)	93 (8)
Health status, n (%)					
Self-rated health					
excellent	645 (6)	373 (62)	246 (12)	26 (<1)	-
very good	2290 (20)	155 (26)	1326 (63)	804 (10)	5 (<1)
good	6358 (55)	4 (<1)	344 (16)	5805 (76)	205 (17)
fair	979 (8)	-	-	275 (4)	704 (58)
poor	20 (<1)	-	-	-	20 (2)
missing	1308 (11)	78 (13)	208 (10)	845 (11)	286 (24)
Chronic diseases (self-reported)					
none	6076 (52)	468 (78)	1386 (66)	3871 (50)	351 (29)
one	3979 (34)	116 (19)	604 (29)	2793 (36)	466 (39)
≥ 2	1545 (13)	24 (4)	121 (6)	1013 (13)	388 (32)
Health behaviors, n (%)					
Physical activity for at least 30 minutes					
≥ 5 days/week	6395 (55)	368 (61)	1330 (63)	4226 (55)	471 (39)
2-4 days/week	2481 (21)	109 (18)	396 (19)	1743 (23)	233 (19)
≤ 1 day/week	761 (7)	27 (5)	93 (4)	512 (7)	129 (11)
missing	1963 (17)	98 (16)	292 (14)	1196 (16)	371 (31)

Characteristic	All	1. Excellent	2. Good	3. Moderate	4. Poor
Smoking status					
never smoker	4453 (38)	238 (40)	802 (38)	2981 (39)	432 (36)
former smoker	5937 (51)	314 (52)	1121 (53)	3890 (51)	612 (51)
current smoker	789 (7)	37 (6)	128 (6)	530 (7)	94 (8)
missing	421 (4)	13 (2)	60 (3)	276 (4)	67 (6)
Alcohol consumption					
abstainer	2123 (18)	78 (13)	258 (12)	1479 (19)	307 (25)
low risk	3931 (34)	198 (33)	742 (35)	2674 (35)	317 (26)
at risk	3958 (34)	238 (40)	863 (41)	2566 (33)	290 (24)
missing	1588 (14)	88 (15)	248 (12)	958 (12)	291 (24)
Physiological markers ^a , n (%)					
BMP in kg / m ²					
<23	1323 (11)	107 (18)	295 (14)	822 (11)	99 (8)
≥ 23 & < 30 ^b	8002 (69)	436 (72)	1560 (74)	5317 (69)	689 (57)
≥ 30	2264 (20)	64 (11)	256 (12)	1533 (20)	411 (34)
Blood pressure in mm Hg					
SBP ≤ 140/ 160 ^c & DBP < 90 ^b	6888 (59)	367 (61)	1271 (60)	4511 (59)	739 (61)
SBP ≤ 140/ 160 ^c & DBP ≥ 90	92 (<1)	3 (<1)	20 (1)	64 (1)	5 (<1)
SBP > 140/ 160 ^c & DBP < 90	3822 (33)	194 (32)	670 (32)	2560 (33)	398 (33)
SBP > 140/ 160 ^c & DBP ≥ 90	774 (7)	42 (7)	145 (7)	528 (7)	59 (5)
CHOL/ HDL ratio					
< 3.5	5561 (48)	310 (51)	1040 (49)	3663 (48)	548 (45)
3.5-4.9 ^b	4540 (39)	220 (37)	820 (39)	3022 (39)	478 (40)
≥ 5	1345 (12)	68 (11)	227 (11)	895 (12)	155 (13)
FEV1/ FVC ratio					
≥ 70 ^b	8860 (76)	473 (79)	1625 (77)	5862 (76)	900 (75)
< 70	2740 (24)	134 (22)	486 (23)	1815 (24)	305 (25)
HbA1C in mmol/ mol (% of total Hb)					
< 48 (< 6.5%) ^b	9208 (79)	523 (87)	1767 (84)	6072 (79)	846 (70)
48-52 (6.5-7%)	424 (4)	7 (1)	43 (2)	288 (4)	86 (7)
53-64 (7-8%)	324 (3)	0 (0)	39 (2)	217 (3)	68 (6)
> 64 (> 8%)	88 (1)	2 (<1)	7 (<1)	57 (1)	22 (2)

Characteristic	All	1. Excellent	2. Good	3. Moderate	4. Poor
Hb in g/ dl (mmol/ L)					
< 121 / 13.7 (< 7.5/ 8.5) ^{a,b}	886 (8)	46 (8)	166 (8)	549 (7)	125 (10)
≥ 121 / 13.7 (≥ 7.5/ 8.5) ^a	10 545 (91)	552 (92)	1921 (91)	7018 (91)	1054 (87)
TSH in mIU/L & fT4 in pmol/L					
TSH<0.5-4.0 & fT4<11-19.5 ^b	2204 (19)	99 (16)	413 (20)	1466 (19)	226 (19)
TSH > 4.0 & fT4 ≥ 11 or <11	427 (4)	24 (4)	61 (3)	292 (4)	50 (4)
TSH < 0.5 & fT4 ≥ 11	81 (1)	6 (1)	8 (<1)	59 (1)	8 (1)
eGFR ^c in ml/min/1.73m ²					
≥ 90 ^b	3809 (33)	179 (30)	622 (29)	2568 (33)	440 (37)
60-89	6577 (57)	375 (62)	1285 (61)	4315 (56)	602 (50)
45-59	898 (8)	40 (7)	166 (8)	594 (8)	98 (8)
< 45	151 (1)	4 (1)	14 (1)	98 (1)	35 (3)
HSI					
≤ 36 ^d	2255 (19)	128 (21)	471 (22)	1486 (19)	170 (14)
> 36	1502 (4)	46 (8)	188 (9)	1031 (13)	237 (20)
MMSE score ^e					
25-30 ^b	10738 (93)	552 (92)	1980 (94)	7178 (94)	1028 (85)
< 25	786 (7)	53 (9)	122 (6)	449 (6)	162 (14)
Sum score physiological markers					
none affected	600 (5)	33 (5)	132 (6)	386 (5)	49 (4)
≤ 2	6606 (57)	369 (61)	1298 (61)	4385 (57)	554 (46)
≥ 3	4394 (38)	202 (33)	670 (32)	2874 (37)	589 (49)

Notes: Percentages may not add up to 100% due to rounding.

^a Missing percentages for all physiological markers were <1% except for FEV1/FVC ratio (31%); TSH and fT4 (75%); and HSI (68%). Blood based markers are reported in the International System of Units (SI) followed by conventional units if used in database. Values marked with \$ are cut offs used to define normal values.

^b Cut-off was adjusted for age.

^c Higher cutoff for SBP was used if participants were aged ≥80.

^d Cut-offs are adjusted for sex, men had higher cut-off.

^e Calculated by the Cockcroft Gault formula using serum creatinin in umol/L, age, weight, and adjusted for sex.

^f Cut-offs are adjusted for level of education.

Abbreviations: BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; CHOL, cholesterol; HDL, high-density lipoprotein; FEV1, forced expiration volume in 1 second; FVC, forced vital capacity; HbA1C, Hemoglobin A1C; Hb, hemoglobin; TSH, thyroid stimulating hormone; fT4, free thyroxine; eGFR, estimated glomeration filtration ratio; HSI, Hepatic Steatosis Index; MMSE, Mini Mental State Examination.

RESULTS

Study population characteristics

Of all 11 600 participants, median age at baseline was 69 years (range 65 to 93), and 47% were male. Of this sample, 34% reported one chronic disease at baseline, 13% reported multimorbidity (≥ 2 chronic diseases), 57% had one or two abnormal physiological markers, and 38% had three or more abnormal physiological markers (Table 1). Over five years of follow-up, 497 people died (4%), and 3721 (32%) were lost to follow-up. The 3010 (26%) participants who were excluded from the analysis in step 2 and 3 due to missing covariates measured at baseline were older, more often female, lower educated, and had relatively less self-reported chronic diseases, but more abnormal values of physiological markers compared to the participants retained in the analysis (completers) (Table 2). One of the reasons for these missing data was that participant with low cognitive abilities (mini mental state examination < 26) had a shorter proxy interview, which was the case in 1261 (42%) of the excluded participants.

Trajectories of SRH over 5 years

Of all evaluated models, four trajectories of SRH with different intercepts, and all slopes close to zero showed the best fit (fit statistics are presented in Appendix C Tables C1 and C2). The four trajectories were identified as excellent, good, moderate, and poor SRH including 607 (5.6%), 2111 (18.8%), 7677 (65.3%), and 1205 (9.6%) participants, respectively (Figure 1; Appendix C Figure C1).

Table 1 presents baseline characteristics of participants in all trajectory groups. People having the highest probability of poor SRH trajectory membership were on average older, more often female, lower educated, more often physically inactive, more often alcohol abstainer, and they had more self-reported chronic diseases compared to people who have highest probabilities of assignment to the excellent, good and moderate SRH trajectories. Concerning objectively measured physiological markers, people having the highest probability of poor SRH trajectory membership had higher BMI, less often high blood pressure, but more often high CHOL/HDL ratio, higher Hb levels, higher HSI index, and they scored lower on cognitive function compared to people with highest probability of

Table 2. Differences between completers and non-completers for baseline variables used in final model.

Characteristic	Completers N = 8590	Non-completers N = 3010	p-value
Demographics			
Age in years, median (IQR 25 - 75) ¹	68 (66 - 72)	69 (67 - 73)	<0.001
Male sex, n (%) ²	4132 (48.1)	1352 (44.9)	0.001
Education, n (%) ²			
low	4955 (57.7)	1608 (53.4)	<0.001
intermediate	1678 (19.5)	359 (11.9)	<0.001
high	1957 (22.8)	282 (9.4)	<0.001
missing percentage	0.5%	26%	
Health status			
Self-rated health, n (%)			
excellent	551 (6.4)	94 (3.1)	<0.001
very good	1982 (23.1)	308 (10.2)	<0.001
good	5274 (61.4)	1084 (36.0)	<0.001
fair	765 (8.9)	214 (7.1)	0.001
poor	16 (0.2)	4 (0.1)	0.129
missing percentage	0%	43%	
Self-reported chronic diseases, n (%) ²			
none	4435 (51.6)	1641 (54.5)	0.003
one	3023 (35.2)	956 (31.8)	0.004
≥ 2	1132 (13.2)	413 (13.7)	0.399
missing percentage	0%	0%	
Health behaviors			
Physical activity for at least 30 minutes, n (%) ²			
≥ 5 days/ week	5732 (66.7)	663 (22.0)	<0.001
2-4 days/ week	2191 (25.5)	290 (9.6)	<0.001
≤ 1 day/ week	667 (7.8)	94 (3.1)	<0.001
missing percentage	0%	65%	
Smoking status, n (%) ²			
never smoker	3349 (39.0)	1104 (36.7)	0.007
former smoker	4628 (53.9)	1309 (43.5)	<0.001
current smoker	613 (7.1)	176 (5.8)	0.007
missing percentage	0%	13%	
Alcohol consumption, n (%) ²			
abstainer	1760 (20.5)	362 (12.0)	<0.001
low-risk alcohol consumption	4224 (49.2)	561 (18.6)	<0.001
at risk alcohol consumption	2606 (30.3)	497 (16.5)	<0.001
missing percentage	0%	43%	
Physiological markers²			
≤ 2 affected	5859 (68.2)	1185 (39.4)	<0.001
≥ 3 affected	2731 (31.8)	1604 (53.3)	<0.001
missing percentage	0%	7%	

¹ Equality of distributions was tested using the Wilcoxon Ranked Sum Test.
² Equality of proportions was tested using the two sample test of proportions.
Abbreviations: n, number of participants; IQR, inter quartile range.

assignment to moderate, good and excellent SRH trajectories. In addition, people with the highest probability for poor SRH trajectory membership had more abnormal values of physiological markers compared to people with highest probability of assignment to moderate, good and excellent SRH trajectories.

Figure 1. Non adjusted trajectories of SRH over five years using 11 600 adults aged 65 years and older of the Lifelines Cohort Study.

Identification of covariates of trajectory membership probability

Table 3 presents the results from multivariate logistic regression analyses on probability of group membership of SRH. Model 2 performed worse compared to model 1 (BIC: -61 942; AIC:1.811). The simplest model with only self-reported covariates (model 1) had lowest BIC (-62 488), but higher AIC (1.807) compared to model 3 that included a sum score of physiological markers as well (BIC:-61 718; AIC: 1.804). However, both models had different sample sizes due to missing values for physiological markers in model 3. Taking into account the exploratory nature of this step in the analysis, type II error (an underfit model) would be more undesirable than type I error (an overfit model). Therefore the covariates included in model 3 were used for the final model (see Table 3, model 3). Final model adjusted for associated covariates

The final trajectory model was modeled by jointly estimating the basic model and the covariates age, sex, educational level, self-reported chronic diseases, physical activity behavior, smoking behavior, alcohol consumption, and the sum score of affected physiological markers as risk factors. The final model assigned 471 (5.5%), 1716 (20.0%), 5637 (65.6%), and 766 (8.9%) people to the excellent, good, moderate, and poor SRH trajectories. The final model including covariates showed best fit statistics of posterior probability of group assignment (Appendix D, Table D1). The basic model overrepresented the proportion of participants with highest probability of poor and moderate SRH trajectory membership, and underrepresented the proportion of people with highest probability of excellent and good trajectory membership, compared to the final model that took into account the effect of covariates (Appendix D, Table D1).

Table 4 presents the odds ratios of each of the evaluated covariates of people with highest probability of poor, moderate, and good SRH trajectory membership using the excellent SRH trajectory as reference category. Increasing number of chronic diseases increased the odds of higher probability of poor SRH trajectory membership relative to the probability of excellent SRH trajectory membership (OR: 10.38; 95% CI: 7.38 - 14.72 for one chronic disease, OR: 37.79; 95% CI 22.35 - 71.75 for two or more chronic diseases). Female gender, low education level, physical inactivity, (former) smoking, alcohol abstinence, and presence of 3 or more abnormal values of physiological markers increased the odds of the probability of poor SRH trajectory membership relative to the probability of excellent SRH trajectory membership (Table 4).

Wald tests implied that all trajectory groups were distinguished by the number of self-reported chronic diseases, alcohol consumption, and the sum score of affected physiological markers (p-values <0.001). However, the results presented in Table 4 should be interpreted with caution as all OR calculations are affected by the covariates that were included in the multinomial model to determine the probability of SRH trajectory membership.

Sensitivity analysis including alteration of the composite measure for multimorbidity without anxiety and depressive disorders did not alter trajectory group sizes, shapes, and odds ratios (results not shown).

Table 3. Regression estimates (relative risk ratios and 95% confidence intervals) of poor SRH relative to excellent SRH from multivariate logistic regression models on SRH trajectory group membership.

Covariate	Excellent SRH	Poor SRH trajectory		
		Model 1 ^a n = 8679	Model 2 ^a n = 8679	Model 3 ^a n = 8590
Age	Ref.	1.01 (0.99; 1.04)	1.02 (0.99; 1.05)	1.01 (0.98; 1.04)
Sex, male	Ref.	Ref.	Ref.	Ref.
female	Ref.	1.44 (1.09; 1.90)	1.66 (1.24; 2.22)	1.46 (1.10; 1.94)
Education low	Ref.	Ref.	Ref.	Ref.
intermediate	Ref.	0.76 (0.55; 1.05)	0.77 (0.56; 1.07)	0.79 (0.57; 1.10)
high	Ref.	0.50 (0.37; 0.68)	0.56 (0.42; 0.77)	0.54 (0.40; 0.74)
Chronic diseases none	Ref.	Ref.	Ref.	Ref.
one	Ref.	7.80 (5.74; 10.61)	7.03 (5.16; 9.57)	7.76 (5.70; 10.58)
≥ 2	Ref.	26.42 (16.12; 43.30)	21.11 (12.80; 34.82)	25.08 (15.28; 41.17)
Physical activity for at least 30 minutes				
≥ 5 days/ week	Ref.	Ref.	Ref.	Ref.
2-4 days/ week	Ref.	1.63 (1.22; 2.18)	1.55 (1.16; 2.08)	1.61 (1.20; 2.15)
≤ 1 day/ week	Ref.	2.82 (1.75; 4.54)	2.55 (1.58; 4.13)	2.85 (1.76; 4.59)
Smoking status never	Ref.	Ref.	Ref.	Ref.
former	Ref.	1.40 (1.07; 1.83)	1.38 (1.05; 1.80)	1.39 (1.06; 1.82)
current	Ref.	1.71 (1.03; 2.85)	1.70 (1.01; 2.84)	1.65 (0.98; 2.78)
Alcohol consumption abstainer	Ref.	Ref.	Ref.	Ref.
low risk	Ref.	0.51 (0.36; 0.71)	0.53 (0.38; 0.75)	0.50 (0.35; 0.71)
at risk	Ref.	0.48 (0.33; 0.69)	0.51 (0.35; 0.74)	0.47 (0.33; 0.69)
Abnormal values of physiological markers ^b				
body composition	Ref.		1.35 (1.03; 1.76)	
cardiovascular function	Ref.		1.36 (1.06; 1.73)	
lung function	Ref.		1.12 (0.84; 1.50)	
glucose metabolism	Ref.		3.77 (1.71; 8.31)	
hematological cond.	Ref.		1.48 (0.95; 2.31)	
endocrine function	Ref.		0.97 (0.53; 1.79)	
renal function	Ref.		0.74 (0.56; 0.97)	
liver function	Ref.		1.78 (1.16; 2.74)	
cognitive function	Ref.		1.53 (1.00; 2.34)	
Sum score of physiological markers				
≤ 2 affected	Ref.			Ref.
≥ 3 affected	Ref.			1.51 (1.16; 1.96)

Notes: Participants with missing data for covariates were excluded from the analyses.

^a Fit statistics: Model 1: AIC: 1.807, BIC: -62488; Model 2: AIC: 1.811, BIC: -61942; Model 3: AIC: 1.804, BIC: -61718.

^b Participants with normal values of the physiological markers were used as the reference category.

Abbreviations: ref, reference category; n, number of participants; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.

Table 4. Odds ratios and 95% confidence intervals per predictor for being member of the good, moderate or poor SRH trajectory group relative to the excellent group (n = 8590a).

Predictor	Odds ratios (95% Confidence Interval)			
	Exc. SRH n = 471	Good SRH n = 1716	Moderate SRH n = 5637	Poor SRH n = 766
Age				
65-69	Ref.	Ref.	Ref.	Ref.
70-74	Ref.	0.99 (0.75; 1.33)	0.93 (0.72; 1.19)	1.03 (0.77; 1.41)
75-79	Ref.	1.38 (0.89; 2.39)	1.33 (0.88; 2.18)	1.34 (0.81; 2.30)
≥ 80	Ref.	1.15 (0.56; 2.59)	1.08 (0.60; 2.31)	1.12 (0.56; 2.78)
Sex				
male	Ref.	Ref.	Ref.	Ref.
female [§]	Ref.	1.03 (0.76; 1.39)	1.21 (0.95; 1.55)	1.43 (1.03; 1.94)
Education				
low	Ref.	Ref.	Ref.	Ref.
intermediate [§]	Ref.	1.10 (0.78; 1.53)	0.87 (0.646; 1.19)	0.76 (0.51; 1.12)
high [§]	Ref.	0.96 (0.73; 1.28)	0.54 (0.41; 0.68)	0.47 (0.33; 0.66)
Chronic diseases				
none	Ref.	Ref.	Ref.	Ref.
one	Ref.	2.11 (1.54; 2.93)	3.55 (2.80; 4.94)	10.38 (7.38; 14.72)
≥ 2	Ref.	1.60 (0.92; 3.30)	5.29 (3.35; 10.52)	37.79 (22.35; 71.75)
Physical activity for at least 30 minutes				
≥ 5 days/ week	Ref.	Ref.	Ref.	Ref.
2-4 days/ week [§]	Ref.	0.99 (0.76; 1.39)	1.35 (1.08; 1.80)	1.61 (1.18; 2.20)
≤ 1 day/ week	Ref.	0.95 (0.54; 1.76)	1.42 (0.90; 2.40)	3.12 (1.76; 5.16)
Smoking status				
never	Ref.	Ref.	Ref.	Ref.
former [#]	Ref.	1.08 (0.82; 1.42)	1.15 (0.91; 1.44)	1.48 (1.11; 1.98)
current [§]	Ref.	1.09 (0.66; 1.95)	1.42 (0.93; 2.30)	1.80 (1.02; 3.16)
Alcohol consumption				
abstainer	Ref.	Ref.	Ref.	Ref.
low risk	Ref.	1.38 (0.93; 2.16)	0.86 (0.62; 1.19)	0.52 (0.35; 0.77)
at risk	Ref.	1.40 (0.97; 2.12)	0.78 (0.57; 1.10)	0.46 (0.31; 0.70)
Sum score of physiological markers				
≤ 2 affected	Ref.	Ref.	Ref.	Ref.
≥ 3 affected	Ref.	0.89 (0.69; 1.21)	1.10 (0.88; 1.45)	1.50 (1.14; 2.03)

Final trajectory model including identified predictors of SRH trajectory membership by multinomial logistic regression analysis (table 2, model 3) adjusted for age (5 year intervals from 65 years old), education, and sex.

^a 3010 of 11,600 participants aged 65 years and older were excluded from the analysis due to missing data on covariates included in the final model.

[§] Wald tests showed no differences between poor and moderate SRH trajectories

[#] Wald tests showed no differences between moderate and good SRH trajectories

Abbreviations: Exc., excellent; Ref., reference category; SRH, self-rated health.

Dual trajectory modeling accounting for non-random attrition showed constant annual attrition probabilities between 10% (good SRH) and 17% (poor SRH) for all trajectory groups (Appendix D, Figure D1). Posterior probability of group assignment did not improve when modeling the trajectories accounting for attrition bias (Appendix D, Table D1).

DISCUSSION

In this sample of an ongoing large cohort study of Dutch community-dwelling older adults, four stable trajectories of SRH over five years were identified. The majority (65.3%) of the participants were classified into the moderate SRH category, followed by good (18.8%), poor (10.2%), and excellent (5.6%) SRH. The results of our study confirmed our a priori hypothesis that the probability of poor SRH trajectory membership was associated with multimorbidity, health risk behaviors, and abnormalities in physiological markers. The number of chronic diseases seems to be one of the key factors that determines someone's probability of SRH trajectory membership, as this was the only covariate under consideration that was significantly associated in all SRH trajectories. In addition, the probability of poor SRH trajectory membership was associated with being female, a low education level, health risk behaviors, and presence of three or more affected physiological markers.

Contrary to previous studies investigating trajectories of SRH, this study identified only stable trajectories of self-rated health of older community-dwelling adults during five years.^{6-8,44} Other studies with comparable measurement intervals, and study duration identified the majority of their participants in the stable trajectories as well, however they also identified small groups with declining and improving trajectories.^{6,8} Sample size was not the limiting factor to identify more groups, however, the posterior diagnostic criteria became worse when adding more than four trajectory groups, indicating four groups was the optimum for our sample. Participants of the current study were older than the populations used in other studies investigating trajectories of SRH. Response shift in SRH is known to occur among older adults.⁴⁵ Compared to their younger counterparts, older adults are suggested to base their SRH more on psychological and life-

style behaviors, and less on functional status and physical health, which might indicate reprioritization response shift.^{46,47} Furthermore, older adults adapt their standards of good health over time, also known as recalibration response shift.⁴⁵ In addition, cognitive strategies to accept negative outcomes, as well as someone's beliefs contribute to enhanced levels of wellbeing, despite negative health outcomes⁴⁸, which can explain the stable trajectories of SRH over time in the present study sample.

Consistent with other studies investigating trajectories of SRH, we found strong associations between increasing numbers of baseline self-reported chronic diseases and the probability of poor SRH trajectory membership.⁶⁻⁸ When participants reported only one chronic disease, they had a two, three-and-half, and ten times higher odds of being a member of the good, moderate, and poor SRH trajectory compared to the probability of excellent SRH trajectory membership, respectively. People suffering two or more self-reported chronic diseases were 38 times more likely for having a higher probability for poor SRH trajectory membership rather than a high probability for excellent SRH trajectory membership. Earlier studies found weaker associations between the probability of poor SRH trajectory membership and the number of chronic diseases.^{7,8} The difference in results might be explained by the different number and combinations of covariates used as predictors in different studies. For instance, previous studies focused on chronic physical health disorders to calculate a composite measure of multimorbidity.^{6,7} For this study, the eleven most burdensome chronic diseases forecasted for the next decades by the Dutch National Institute for Public Health and the Environment were used to measure chronic diseases, which included depression and anxiety disorders. The inclusion of depression and anxiety disorders in our composite measure of chronic diseases may have led to the strong associations between self-rated chronic diseases and the probability of poor SRH trajectory membership in the present study, because depressive symptoms are considered a risk factor for poor SRH.⁴⁹ However, sensitivity analyses excluding depression and anxiety disorders in the composite score for chronic diseases led to similar results. Therefore, it is not expected that the differences in composite measures for chronic diseases explain the differences in magnitude of odds for the probability of poor SRH trajectory membership with increasing number of chronic diseases found in the present study compared to previous studies.

Strengths of this study are the large sample size, and short measurement intervals for SRH that contribute to the robustness of the findings. In addition, the use of physiological markers next to self-reported data was, to the best of our knowledge, not previously investigated in combination with trajectory analyses. There were limitations as well. Firstly, although we found a strong association between self-reported diseases and higher probability of poor SRH trajectory membership, we cannot rule out reverse causation. The presented odds ratios only measure relative change on group level and are not suited to generalize to individual probability of group membership. It is therefore hard to translate these results into concrete clinical implications, as there will always be people having multimorbidity combined with excellent self-rated health. Second, in this older population, the use of self-reported measurements used for measuring the number of chronic diseases may have led to an over- or underestimation of the prevalence of diseases due to non-differential misclassification bias. Finally, attrition may have threatened the generalizability of our results.⁵⁰ However, sensitivity analysis with trajectories jointly modeled with attrition⁵¹ did not improve group allocation probabilities. In addition, constant annual attrition probabilities below 20% for all groups were identified, which led us to conclude that attrition rates were constant among all trajectory groups.

IMPLICATIONS AND CONCLUSIONS

The present study identified four stable trajectories of SRH over five years in Dutch community-dwelling, older adults where the majority of the sample had moderate SRH. Being female, lower levels of education, health risk behaviors (smoking, physical inactivity, and alcohol abstinence), and presence of three or more abnormal physiological markers were associated with higher probability of poor SRH trajectory membership. The identified modifiable determinants may provide a basis for future preventive strategies.

REFERENCES

1. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Soc Sci Med*. 2009;69(3):307–316. doi:10.1016/j.socscimed.2009.05.013
2. Tissue T. Another look at self-rated health among the elderly. *J Gerontol*. 1972;27(1):91–94. doi:10.1093/geronj/27.1.91
3. DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality Prediction with a Single General Self-Rated Health Question A Meta-Analysis. *J Gen Intern Med*. 2005;20:267–275. doi:10.1111/j.1525-1497.2005.0291.x
4. Idler EL, Benyamini Y. Self-Rated Health and Mortality : A Review of Twenty-Seven Community Studies. *J Health Soc Behav*. 1997;38(1):21–37. doi:10.1007/s10290-009-0045-y
5. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol*. 2010;6:109–138. doi:10.1146/annurev.clinpsy.121208.131413
6. Schmitz N, Gariépy G, Smith KJ, et al. Trajectories of self-rated health in people with diabetes: Associations with functioning in a prospective community sample. *PLoS One*. 2013;8(12):1–7. doi:10.1371/journal.pone.0083088
7. Lee HL, Huang HC, Lee M Der, Chen JH, Lin KC. Factors affecting trajectory patterns of self-rated health (SRH) in an older population—A community-based longitudinal study. *Arch Gerontol Geriatr*. 2012;54(3):334–341. doi:10.1016/j.archger.2011.10.009
8. Ayyagari P, Ullrich F, Malmstrom TK, et al. Self-Rated Health Trajectories in the African American Health Cohort. *PLoS One*. 2012;7(12). doi:10.1371/journal.pone.0053278
9. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and Other Tools) Silver Spring: Food and Drug Administration; Bethesda: National Institutes of Health, 2016.
10. Stenholm S, Pentti J, Kawachi I, Westerlund H, Kivimäki M, Vahtera J. Self-rated health in the last 12 years of life compared to matched surviving controls: The health and retirement study. *PLoS One*. 2014;9(9). doi:10.1371/journal.pone.0107879
11. Stenholm S, Kivimäki M, Jylhä M, et al. Trajectories of self-rated health in the last 15 years of life by cause of death. *Eur J Epidemiol*. 2016;31(2):177–185. doi:10.1007/s10654-015-0071-0
12. Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2015;44(4):1172–1180. doi:10.1093/ije/dyu229
13. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473–483.
14. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol*. 1998;51(11):1055–1068.
15. Eriksson I, Undén A-L, Elofsson S. Research on self-rated health measurement scale. *Int J Epidemiol*. 2001;30:326–333.
16. DeSalvo KB, Fisher WP, Tran K, Bloser N, Merrill W, Peabody J. Assessing measurement properties of two single-item general health measures. *Qual Life Res*. 2006;15(2):191–201. doi:10.1007/s11136-005-0887-2
17. Cox B, van Oyen H, Cambois E, et al. The reliability of the Minimum European Health Module. *Int J Public Health*. 2009;54(2):55–60. doi:10.1007/s00038-009-7104-y
18. Unesco Institute for Statistics. ISCED 2011, International Standard Classification of Education. Montreal, Canada; 2012.
19. Divisie Sociale en ruimtelijke statistieken, Sector Sociaal-economisch totaalbeeld, Statistiek CB voor de. SOI 2016, Standaard Onderwijsindeling 2016. Den Haag / Heerlen, The Netherlands; 2017.
20. Kemper HCG, Ooijendijk WTM, Stiggelbout M. Consensus over de Nederlandse norm voor gezond bewegen. *Tijdschr Soc Gezondheidsz*. 2000;78(3):180–183.
21. Gezondheidsraad. Richtlijnen Goede Voeding 2015. Den Haag; 2015.
22. National Institute on Alcohol Abuse and Alcoholism. Rethinking Drinking Alcohol and Your Health. Bethesda MD; 2016. <https://www.rethinkingdrinking.niaaa.nih.gov>.
23. Mathers J, Deary I, Kuh D, et al. Guidelines for biomarkers of healthy ageing. 2015:93. <http://www.mrc.ac.uk/documents/pdf/biomarkers-of-healthy-ageing/>.

24. Winter JE, Macinnis RJ, Wattanapenpaiboon N, Nowson CA. BMI and all-cause mortality in older adults: a meta-analysis. *Am J Clin Nutr.* 2014;99:875-890. doi:10.3945/ajcn.113.068122.1
25. Mannino DM, Diaz-Guzman E. Interpreting lung function data using 80% predicted and fixed thresholds identifies patients at increased risk of mortality. *Chest.* 2012;141(1):73-80. doi:10.1378/chest.11-0797
26. Sorino C, Sherrill D, Guerra S, et al. Prognostic value of FEV1/FEV6 in elderly people. *Clin Physiol Funct Imaging.* 2011;31(2):101-107. doi:10.1111/j.1475-097X.2010.00984.x
27. Fried LP, Xue QL, Cappola AR, et al. Nonlinear multisystem physiological dysregulation associated with frailty in older women: Implications for etiology and treatment. *Journals Gerontol - Ser A Biol Sci Med Sci.* 2009;64(10):1049-1057. doi:10.1093/gerona/glp076
28. Rodondi N, Elzen WPJ Den, Bauer DC, et al. CLINICIAN ' S CORNER Subclinical Hypothyroidism and the Risk of Coronary Heart Disease and Mortality. *JAMA.* 2010;304(12):1365-1374. doi:10.1001/jama.2010.1361
29. Pearce SHS, Razvi S, Yadegarfar ME, et al. Serum Thyroid Function, Mortality and Disability in Advanced Old Age: The Newcastle 85+ Study. *J Clin Endocrinol Metab.* 2016;101(11):4385-4394. doi:10.1210/jc.2016-1935
30. Mammen JS, McGready J, Ladenson PW, Simonsick EM. Unstable Thyroid Function in Older Adults Is Caused by Alterations in Both Thyroid and Pituitary Physiology and Is Associated with Increased Mortality. *Thyroid.* 2017;27(11):thy.2017.0211. doi:10.1089/thy.2017.0211
31. Montesanto A, De Rango F, Berardelli M, et al. Glomerular filtration rate in the elderly and in the oldest old: Correlation with frailty and mortality. *Age (Omaha).* 2014;36(3):1503-1514. doi:10.1007/s11357-014-9641-4
32. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009;150(9):604-612. doi:10.7326/0003-4819-150-9-200905050-00006
33. Levey AS, Stevens LA. Estimating GFR using the CKD Epidemiology Collaboration (CKD-EPI) creatinine equation: more accurate GFR estimates, lower CKD prevalence estimates, and better risk predictions. *Am J Kidney Dis.* 2010;55(4):622-627. doi:10.1053/j.ajkd.2010.02.337
34. Kahl S, Strassburger K, Nowotny B, et al. Comparison of Liver Fat Indices for the Diagnosis of Hepatic Steatosis and Insulin Resistance. *PLoS One.* 2014;9(4):e94059. doi:10.1371/journal.pone.0094059
35. Lee J, Kim D, Jung H, et al. Hepatic steatosis index : A simple screening tool reflecting nonalcoholic fatty liver disease. *Dig Liver Dis.* 2010;42(7):504-509. doi:10.1016/j.dld.2009.08.002
36. Schmand B, Lindeboom J, Hooijer C, Jonker C. Relation between education and dementia : the role of test bias revisited. *J Neurol Neurosurg psychiatry.* 1995;59:170-174.
37. Nagin DS, Jones BL, Lima Passos V, Tremblay RE. Group-based multi-trajectory modeling. *Stat Methods Med Res.* 2016:096228021667308. doi:10.1177/0962280216673085
38. Schwarz G. Estimating the dimension of a model. *Ann Stat.* 1978;6(2):461-464.
39. Wasserman L. Bayesian Model Selection and Model Averaging. *J Math Psychol.* 2000;44:92-107.
40. Nagin D. Group-Based Modeling of Development. Cambridge, Massachusetts: Harvard University Press; 2005.
41. Dziak JJ, Coffman DL, Lanza ST. Sensitivity and Specificity of Information Criteria. *Brief Bioinform* 2019;6:1-13.
42. Jones BL, Nagin DS. A Stata Plugin for Estimating Group-Based Trajectory Models. Pittsburgh; 2012.
43. Jones BL, Nagin DS. A Note on a Stata Plugin for Estimating Group-based Trajectory Models. *Sociol Methods Res.* 2013;42(4):608-613. doi:10.1177/0049124113503141
44. Sacker A, Wiggins RD, Bartley M, McDonough P. Self-rated health trajectories in the United States and the United Kingdom: A comparative study. *Am J Public Health.* 2007;97(5):812-818. doi:10.2105/AJPH.2006.092320
45. Spuling SM, Wolff JK, Wurm S. Response shift in self-rated health after serious health events in old age. *Soc Sci Med.* 2017;192:85-93. doi:10.1016/j.socscimed.2017.09.026
46. Hoeymans N, Feskens EJM, Kromhout D, van den Bos GAM. Aging And Estimating

- The Relationship Between Functional Status And Self-Rated Health In Elderly Men. *Soc Sci Med.* 1997;45(10):1527-1536.
47. Leinonen R, Heikkinen E, Jylhä M. Self-rated health and self-assessed change in health in elderly men and women - A five-year longitudinal study. *Soc Sci Med.* 1998;46(4-5):591-597. doi:10.1016/S0277-9536(97)00205-0
 48. De Ridder D, Geenen R, Kuijer R, van Middendorp H. Psychological adjustment to chronic disease. *Lancet.* 2008;372(9634):246-255. doi:10.1016/S0140-6736(08)61078-8
 49. Han B. Depressive Symptoms and Self-Rated Health in Community-Dwelling Older Adults: A Longitudinal Study. *J Am Geriatr Soc.* 2002;50:1549-1556.
 50. Genbäck M, Ng N, Stanghellini E, Luna X De. Predictors of decline in self - reported health : addressing non - ignorable dropout in longitudinal studies of aging. *Eur J Ageing.* 2018;15(2):211-220. doi:10.1007/s10433-017-0448-x
 51. Haviland AM, Jones BL, Nagin DS. Group-based Trajectory Modeling Extended to Account for Nonrandom Participant Attrition. *Sociol Methods Res.* 2011;40(2):367-390. doi:10.1177/0049124111400041

SUPPLEMENTARY FILES – CHAPTER 5

APPENDIX A. Biomarker specification

Table A1: Cut-offs used to define normal and affected values for biomarkers per organ system.

Lung function		
FEV1 (L) / FVC (L) ratio (multiplied by 100%)	Cut-off scores used (Quanjer et al., 2012): Normal: ≥ 70% Affected < 70%	
Renal function		
eGFR (in mL/min/1.73m2)	Estimated with the Cockcroft Gault formula using serum creatine in µmol/L (adjusted for age, sex, weight) (Cockcroft & Gault, 1976). Cut-off scores used (Traynor, Mactier, Geddes, & Fox, 2006): Normal: ≥90 mL/min/1.73m² Affected: <90 mL/min/1.73m²	
Endocrine function¹		
TSH (mIU/L)	Normscores (lab standards UMCG): Low: <0.5 mIU/L Normal: 0.5 – 4.0 mIU/L High: ≥ 4.0 mIU/L	
fT4 (pmol/L)	Normscores (Boesten et al., 2012): Low: < 11.0 pmol/L Normal: 11.0 – 19.5 pmol/L High: > 19.5 pmol/L	
Immune function		
Hb (mmol/L)	Different cut-offs used for men and women (lab standards UMCG). Cut off used: Men: Normal: ≥8.5 and ≤11 mmol/L Affected: <8.5, >11 mmol/L Women: Normal: ≥7.5 and ≤10 mmol/L Affected: <7.5, >10 mmol/L	
Liver function		
Hepatic Steatosis Index	Cut off used (Lee et al., 2010; Meerns et al., 2015): Normal: ≤ 36 Affected >36	
Cognitive function		
MMSE	Adjusted for level of education: primary education or less (max.6 years) and secondary or higher (>6 years) (Schmand, Lindeboom, Hooijer, & Jonker, 1995). Cut-off used: ≤ Primary: Normal: ≥25 Affected: <25 ≥ Secondary: Normal ≥27 Affected: <27	
Body composition		
BMI (for Caucasian)	Age adjusted BMI cutoffs were used (Winter, Macinnis, Wattanapenpaiboon, & Nowson, 2014). Cut offs used: Normal: ≥23.0 BMI <30 Affected: <23 & BMI ≥30	
Cardiovascular function²		
SBP (mmHg)	Adjusted for age ((European Society of Hypertension/ European Society of Cardiology, 2014)) Cut-offs used: Aged <80 Normal: ≤ 140 mmHg High: >140 mmHg Aged ≥80: Normal: ≤ 160 mmHg High: >160 mmHg	
DBP (mmHg)	Cut-off used (European Society of Hypertension/ European Society of Cardiology, 2014): Normal: <90 mmHg High: ≥90 mmHg	
Total cholesterol (mmol/L)/ HDL (mmol/L) ratio	Cut offs used (European Society of Cardiology / European Atherosclerosis Society, 2016; Landelijke werkgroep Cardiovasculair risicomanagement, 2012): Normal: <5.0 High ≥ 5.0	
Glucose metabolism		
HbA1c (mmol (HbA1c) / mol (Hb))	Cut offs used (Fried et al., 2009): Normal: < 48 mmol/mol (corresponding to 6,5% of total Hb) Affected: ≥ 48 mmol/mol	

Abbreviations: FER, forced expiratory ratio; eGFR, estimated glomerular filtration rate; TSH, thyroid stimulating hormone; FSH, follicle-stimulating hormone; HB, hemoglobin; MMSE, mini mental state examination; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; HDL, high density lipoprotein; HbA1C, glycated hemoglobin.

1. TSH cut-offs were interpreted with FT4; both TSH and FT4 should be in the normal range to score 'normal' endocrine function.
2. Blood pressure was interpreted with cholesterol levels; both diastolic and systolic blood pressure and cholesterol/HDL ratio should be in the normal range to score 'normal' cardiovascular function.

References:

Boesten, L. S. M., Brugts, M. P., & van Rossum, A. P. (2012). Schildklierdiagnostiek: discrepantie FT4 en TSH. *Nederlands Tijdschrift Voor Geneeskunde*, 156, A4167V.

Cockcroft, D. W., & Gault, H. (1976). Prediction of Creatinine Clearance from Serum Creatinine. *Nephron*, 16(1), 31–41. European Society of Cardiology / European Atherosclerosis Society. (2016). 2016 ESC/EAS guidelines for the management of dyslipidemias. *European Heart Journal*, 37, 2999–3058.

European Society of Hypertension/ European Society of Cardiology. (2014). 2014 guidelines for the management of arterial hypertension. *Primary Care Cardiovascular Journal*, 7(2), 85–88.

Fried, L. P., Xue, Q. L., Cappola, A. R., Ferrucci, L., Chaves, P., Varadhan, R., ... Bandeen-Roche, K. (2009). Nonlinear multisystem physiological dysregulation associated with frailty in older women: Implications for etiology and treatment. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 64(10), 1049–1057.

Landelijke werkgroep Cardiovasculair risicomanagement. (2012). Dutch standard of cardiovascular riskmanagement. *Huisarts Wet*, 55(1), 14–28.

Lee, J., Kim, D., Jung, H., Lee, C., In, J., Kim, W., ... Lee, H. (2010). Hepatic steatosis index: A simple screening tool reflecting nonalcoholic fatty liver disease. *Digestive and Liver Disease*, 42(7), 504–509.

Meems, L. M. G., De Borst, M. H., Postma, D. S., Vonk, J. M., Kremer, H. P. H., Schuttelaar, M. L. A., ... De Boer, R. A. (2015). Low levels of Vitamin D are associated with multimorbidity: Results from the LifeLines Cohort Study. *Annals of Medicine*, 47(6), 474–481.

Quanjer, P. H., Stanojevic, S., Cole, T. J., Baur, X., Hall, G. L., Culver, B. H., ... Schindler, C. (2012). Multi-ethnic reference values for spirometry for the 3–95-yr age range: The global lung function 2012 equations. *European Respiratory Journal*, 40(6), 1324–1343.

Schmand, B., Lindeboom, J., Hooijer, C., & Jonker, C. (1995). Relation between education and dementia: the role of test bias revisited. *Journal of Neurology, Neurosurgery and Psychiatry*, 59, 170–174.

Traynor, J., Mactier, R., Geddes, C. C., & Fox, J. G. (2006). How to measure renal function in clinical practice: Estimated glomerular filtration rate in general practice. *Bmj*, 333(7574), 918.2.

Winter, J. E., Macinnis, R. J., Wattanapenpaiboon, N., & Nowson, C. A. (2014). BMI and all-cause mortality in older adults: a meta-analysis. *Am J Clin Nutr*, 99, 875–890.

APPENDIX B. Flowchart of study sample

Figure B1. Flow of selection of study sample.

APPENDIX C. Basic model specifications

Figure C1. Basic mean trajectory groups of SRH (a), and observed individual trajectories per trajectory group (b-e) over five years of 11.600 people aged 65 years and older of the Lifelines Cohort.

a. Dots represent the mean observed value per measurement moment; solid lines represent fit lines; dotted lines represent 95% confidence intervals of the fit lines.

b-e. Jittering was used for adding random noise to make all individual scores integer to avoid overlap of individual trajectories for people with identical trajectories.

Table C1. Bayesian Information Criteria and probability estimation Jeffreys's scale of evidence for Bayes factors of crude trajectory calculations with fixed quadratic growth terms used to select adequate number of groups.

No. of groups	BIC (n=11,600)	Probability of correct model
2	-35016.79	0
3	-33001.46	0
4	-32240.77	0.98
5	-32259.49	< 0.01
6	-32278.20	< 0.01
7	-32244.54	0.02

Table C2. Posterior diagnostics of model performance of basic trajectory model.

Group	Model estimate (π^A)	95% CI	Proportion classified (p^A)	Ave. PP	Odds correct classification
1	.056	(.050; .062)	.052	.892	138.3
2	.653	(.642; .664)	.662	.941	8.4
3	.102	(.095; .109)	.104	.852	50.7
4	.188	(.179; .198)	.182	.863	27.1

APPENDIX D. Sensitivity analysis

Figure D1. Trajectories of SRH jointly modelled with attrition.

Table D1. Comparison of posterior probability of assignment for the basic model, the model including covariates / risk factors, and the trajectory model that jointly modelled attrition (sensitivity analysis).

Group Allocation	N (%)	Excellent	Good	Moderate	Poor
Basic model (step 1): posterior probability of assignment					
Excellent	607 (5.6)	0.89	<0.01	0.01	0.05
Good	2111 (18.8)	0.11	0.86	0.06	<0.01
Moderate	8762 (65.3)	<0.01	0.15	0.91	0.08
Poor	1205 (10.2)	<0.01	<0.01	0.03	0.85
Model with covariates (step 3): posterior probability of assignment					
Excellent	471 (5.5)	0.91	0.05	<0.01	<0.01
Good	1716 (20.0)	0.09	0.87	0.04	<0.01
Moderate	5637 (65.6)	<0.01	0.08	0.95	0.09
Poor	766 (8.9)	<0.01	<0.01	0.02	0.91
Model with attrition (sensitivity analysis): posterior probability of assignment					
Excellent	609 (5.7)	0.90	0.05	0.01	<0.01
Good	2123 (18.7)	0.10	0.86	0.05	<0.01
Moderate	8762 (65.3)	<0.01	0.09	0.91	0.14
Poor	1191 (10.3)	<0.01	<0.01	0.04	0.86

*Rows may add to more than 1.0 due to rounding.

CHAPTER 6

Determinants of trajectories of fatigability and mobility among older medical patients during and after hospitalization, an explorative study

6.

Marlies Feenstra
Barbara C van Munster
Nynke Smidt
Sophia E de Rooij

ABSTRACT

Background: Fatigability is an important marker of functional decline in community-dwelling older people, yet its relationship with functional decline after hospitalization is unclear. The objectives of this study were to identify trajectories of fatigability and mobility over time and to examine the association between demographic and clinical characteristics and these trajectories in medical patients aged 70 years and older admitted to a Dutch tertiary care teaching hospital.

Methods: In this prospective cohort study with baseline (in-hospital), discharge, three-, and six-months post discharge follow-up measurements, fatigability was assessed by the physical subscale of the Pittsburgh Fatigability Scale (PFS). Mobility was assessed by the De Morton Mobility Index (DEMMI). Group-based trajectory modeling was used to identify joint trajectories of fatigability and mobility. Covariates included demographic (age, gender, living situation, education) and clinical characteristics (functional status, frailty status, depression, comorbidity, length of hospital stay).

Results: Among 44 patients, three distinct fatigability trajectories and two mobility trajectories were identified over the course from hospital admission up to six months after discharge. Subsequently, three joint trajectories were identified, including low fatigability and high mobility (11%), improving fatigability and high mobility (52%), and high fatigability and low mobility (36%). Controlling for baseline functional status, patients with a lower comorbidity score (OR: 0.27, 95%CI: 0.10; 0.74) and higher frailty status (OR: 1.36, 95%CI: 1.07; 1.74) were more likely to be a member of the high fatigability and low mobility trajectories.

Conclusions: From hospital admission up to six months after discharge, three distinct trajectories of fatigability and mobility were identified among older medical patients. Our results should be interpreted with caution due to the small sample size, but may inspire other researchers to determine the value of fatigability assessment in identifying older medical patients at risk for developing mobility problems.

BACKGROUND

Fatigue is the most prevalent symptom among older hospitalized patients, with a prevalence rate of around 70%.^{1,2} After hospital discharge, many older people still experienced fatigue, often leading to disruption of activities of daily living (ADL).³ Hence, it can be assumed that older hospitalized patients experience high prevalence of fatigability as well, which is defined as perceptions of fatigue while performing activities of a certain intensity and duration.⁴ Over the past decade, fatigability has emerged as an important marker of functional decline in community-dwelling older people without disabilities⁵⁻⁷, yet its relationship with functional decline after hospitalization is unclear.

Functional decline is a serious consequence of hospitalization among older patients, with a prevalence rate of around 30% at the moment of discharge.⁸ Recovery rates of functional decline varied from 30% to 70% at three months after discharge from the hospital.⁹⁻¹² Low mobility during hospital stay is an important predictor of adverse outcomes after hospitalization such as functional decline, re-hospitalization, institutionalization, and death.^{13,14} In addition, loss of mobility and exhaustion or fatigue are considered as early manifestations of frailty, contributing to functional decline and dependence in older people.^{15,16} The association between mobility and fatigability among hospitalized older patients has not been directly investigated yet, but it may provide interesting information to design targeting interventions to prevent adverse outcomes after hospitalization of older patients.

Therefore, this study aimed to investigate whether trajectories of fatigability are associated with trajectories of mobility from hospital admission up to six months post-discharge in medical patients aged 70 years and older. A second aim is to investigate the determinants of these combined trajectories.

METHODS

Participants

Patients admitted to medical wards of a tertiary care teaching hospital in the Netherlands were recruited for participation within the first four days of hospital admission. Patients younger than 70 years, those who did not speak Dutch, those who were cognitively impaired due to dementia or delirium, and those who did not provide informed consent were excluded from participation.

Procedure

Every weekday between April 2018 and July 2019, all consecutive patients admitted to the hospital were screened for eligibility by research staff. The research staff introduced the study to eligible patients and handed them written information about the study procedure. The next day patients were visited again to ask for their decision to participate. All patients signed informed consent before enrollment. The study was approved by the local research ethical committee (registration number: 2017.667).

Data collection

Baseline assessment was conducted within the first four days of hospital admission by trained research staff. Data collection consisted of comprehensive questionnaire and physical assessments. Follow-up assessments were conducted within 24 hours before discharge (only when discharge was at least three days after baseline assessment), three months after discharge (home visit), and six months after discharge (home visit).

Demographic variables included age (in years), gender (male, female), living situation (living alone, co-habiting) and educational level (≤ 12 years; >12 years of education).

Clinical variables included (1) Frailty status measured by the Frailty Index including 34 health-related items such as psychosocial, cognitive, and general functioning, and geriatric symptoms.¹⁷ Item scores were rescaled into values between 0 and 1, with higher scores representing worse functioning. The total score represents the proportion of present

items as a proportion of all items measured, resulting in a Frailty Index score ranging from 0 to 1. A detailed item description is presented in Supplementary Table S1; (2) Pre-hospital functional status (re-call based) measured by the 15-item Katz ADL Index score including basic and instrumental ADL¹⁸; (3) Depressive symptoms were screened by using the two questions 'During the past month, have you often been bothered by feeling down, depressed or hopeless?' and 'During the past month, have you often been bothered by little interest or pleasure in doing things?'.¹⁹ If one or two of these screening questions were answered with 'yes', the 15-item GDS was subsequently assessed using a cut-off of five items for depression^{20,21}; (4) Chronic conditions were measured by the Charlson Comorbidity Index (CCI)²²; and (5) Length of hospital stay was obtained from the hospital medical records after hospital discharge.

Fatigability was measured by the Pittsburgh Fatigability Scale (PFS).²³ The PFS was originally developed in the United States as a self-administered scale consisting of 10-items on which respondents indicated the level of perceived physical and mental fatigue that they expect to perceive immediately after performing the activity on a scale from 0 (no fatigue at all) to 5 (extreme fatigue). Activities include physical activities, household activities, sedentary activities, and social activities. Total physical and mental fatigability scores range from 0–50, with higher scores indicating higher fatigability. The PFS was recently translated into Dutch and validated in older hospitalized patients, showing good internal consistency (range α : 0.80 – 0.92), and test-retest reliability (ICC: 0.80–0.81).²⁴ For this study, only the PFS physical score was used.

Mobility was measured by the 'De Morton Mobility Index' (DEMMI).²⁵ The DEMMI is a physical performance test designed in Australia for older medical patients consisting of 15 mobility tasks that should be performed by the patient while lying (3 items), sitting (3 items), standing (4 items) and moving (5 items).²⁵ A trained research assistant rated for each item whether the patient was able to perform the activity without help, with minimal assistance, or not at all. The total score ranged from 0 (no mobility) to 100 (full mobility). The Dutch version of the DEMMI was successfully examined for inter-rater reliability (0.85) and construct validity with other physical performance tests (ρ ranging from 0.73 to 0.74) among older hospitalized patients.²⁶

Statistical analysis

First descriptive statistics including demographic and clinical characteristics of the total study sample were calculated. Next, group-based trajectory modeling (GBTM) was applied to identify trajectories of fatigability and mobility. Here, a stepwise approach was used:²⁷

Step 1, separate models: separate fatigability and mobility trajectories were plotted using a censored normal model with fixed cubic growth terms for both models. The number of trajectory groups was selected using Bayesian Information Criterion (BIC).²⁸ Second, higher or lower order growth terms were added, and optimal trajectory shape was determined based on posterior diagnostic criteria such as 95% confidence intervals, odds of correct classification, and posterior probability of assignment. These posterior diagnostic criteria reflect the probability of a person belonging to the selected trajectory and are explained in detail elsewhere.²⁸

Step 2, dual models: Next, the trajectories of fatigability and mobility obtained in the first step were simultaneously modeled. This so called dual trajectory model linked the trajectories of fatigability with the trajectories of mobility by estimating the conditional probabilities of membership in the trajectory groups of fatigability, given membership in one of the specific trajectory groups of mobility.

Step 3, adding covariables: Finally multinomial logistic regression analyses were done to explore the associations between the identified dual trajectories and demographic and clinical co-variables. Age, sex, educational level, living situation, functional status, frailty status, depression, and length of hospital stay were alternately (univariably) added as independent variable using the estimated dual trajectory groups as dependent outcome. Given the complex interactions between fatigability, immobilization, disability, frailty, and comorbidity^{15, 16, 29}, a multivariable model was built including these variables to explore potential confounding effects. For these variables, presence of multicollinearity was checked using Pearson's correlations. Odds ratios and 95% confidence intervals of each co-variable per trajectory group were presented.

Subjects with missing data at all time points were excluded from the

analysis. In the trajectory analysis, missing data for some time points were imputed using maximum likelihood estimation. Items of the PFS were imputed using item mean imputation when a. the question whether the activity had been done in the past month was present, and b. maximum three items of the PFS physical subscale were missing.³⁰ For all analyses Stata Statistical Software release 14 was used (StataCorp. 2015. College Station, Texas, USA) using the Traj plug-in.³¹

RESULTS

Characteristics of study participants

Of the 843 eligible hospitalized medical patients aged 70 years and older, 46 gave written consent to participate in our study. Two patients were excluded from the analyses due to withdrawn consent during baseline assessment, leaving 44 patients in the analytic sample. A flowchart of study participants is presented in Figure 1. Demographic and clinical characteristics of study participants are presented in Table 1. Among all participants, sex was equally distributed, median age was 75 years ranging from 70 to 88 years, and the median length of hospitalization was five days (IQR: 3; 8). The majority of the study population lived together (70%) and had one or more disabilities in performing ADL before hospital admission (60%). The minority of the study population had less than twelve years of education (42%) and had five or more depressive symptoms at baseline assessment (18%). The median comorbidity index score of the total group was 2 (IQR: 1; 3) and a detailed description comorbidity distribution is provided in Supplementary Table S2.

Baseline characteristics of patients with missing data on the PFS physical subscale did not differ from patients with complete data (Supplementary Table S3), nor did we identify missing assessments of only the most vulnerable patients. Consequently, these types of missing data were assumed to be missing completely at random. Therefore the data of five patients (11%) with less than three missing items of the PFS physical subscale were imputed, and maximum likelihood estimation was applied in the trajectory analyses when one or more assessments were missing.

Separate fatigability and mobility trajectory models

Among all evaluated trajectory models, a 3-group trajectory solution was selected as the optimal model for fatigability and a 2-group trajectory solution was selected as the optimal model for mobility. The posterior

diagnostic criteria of the selected models were good, indicated by the margins of error <1%, the odds of correct classification >5, and the posterior probabilities of assignment ranging from 0.85 to 0.94 for the fatigability trajectories and from 0.95 to 0.96 for the mobility trajectories.

Figure 1. Flowchart of study participants.

Table 1. Baseline characteristics of all participants (n=44) and according to the joint trajectories of fatigability and mobility.

Baseline characteristic	Total	Low fatigability high mobility	Trajectory group [*]	
			Improving fatigability high mobility	High fatigability low mobility
N (%)	44 (100)	5 (11)	23 (52)	16 (36)
Demographic characteristics				
Age in years, median (IQR)	75 (73; 81)	75 (74; 81)	75 (71; 79)	79 (73; 83)
Female gender	22 (50)	2 (9)	10 (45)	10 (45)
<12 years of education	18 (42)	1 (6)	7 (39)	10 (56)
Missing	1 (3)			
Living alone	13 (30)	2 (15)	5 (38)	6 (46)
Clinical characteristics				
Frailty [†] , median (IQR)	0.26 (0.12; 0.34)	0.12 (0.04; 0.20)	0.15 (0.10; 0.32)	0.35 (0.29; 0.49)
≥1 disabilities	29 (66)	1 (3)	13 (45)	15 (52)
Missing	1 (3)			
≥1 disabilities basic ADL	18 (41)	0 (-)	8 (44)	10 (56)
≥1 disabilities iADL	26 (60)	1 (4)	12 (46)	13 (50)
Missing	1 (3)			
Depressive symptoms [‡]	8 (18)	1 (13)	2 (25)	5 (63)
Comorbidity [§] , median (IQR)	2 (1; 3)	3 (2; 5)	2 (1; 3)	2 (1; 3)
Missing	2 (5)			
LoS in days, median (IQR)	5 (3; 8)	4 (2; 5)	4 (3; 7)	7 (5; 9)
Fatigability , mean (SD)	31 (9)	14 (3)	29 (7)	37 (5)
Missing	9 (20)			
Mobility [¶] , mean (SD)	65 (23)	76 (24)	78 (13)	44 (18)
Missing	5 (11)			

Results are presented as n (%) unless indicated otherwise. Percentages may not add up to 100 due to rounding. The number of missing values at baseline are presented, unless there were no missing values.

^{*} Trajectory groups were estimated by jointly modeling fatigability and mobility over four waves from hospital admission up to six months after discharge.

[†] Frailty was assessed by the Frailty Index.

[‡] Depressive symptoms were positive when the Geriatric Depression Score was ≥5.

[§] Comorbidity was assessed by the Charlson Comorbidity Index.

^{||} Fatigability was assessed by the physical subscale of the Pittsburgh Fatigability Scale.

[¶] Mobility was assessed by the De Morton Mobility Index.

Abbreviations: ADL, activities of daily living; IQR, interquartile range; LoS, length of stay.

Fatigability trajectories

Using repeated measures of four waves of the PFS physical subscale, three fatigability trajectories were identified: stable low fatigability (n: 5, 13%), improving fatigability (n: 19, 43%), and stable high fatigability (n: 20, 45%).

Mobility trajectories

Using repeated measures of four waves of the DEMMI, two mobility trajectories were identified: stable low mobility (n: 15, 34.2%) and stable high mobility (n: 29, 65.8%).

Dual trajectory model of fatigability and mobility

Table 2 represents the posterior probabilities of each fatigability trajectory membership given the mobility trajectory (A) and the posterior

Table 2. Comparison of posterior probability of assignment for the single mobility model, the single fatigability model and the trajectory model that jointly modelled mobility and fatigability.

A. Fatigability given mobility trajectory: posterior probability of assignment				
		Fatigability trajectory groups		
		Stable low (n=5)	Improving (n=19)	Stable high (n=20)
Mobility trajectory groups				
Low (n=15)		0.05	0.00	0.95
High (n=29)		0.16	0.66	0.17
B. Mobility given fatigability trajectory: posterior probability of assignment				
		Mobility trajectory groups		
		Low (n=15)	High (n=29)	
Fatigability trajectory groups				
Stable low (n=5)		0.15	0.86	
Improving (n=19)		0.00	1.00	
Stable high (n=20)		0.74	0.26	
C. Joint fatigability mobility model: posterior probability of assignment				
Group Allocation	n (%)	Low fatigability high mobility	Improving fatigability high mobility	High fatigability low mobility
Low fatigability high mobility	5 (11)	0.79	0.03	<0.001
Improving fatigability high mobility	23(52)	0.06	0.79	0.01
High fatigability low mobility	16 (36)	<0.001	0.01	0.86

Figure 2. Joint trajectories of fatigability and mobility. Fatigability was assessed using the Pittsburgh Fatigability Score physical subscale (PFS, score range: 0 - 50). Mobility was assessed using the De Morton Mobility Index (DEMMI, score range 0 - 100). Higher scores represent higher fatigability and mobility. Dots represent mean predicted PFS and DEMMI scores. Dashed lines represent 95% confidence intervals mean predicted PFS and DEMMI scores.

probabilities of each mobility trajectory given the fatigability trajectories (B). Here we see that the fatigability trajectory largely determined the mobility trajectory someone is allocated to: 86% of the patients who were assigned to the stable low fatigability trajectory were assigned to the high mobility trajectory as well ($p<0.001$); all of the patients (100%) assigned to the improving fatigability trajectory were assigned to the high mobility trajectory too ($p<0.001$); and, 74% of the patients who were assigned to the high fatigability trajectory were assigned to the low mobility trajectory ($p<0.001$). Figure 2 presents these three trajectory combinations resulted from jointly modeling the fatigability and mobility trajectories: 1. Low fatigability and high mobility (11%); 2. Improving fatigability and high mobility (52%), and 3. High fatigability and low mobility (36%). Baseline characteristics of the study participants by

these trajectory combinations are presented in Table 1. Table 2–C presents the posterior probabilities of assignment to each of these dual trajectory groups.

Associations between dual trajectory and covariables

The associations between baseline demographic and clinical characteristics and dual trajectory group membership probabilities using the most favorable trajectory group as the reference group are presented in Table 3.

Table 3. Regression estimates (odds ratio's and 95% confidence intervals) of the associations between the dual trajectories^a and demographic and clinical characteristics.

Characteristic	Low fatigability high mobility	Improving fatigability high mobility	High fatigability low mobility
Univariable analysis			
Demographic characteristics			
Age, per year	Ref.	0.94 (0.79; 1.13)	1.03 (0.86; 1.23)
Female	Ref.	0.87 (0.12; 6.21)	0.00 (0.05; 3.12)
Male	Ref.	Ref.	Ref.
<12 years of education	Ref.	0.57 (0.05; 6.08)	0.12 (0.01; 1.43)
≥12 years of education	Ref.	Ref.	Ref.
Living alone	Ref.	0.42 (0.05; 3.22)	0.90 (0.12; 7.03)
Living together	Ref.	Ref.	Ref.
Clinical characteristics			
Frailty [†] , per 0.01 point	Ref.	1.08 (0.96; 1.21)	1.26 (1.08; 1.46)
No baseline disabilities	Ref.	Ref.	Ref.
≥1 baseline disabilities	Ref.	5.78 (0.55; 60.60)	60.00 (3.04; 1185)
No depressive symptoms	Ref.	Ref.	Ref.
Depressive symptoms [‡]	Ref.	0.38 (0.03; 5.27)	1.82 (0.16; 20.71)
Comorbidity [§] , per point	Ref.	0.74 (0.47; 1.19)	0.77 (0.48; 1.26)
Length of stay, per day	Ref.	1.11 (0.78; 1.59)	1.25 (0.87; 1.79)
Multivariable analysis			
Frailty [†] , per 0.01 point	Ref.	1.09 (0.91; 1.30)	1.36 (1.07; 1.74)
No baseline disabilities	Ref.	Ref.	Ref.
≥1 baseline disabilities	Ref.	5.04 (0.23; 109.14)	9.89 (0.14; 686.83)
Comorbidity [§] , per point	Ref.	0.56 (0.41; 1.12)	0.27 (0.10; 0.74)

^a Trajectory groups were estimated by jointly modeling fatigability and mobility over four waves from hospital admission up to six months after discharge. The low fatigability high mobility trajectory was used as the reference category.

[†] Frailty was assessed using the Frailty Index.

[‡] Depressive symptoms were positive when the Geriatric Depression Score was ≥5.

[§] Comorbidity was assessed using the Charlson Comorbidity Index.

The results of the univariable multinomial logistic regression analysis showed that presence of ADL disabilities before hospital admission (OR: 60.00, 95%CI: 3.04; 1185.03) and a higher frailty index (OR: 1.26, 95% CI: 1.08; 1.46) more often lead to membership of the high fatigability and low mobility trajectory group. The results of the multivariable multinomial logistic regression analysis showed that, after adjustment for baseline ADL disabilities, patient with a lower comorbidity index score (OR per point increase: 0.27, 95%CI 0.10; 0.74) and higher frailty status (OR per 0.01 point increase: 1.36, 95%CI: 1.07; 1.74) more often belonged to the high fatigability and low mobility trajectory group (Table 3). For completeness, the results of all univariable analyses using the improving fatigability, high mobility as the reference group, the univariable regression estimates for the separate mobility and fatigability trajectories, and the correlation matrix of all covariables are presented in Supplementary Tables S4, S5, S6, and S7, respectively.

DISCUSSION

The current study identified three distinct fatigability trajectories and two mobility trajectories from hospital admission up to six months after discharge. Combining these two models revealed three joint trajectory groups of fatigability and mobility. The smallest group (11%) had the most favorable outcome with low fatigability and high mobility; the majority of the patients (52%) was allocated to the improving fatigability and high mobility trajectories; and 32% of the patients were allocated to the least favorable trajectories of both low fatigability and low mobility. Controlling for baseline ADL disability, patients with a lower comorbidity index score and higher frailty index scores were more likely to be a member of the least favorable joint trajectory group: the high fatigability and low mobility trajectories.

Univariable analyses identified baseline ADL disability and frailty as determinants of the least favorable joint trajectory group. However, the multivariable analysis identified a lower comorbidity index score as a determinant of the least favorable combined fatigability and mobility trajectory group as well. This finding was not surprising given the

complex and cyclic interplay of various factors that contribute to the manifestation of frailty such as immobility and fatigue: comorbidity and disability often co-occur together with frailty, but they are not conditional.^{15,28} In our study, patients with the highest probability of belonging to the most favorable trajectory group had more severe comorbidities (kidney disease, malignant tumor, metastatic tumor) that had a weighting factor of three or six on the total comorbidity index score. Lung disease and heart failure, on the other hand, were more common in people in the other two trajectory groups, which in themselves are less lethal than, for example, a metastatic tumor, but have a direct and profound impact on mobility and fatigability, which may explain the found associations between a lower baseline comorbidity index score and a higher likelihood of belonging to the high fatigability and low mobility trajectory group.

Where previous research has identified fatigability to be a marker of mobility decline in community-dwelling older adults^{5,6}, our findings suggest that fatigability is associated with mobility levels of older medical patients as well. More specifically, patients with initial high, but improving fatigability levels over the six months following hospital discharge were more likely to have high mobility levels over the same period. By contrast, patients in the constant high fatigability trajectory were more likely to keep low mobility levels over time. However, our results should be interpreted with caution, due to the small study sample, and larger studies are strongly recommended to determine the value of fatigability assessment in identifying older medical patients at risk for developing mobility problems.

Strengths of this study included repeated measures of comprehensive fatigability and mobility assessments that were collected by face-to-face contact, allowing us to investigate the development and interactions of these constructs longitudinally and secured high reliability of the collected data.

Despite this study was initially designed to investigate associations between fatigue, mobility and clinical patient characteristics, the low inclusion rates have led to underpowered results and thus unreliable estimates which limits to draw firm conclusion.³² Nevertheless, we have learned important lessons that can be of added value for other

researchers aiming to design comparable studies in vulnerable groups: 1. Develop targeted study protocols with a limited amount of performance-based assessments to limit the required effort of the older patient to increase the patient's commitment to participate.³³ Our extensive protocol including lung function analysis, bio-impedance analysis, mobility tracking, and blood sampling as well may have deterred the most vulnerable patients from participation; 2. Try to achieve cooperation and support from the attending professionals where possible, such as nursing or physiotherapeutic care.³⁴ In our case, we got a lot of support from the attending professionals when we had to extend the study duration by one year to increase the inclusion rates; 3. Provide permanent and committed research staff during the study to ensure a good atmosphere and familiar faces, which may lead to higher commitment.³³ In our study, the low dropout rate other than death may have been due to the good relationship with our participants, which they appreciated.

CONCLUSION

Older medical patients showed three major distinct trajectories of fatigability and mobility from hospital admission up to six months after discharge. Controlling for baseline disabilities in ADL, a lower comorbidity index score and a higher baseline frailty index score were associated with the high fatigability and low mobility trajectories. Due to the small sample size in the current study, larger studies are needed to determine the value of fatigability assessment in identifying older medical patients at risk for developing mobility problems.

REFERENCES

1. Henoeh I, Sawatzky R, Falk H, et al. Symptom distress profiles in hospitalized patients in sweden: A cross-sectional study. *Res Nurs Heal*. 2014;37(6):512-523. doi:10.1002/nur.21624
2. Van Seben R, Reichardt LA, Aarden JJ, et al. The Course of Geriatric Syndromes in Acutely Hospitalized Older Adults: The Hospital-ADL Study. *J Am Med Dir Assoc*. 2019;20(2):152-158.e2. doi:10.1016/j.jamda.2018.08.003
3. Van Seben R, Reichardt LA, Essink DR, Van Munster BC, Bosch JA, Buurman BM. "I Feel Worn Out, as if i Neglected Myself": Older Patients' Perspectives on Post-hospital Symptoms after Acute Hospitalization. *Gerontologist*. 2019;59(2):315-326. doi:10.1093/geront/gnx192
4. Eldadah BA. Fatigue and Fatigability in Older Adults. *PM R*. 2010;2(5):406-413. doi:10.1016/j.pmrj.2010.03.022
5. Simonsick EM, Glynn NW, Jerome GJ, Shardell M, Schrack JA, Ferrucci L. Fatigued, but Not Frail: Perceived Fatigability as a Marker of Impending Decline in Mobility-Intact Older Adults. *J Am Geriatr Soc*. 2016;64(6):1287-1292. doi:10.1111/jgs.14138
6. Simonsick EM, Schrack JA, Santanasto AJ, Studenski SA, Ferrucci L, Glynn NW. Pittsburgh Fatigability Scale: One-Page Predictor of Mobility Decline in Mobility-Intact Older Adults. *J Am Geriatr Soc*. 2018;66(11):2092-2096. doi:10.1111/jgs.15531
7. Wanigatunga AA, Simonsick EM, Zipunnikov V, et al. Perceived Fatigability and Objective Physical Activity in Mid- to Late-Life. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2018;73(5):630-635. doi:10.1093/gerona/glx181
8. Loyd C, Do ADM, Zhang Y, et al. Prevalence of Hospital-Associated Disability in Older Adults : A Meta-analysis. *J Am Med Dir Assoc*. 2019. doi:10.1016/j.jamda.2019.09.015
9. Chen CC-H, Wang C, Huang G-H. Functional Trajectory 6 Months Posthospitalization. *Nurs Res*. 2008;57(2):93-100. doi:10.1097/01.nnr.0000313485.18670.e2
10. Boyd C, Ricks M, Fried LP, Guralnik JM, Xue Q, Bandeen-Roche K. Functional decline and recovery of activities of daily living in hospitalized, disabled older women: the Women's Health and Aging Study I. *J Am Geriatr Soc*. 2009;57(10):1757-1766. doi:10.1111/j.1532-5415.2009.02455.x.Functional
11. Dasgupta M, Brymer C. Poor functional recovery after delirium is associated with other geriatric syndromes and additional illnesses. *Int Psychogeriatrics*. 2015;27(5):793-802. doi:10.1017/S1041610214002658
12. Huang HT, Chang CM, Liu LF, Lin HS, Chen CH. Trajectories and predictors of functional decline of hospitalised older patients. *J Clin Nurs*. 2013;22(9-10):1322-1331. doi:10.1111/jocn.12055
13. Brown CJ, Friedkin RJ, Inouye SK. Prevalence and Outcomes of Low Mobility in Hospitalized Older Patients. *J Am Geriatr Soc*. 2004;52:1263-1270.
14. Volpato S, Cavalieri M, Sioulis F, et al. Predictive Value of the Short Physical Performance Battery Following Hospitalization in Older Patients. *Journals Gerontol Ser A Biol Sci Med Sci*. 2011;66A(1):89-96. doi:10.1093/gerona/glq167
15. Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G. Untangling the Concepts of Disability, Frailty, and Comorbidity: Implications for Improved Targeting and Care. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2004;59(3):255-263. doi:10.1093/gerona/59.3.m255
16. Xue QL. The Frailty Syndrome: Definition and Natural History. *Clin Geriatr Med*. 2011;27(1):1-15. doi:10.1016/j.cger.2010.08.009
17. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8(1):24. doi:10.1186/1471-2318-8-24
18. Laan W, Zuithoff NPA, Drubbel I, et al. Validity and reliability of the Katz-15 scale to measure unfavorable health outcomes in community-dwelling older people. *J Nutr Health Aging*. 2014;18(9):848-854. doi:10.1007/s12603-014-0479-3
19. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression: Two questions are as good as many. *J Gen Intern Med*. 1997; 12(7): 439-45. doi: 10.1046/j.1525-1497.1997.00076.x
20. Sheikh, J.I., & Yesavage JA. Geriatric Depression Scale (GDS). Recent evidence and development of a shorter version. In: Brink TL, ed.

- Gerontology: A Guide to Clinical Assessment and Intervention. New York: NY: The Haworth Press, Inc.; 1986:165–173.
21. Pocklington C, Gilbody S, Manea L, McMillan D. The diagnostic accuracy of brief versions of the Geriatric Depression Scale: a systematic review and meta-analysis. *Int J Geriatr Psychiatry*. 2016;31(8):837–857. doi:10.1002/gps.4407
 22. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373–383
 23. Glynn NW, Santanasto AJ, Simonsick EM, et al. The Pittsburgh fatigability scale for older adults: Development and validation. *J Am Geriatr Soc*. 2015;63(1):130–135. doi:10.1111/jgs.13191
 24. Feenstra M, Smidt N, van Munster BC, Glynn NW, de Rooij SE. Translation and validation of the Dutch Pittsburgh Fatigability Scale for older adults. *BMC Geriatr*. 2020;20(234). doi:https://doi.org/10.1186/s12877-020-01630-8
 25. de Morton NA, Davidson M, Keating JL. The de Morton Mobility Index (DEMMI): An essential health index for an ageing world. *Health Qual Life Outcomes*. 2008;6(1):63. doi:10.1186/1477-7525-6-63
 26. Jans MP, Slootweg VC, Boot CR, De Morton NA, Van Der Sluis G, Van Meeteren NL. Reproducibility and validity of the Dutch translation of the de Morton Mobility Index (DEMMI) used by physiotherapists in older patients with knee or Hip osteoarthritis. *Arch Phys Med Rehabil*. 2011;92(11):1892–1899. doi:10.1016/j.apmr.2011.05.011
 27. Nagin D. *Group-Based Modeling of Development*. Cambridge, Massachusetts: Harvard University Press; 2005.
 28. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol*. 2010;6:109–138. doi:10.1146/annurev.clinpsy.121208.131413
 29. Theou O, Rockwood MRH, Mitnitski A, Rockwood K. Disability and comorbidity in relation to frailty: How much do they overlap? *Arch Gerontol Geriatr*. 2012;55(2):e1. doi:10.1016/j.archger.2012.03.001
 30. Cooper R, Popham M, Santanasto AJ, Hardy R, Glynn NW, Kuh D. Are BMI and inflammatory markers independently associated with physical fatigability in old age? *Int J Obes*. 2019;43(4):832–841. doi:10.1038/s41366-018-0087-0
 31. Jones BL, Nagin DS. A Note on a Stata Plugin for Estimating Group-based Trajectory Models. *Sociol Methods Res*. 2013;42(4):608–613. doi:10.1177/0049124113503141
 32. Hackshaw A. Small studies: Strengths and limitations. *Eur Respir J*. 2008;32(5):1141–1143. doi:10.1183/09031936.00136408
 33. Portegijs E, Karavirta L, Saajanaho M, Rantalainen T, Rantanen T. Assessing physical performance and physical activity in large population-based aging studies: home-based assessments or visits to the research center? *BMC Public Health*. 2019;19(1):1570. doi:10.1186/s12889-019-7869-8
 34. Mody L, Miller DK, McGloin JM, et al. Recruitment and Retention of Older Adults in Aging Research. *J Am Geriatr Soc*. 2008;56(12):2340–2348. doi:10.1111/j.1532-5415.2008.02015.x. Recruitment

SUPPLEMENTARY FILES – CHAPTER 6

Table S1. Item description used for the Frailty Index.

Deficits were rescaled into values between 0 and 1, representing the presence of the deficits measured. The total score represents present deficits as a proportion of all deficits measured, resulting in a score ranging from 0 to 1.

Variable	Scoring
ADL / iADL domain	
1. Help bathing	Yes = 1
2. Help dressing	No = 0
3. Help grooming	
4. Help using toilet	
5. Difficulty getting out chair	
6. Help getting out chair	
7. Help walking	
8. Help eating	
9. Help using phone	
10. Help transport	
11. Help shopping	
12. Help meal preparations	
13. Help housework	
14. Help taking medications	
15. Help finances	
Geriatric syndrome	
16. Urinary incontinence	Yes = 1
17. Visual impairment	No = 0
18. Hearing impairment	
19. Fall in past month	
20. Unintentional weight loss	
21. Concurrent use of ≥5 medications	

Psychosocial domain

22. Feel everything is an effort	Yes = 1 No = 0
23. Difficulty get going	
24. Self-rated health	Poor = 1 Fair = 0.75 Good = 0.5 Very Good = 0.25 Excellent = 0
25. Did you feel downhearted and blue?	Both yes = 1
26. Have you dropped many of your interests and activities?	One yes = 0.5 Both no = 0

Symptoms

27. Pain / discomfort	No pain / discomfort = 0 Moderate = 0.5 Extreme = 1
28. Mood	Not anxious / depressed = 0 Moderately = 0.5 Extreme = 1
29. Loss of appetite*	Yes = 1 No = 0

Physical domain

30. Housing situation	Independent / conventional = 0 Transitional = 0.5 Institutionalized = 1
31. Living situation	Living alone = 1 Cohabiting/married = 0
32. Walk outside for 5 minutes	Impossible = 1 Only with help of somebody else = 0.75 Much effort = 0.5 Some effort = 0.25 No effort = 0
33. Physical exercise for at least 30 minutes	Never = 1 Monthly = 0.75 Weekly = 0.5 3 days/week = 0.25 Daily = 0

Cognitive domain

34. MOCA†	Total score ≥26 = 0 Total score ≤25 = 1
-----------	--

Notes: All baseline assessments referred to the situation two weeks before hospital admission unless indicated otherwise. (i)ADL, (instrumental) Activities of Daily Living; MOCA, Montreal Cognitive Assessment.

* Baseline assessment referred to a month before hospital admission.

† The MOCA was assessed within the first four days of hospital admission.

Table S2. Charlson Comorbidity Index item scores within the joint trajectories of fatigability and mobility

	Total	Trajectory group ^a		
		Low fatigability high mobility	Improving fatigability high mobility	High fatigability low mobility
Total n (%)	44 (100)	5 (11)	23 (52)	16 (36)
Comorbidity Index, median (IQR)	2 (1; 3)	3 (2; 5)	2 (1; 3)	2 (1; 3)
Acute myocardial infarction [*]	4	-	2	2
Congestive heart failure [*]	11	-	6	5
Peripheral vascular disease [*]	3	1	1	1
Cerebrovascular accident [*]	4	-	4	-
Dementia [*]	-	-	-	-
Chronic lung disease [*]	18	-	9	8
Rheumatic disease [*]	4	1	2	1
Peptic ulcer [*]	2	1	-	1
Mild liver disease [*]	-	-	-	-
Mild to moderate diabetes [*]	4	-	2	2
Diabetes with chronic complications [†]	1	-	-	1
Hemi- / paraplegia [†]	-	-	-	-
Kidney disease [†]	1	1	-	-
Malignant tumors [†]	6	1	4	1
Leukemia [†]	1	-	1	-
Lymphoma [†]	2	-	-	2
Moderate to serious liver disease ^{**}	1	-	1	-
Solid, metastatic tumor [†]	3	1	1	1
AIDS [‡]	-	-	-	-
Missing CCI data	2	1	-	1

Results are presented as n (%) unless indicated otherwise.
* counts for 1 point to the Charlson Comorbidity Index total score.
† counts for 2 points to the Charlson Comorbidity Index total score.
** counts for 3 points to the Charlson Comorbidity Index total score.
‡ counts for 6 points to the Charlson Comorbidity Index total score.

Table S3. Group characteristics of patients with missing and complete data for the main outcomes.

Baseline characteristic	Not missing DEMMI data	Missing DEMMI data	Not missing PFS data	Missing PFS data
n (%)	39 (89)	5 (11)	35 (80)	9 (20)
Demographic characteristics				
Age in years, mean (SD)	77 (5.5)	74 (7.5)	76.7 (5.5)	77.2 (5.9)
Female gender	20 (51)	2 (40)	20 (57)	2 (22)
<12 years of education	16 (42)	2 (40)	14 (41)	4 (44)
Living alone	12 (31)	1 (20)	12 (34)	1 (11)
Clinical characteristics				
Frailty [†] , mean (SD)	0.25 (0.15)	0.31 (0.15)	0.26 (0.14)	0.24 (0.15)
≥1 disabilities	26 (67)	3 (75)	24 (69)	5 (62)
≥1 disabilities basic ADL	15 (38)	3 (60)	15 (43)	3 (33)
≥1 disabilities iADL	23 (59)	3 (75)	22 (63)	4 (50)
Depressive symptoms [‡]	7 (18)	1 (20)	7 (20)	1 (11)
Comorbidity [§] , mean (SD)	2.4 (2.0)	1.8 (0.8)	2.5 (2.0)	1.2 (1.3)
LoS in days, mean (SD)	6.2 (4.2)	6.6 (4.3)	6 (4.1)	7.2 (4.7)
Fatigability , mean (SD)	30.7 (9.0)	27.7 (8.5)	-	-
Mobility , mean (SD)	-	-	66 (22)	59 (28)

Results are presented as n (%) unless indicated otherwise.
†. Frailty was assessed by the Frailty Index.
‡. Depressive symptoms were positive when the Geriatric Depression Score was ≥5.
§. Comorbidity was assessed by the Charlson Comorbidity Index.
|| Fatigability was assessed by the physical subscale of the PFS.
|| Mobility was assessed by the DEMMI.
Abbreviations: ADL, activities of daily living; DEMMI, De Morton Mobility Index; IQR, interquartile range; LoS, length of stay; PFS, Pittsburgh Fatigability Scale.

Table S4. Univariable regression estimates (odds ratio's and 95% confidence intervals) of the associations between the dual fatigability and mobility trajectories[#] and demographic and clinical characteristics using the improving fatigability high mobility trajectory group as the reference group.

Characteristic	Low fatigability high mobility	Improving fatigability high mobility	High fatigability low mobility
Age, per year	1.06 (0.89; 1.27)	Ref.	1.09 (0.97; 1.23)
Sex			
Female	0.87 (0.12; 6.21)	Ref.	2.17 (0.59; 7.99)
Male		Ref.	
Education			
<12 years of education	1.75 (0.16; 18.62)	Ref.	0.22 (0.05; 0.88)*
≥12 years of education		Ref.	
Living situation			
Living alone	2.40 (0.31; 18.55)	Ref.	2.16 (0.52; 8.90)
Living together		Ref.	
Functional status			
No baseline disabilities		Ref.	
≥1 baseline disabilities	0.17 (0.02; 1.82)	Ref.	10.38 (1.16; 93.29)*
Frailty [†] , per 0.01 point	0.93 (0.83; 1.04)	Ref.	1.17 (1.06; 1.29)*
Depressive symptoms			
No depressive symptoms		Ref.	
Depressive symptoms	2.63 (0.19; 36.34)	Ref.	4.77 (0.79; 28.72)
Comorbidity [§] , per point	1.35 (0.84; 2.15)	Ref.	1.04 (0.73; 1.49)
Length of stay, per day	0.90 (0.63; 1.28)	Ref.	1.12 (0.96; 1.31)

[#] Trajectory groups were estimated by jointly modeling fatigability and mobility over four waves from hospital admission to six months post discharge. The improving fatigability high mobility trajectory was used as the reference category.

[†] Frailty was assessed using the Frailty Index.

[§] Comorbidity was assessed using the Charlson Comorbidity Index.

* p: 0.05

Table S5. Univariable regression estimates (odds ratio's and 95% confidence intervals) of the associations between the fatigability trajectories and demographic and clinical characteristics.

Characteristic	Low fatigability	Improving fatigability	Stable high fatigability
Age, per year	Ref.	0.96 (0.80; 1.15)	1.00 (0.84; 1.19)
Sex	Ref.		
Female	Ref.	1.14 (0.15; 8.59)	0.90 (0.05; 2.68)
Male	Ref.	Ref.	Ref.
Education	Ref.		
<12 years of education	Ref.	0.70 (0.06; 7.85)	0.82 (0.01; 1.58)
≥12 years of education	Ref.	Ref.	Ref.
Living situation	Ref.		
Living alone	Ref.	0.88 (0.03; 2.47)	1.00 (0.14; 7.39)
Living together	Ref.	Ref.	Ref.
Functional status	Ref.		
No baseline disabilities	Ref.	Ref.	Ref.
≥1 baseline disabilities	Ref.	5.50 (0.51; 59.01)	34.00 (2.44; 474.55)**
Frailty [†] , per 0.01 point	Ref.	1.00 (1.00; 1.00)	1.21 (1.06; 1.39)**
Depressive symptoms	Ref.		
No depressive symptoms	Ref.	Ref.	Ref.
Depressive symptoms	Ref.	0.47 (0.03; 6.57)	1.33 (0.12; 14.90)
Comorbidity [§] , per point	Ref.	0.79 (0.49; 1.25)	0.72 (0.44; 1.18)
Length of stay, per day	Ref.	1.13 (0.80; 1.61)	1.20 (0.85; 1.71)
Baseline mobility [§] , per point	Ref.	1.01 (0.95; 1.07)	0.93 (0.87; 0.99)*

[†] Frailty was assessed using the Frailty Index.

[§] Comorbidity was assessed using the Charlson Comorbidity Index.

[§] Mobility was assessed using the De Morton Mobility Index.

* p≤ 0.05, ** p≤0.01

Table S6. Univariable regression estimates (odds ratio's and 95% confidence intervals) of the associations between the mobility trajectories and demographic and clinical characteristics.

Characteristic	High mobility	Low mobility
Age, per year	Ref.	1.17 (1.03; 1.32)*
Sex	Ref.	
Female	Ref.	0.54 (0.15; 1.92)
Male	Ref.	Ref.
Education	Ref.	
<12 years of education	Ref.	0.87 (0.07; 0.96)*
≥12 years of education	Ref.	Ref.
Living situation	Ref.	
Living alone	Ref.	2.10 (0.55; 7.99)
Living together	Ref.	Ref.
Functional status	Ref.	
No baseline disabilities	Ref.	Ref.
≥1 baseline disabilities	Ref.	12.13 (1.40; 105.25)*
Frailty [†] , per 0.01 point	Ref.	1.13 (1.05; 1.22)**
Depressive symptoms	Ref.	
No depressive symptoms	Ref.	Ref.
Depressive symptoms	Ref.	4.33 (0.87; 21.60)
Comorbidity [#] , per point	Ref.	1.08 (0.78; 1.49)
Length of stay, per day	Ref.	1.20 (1.01; 1.41)*
Baseline fatigability [§] , per point	Ref.	1.23 (1.06; 1.44)**

[†] Frailty was assessed using the Frailty Index.

[#] Comorbidity was assessed using the Charlson Comorbidity Index.

[§] Fatigability was assessed using the Pittsburgh Fatigability Scale.

* $p \leq 0.05$, ** $p \leq 0.01$

Table S7. Pearson Correlation Coefficients of the independent covariables.

Variables	1	2	3	4	5	6	7	8	9
1 Age	1.00								
2 Sex	0.03	1.00							
3 Education	-0.31	0.37	1.00						
4 Living situation	-0.27	0.40	0.35	1.00					
5 Frailty	0.07	-0.15	-0.37	-0.24	1.00				
6 ADL disabilities [*]	-0.07	-0.20	-0.24	0.01	0.56	1.00			
7 Depressive symptoms [*]	0.19	-0.18	-0.15	0.06	0.31	0.31	1.00		
8 Comorbidity	-0.07	0.02	0.24	-0.12	0.33	0.22	0.30	1.00	
9 LoS	0.12	0.16	-0.11	0.01	0.38	0.21	0.23	0.32	1.00

Abbreviations: ADL, Activities of Daily Living; LoS, Length of hospital Stay.

^{*} Scoring as used in the analyses.

CHAPTER 7

General Discussion

The main objective of this dissertation was to evaluate trajectories of health status over time among older people. The first part was devoted to examining relevant psychometric properties of various instruments representing aspects of health status in older people. The second part concerned the actual health trajectories of both community-dwelling older people and hospitalized older patients. This general discussion will first present the main findings; second, a reflection on the findings compared to the existing literature and methodological strengths and weaknesses will be discussed, and; third, attention is paid to clinical implications and future research directions to which the results of this dissertation constitute a stepping stone.

MAIN FINDINGS

In Chapter 2 the agreement between self-perceived and measured change in Activities of Daily Living (ADL) status by the 15-item Katz ADL Index from pre-hospital admission to three and twelve months post discharge was investigated. Poor agreement was found between self-perceived and formally measured change in ADL status from pre-hospital admission over both timeframes.

In Chapter 3 the reproducibility and responsiveness were investigated of two most widely used frailty instruments: the Frailty Index (FI) and the Frailty Phenotype (FP). By using an anchor-based method, the Frailty Index showed good reproducibility and responsiveness over both three and twelve months timeframes. The Frailty Phenotype performed less well, as indicated by poor to moderate reproducibility and a sufficient responsiveness over a timeframe of 12 months only.

In Chapter 4, the Pittsburgh Fatigability Scale (PFS) was translated into Dutch and subsequently validated in a cohort of older hospitalized patients. After extensive translation and testing procedures, the Dutch version of the PFS demonstrated good content validity, construct validity, and test-retest reliability for measuring perceived physical and mental fatigability in hospitalized adults aged 70 years and older.

In Chapter 5 the course of health status in community-dwelling older people was investigated. Four stable trajectories of self-rated health (SRH) over time were identified. Poor SRH trajectory membership was associated with female sex, lower educational levels, the presence of two or more chronic conditions, unhealthy life style behavior, and abnormalities in three or more physiological markers.

In the final study, Chapter 6, dual trajectory modeling was used to explore the relation between fatigability and mobility in older medical patients. Three distinct trajectory combinations of fatigability and mobility from hospital admission up to six months after discharge were identified: the high fatigability and low mobility trajectory group, the low fatigability and high mobility trajectory group, and the improving fatigability, low mobility trajectory group. Here, a lower comorbidity index score and a higher baseline frailty index score were associated with the least favorable trajectory combination.

In summary, three important lessons have been learned with this dissertation, which will be reflected on in the next parts of this chapter:

1. Common used instruments of health status in older hospitalized people, although used as such, were not necessarily suitable for measuring change in health status over time (Chapters 2 and 3);
2. Fatigability can be measured validly and reliably in older hospitalized patients, but the added value as a determinant of health status has not yet been conclusively demonstrated (Chapters 4 and 6);
3. Certain determinants that can be traced back to frailty have a negative influence on health status trajectories in older people (Chapters 5 and 6).

REFLECTION ON THE FINDINGS

Common instruments to measuring health status over time in older people

As already mentioned in the introduction, health status is a difficult construct to operationalize. In addition, the purpose of the development of an instrument and its measurements properties have implications for their application.¹ Still, instruments to assess health status in older people are used to evaluate the effects of interventions, despite they are not being developed or tested for responsiveness. A first example is the frequent use of the Katz ADL Index to operationalize recovery of functional decline in ADL by comparing the number of ADL disabilities before and after hospitalization.²⁻⁵ Among hospitalized older people, high proportions of misclassification existed between change scores of the Katz ADL Index and patients' perceptions of change in ADL in both directions (Chapter 2). Consequently, detecting any change in ADL status based on the calculation of disabilities according to the Katz ADL Index does not necessarily mean that a patient perceived himself deteriorated or improved in the same direction as well. The found discrepancies may be indicative for the occurrence of response shift, which may lead to an over- or underestimation of the effects if unadjusted.^{6,7} These results contribute to the opinion that the Katz ADL Index score is not suitable to measure change in ADL status over time that is relevant to the individual patient, and should therefore not be used for evaluative purposes in individuals.

Another example of unintended use of instruments concerns frailty status, which is often used as a determinant of negative health outcomes in older hospitalized people. Frailty is defined as a state of increased vulnerability due to a reduced ability to maintain homeostasis after a stressor event as a consequence of cumulative decline in multiple physiological systems during a lifetime.⁸ The gold standard for measuring frailty in a clinical setting is the comprehensive geriatric assessment (CGA), yet the use of an index based on CGA outcomes that can be used for research purposes has to be further validated.⁸ The FI and the FP, originally developed for screening and risk-assessment purposes, are the most widely used instruments used to establish frailty status in research.^{9,10} Recently, these instruments have been used to evaluate outcomes of interventions aiming at slowing down the development of frailty.^{11,12} Because information about the responsiveness was generally lacking, it remained unclear which

frailty instrument was most appropriate to measure the effectiveness of interventions.^{13–15} To contribute to the body of knowledge, the reliability and responsiveness of the FI and FP were investigated (Chapter 3). Our results indicated that the Frailty Index is more sensitive to measuring change over time than the Frailty Phenotype. Consequently, the Frailty Index seems more suitable for evaluative purposes in studies among older hospitalized people.

Fatigability as a determinant of health-status

In the search of relevant outcome measures of older hospitalized patients, fatigability emerged as a particular interesting outcome. On the one hand, this was prompted by the high incidence of fatigue in hospitalized older people, which is known from the literature^{16,17}, and what was also observed in the Seniorlines cohort. On the other hand, the existing studies of fatigability among community-dwelling older people indicated that fatigability appeared to be a promising measure for detecting people who are susceptible to functional decline.^{18–20} To be able to assess fatigability in our Dutch population of older hospital patients (Chapter 6), the Pittsburgh Fatigability Scale (PFS) was first successfully translated and validated (Chapter 4). Although the study sample was small, in Chapter 6, clear patterns of three distinct fatigability trajectories were identified. However, the absence of other studies that investigated trajectories of fatigability over time limited mutual comparisons. Therefore, no clear-cut conclusions could be drawn about the value of fatigability assessment in identifying older medical patients at risk for developing health-status decline due to mobility problems.

Determinants of poor health status over time

When investigating the course of health status in community-dwelling older people, four stable trajectories of SRH over time among 12,685 community-dwelling older people were identified (Chapter 5). Although the majority of the people had highest probability of belonging to the moderate SRH trajectory, the main focus was the group with the highest probability of poor SRH, as poor SRH is a known predictor of mortality among older people.^{21,22} In this cohort, poor SRH trajectory membership was associated with female sex, lower educational levels, presence of multimorbidity (defined as two or more chronic conditions), unhealthy life style behavior, and abnormalities in three or more physiological

markers, of which the latter is indicative of pre-frailty.²³ In the cohort of older hospitalized people, controlling for baseline disabilities in ADL, a lower comorbidity index score (based on the Charlson Comorbidity Index (CCI)) and a higher baseline frailty index score were associated with the least favorable trajectory combination: the high fatigability and low mobility trajectory group. The seemingly unexpected inverse association between a lower baseline comorbidity index score and a higher likelihood of the worst outcome is likely explained by the fact that patients in the most favorable trajectory group often had severe chronic disease, such as metastatic tumor and kidney disease. These conditions had higher weighting factors than for instance chronic lung disease or rheumatic disease, which were more present among the patients of the high fatigability and low mobility trajectory group. The weights of the CCI were originally added for the relative risk of mortality. However, our outcomes were fatigability and mobility which might have had a different weighting. For instance, the relative risk of mobility decline within six months in patients having chronic lung disease or rheumatic disease is probably higher than the relative risk of dying. Consequently, to clarify the association between co- or multimorbidity and health status trajectories, the use of a generic outcome that counts the co-occurrence of multiple chronic conditions within a person without weighting, such as the Index of Co-Existent Disease²⁴, is recommended.²⁵

In summary, the determinants found in Chapters 5 and 6 that were associated with unfavorable health status trajectories can be traced back to frailty. Frailty is considered as an accumulation of measurable deficits that affect physiological reserves of underlying organ systems to such an extent that there is less resilience to deal with other external stressors, such as hospitalization.⁸ In chapter 5, these deficits were reflected by abnormalities in physiological markers and the presence of multiple chronic conditions. Deviating physiological markers are directly linked to inefficient functioning of the underlying physiological systems that affect frailty.²³ Presence of chronic conditions have an impact on physiological reserves as well, which cannot be used when acute stressors, such as hospitalizations, occur.²⁶ In chapter 6, the number of deficits were directly assessed by using the Frailty Index, which is based on deficits associated with health status.²⁷

In addition, a lower educational level, female sex, and an unhealthy lifestyle were determinants of poor health status of older people in chapter 5. Previous literature has already proven the protective effect of higher education on poor health status^{28–30}, even in older age groups.^{29,31} Regarding the relation between sex and poor health status over time, there is no conclusive evidence, as the literature shows ambiguous results.^{30,32,33} However, sex and educational level in older adults are not considered modifiable determinants. From a pragmatic perspective, therefore, the most interesting determinant is probably the finding that unhealthy lifestyle behavior was associated with the poor health status trajectory, as it may provide a base for preventing or reversing poor health status.

Older people with multimorbidity or frailty may benefit from specific intervention programs aiming at improving health behavior and health status.^{11,34–36} A community based self-management program among people with multimorbidity including education and lifestyle activities, lead to a small decrease in length of hospital stay within four to 6 months after the intervention³⁴. Among frail older people recruited from the community, long term care facilities, retirement homes and hospitals, no effects of health care utilization or care costs were found.^{11,35} Intervention programs that include physical exercise showed the greatest improvements in physical functioning and functional ability in both frail older people^{11,35} and in older people with multimorbidity.³⁴ More specifically, multidomain intervention programs, for instance combining resistance, aerobic, balance, and flexibility exercises, lasting for more than 5 months with a frequency of three times per week for 30 to 45 minutes per session were most beneficial³⁵, and group training seems to be more effective than individual programs.^{11,36}

METHODOLOGICAL CONSIDERATIONS

Research into older people, and in elderly hospital patients in particular, poses several challenges. The older populations consists of a heterogeneous group of people with varying levels of vulnerability, fitness, and number and severity of chronic conditions potentially leading

to selective response of only the fittest older patients. In addition, the risk of dropout, for example due to death or mobility constraints, is higher than in younger populations.³⁷ In this section, the applied strategies to increase recruitment, as well as strategies to reduce the risk of attrition bias are discussed. Finally, the methodological strengths and limitations of the methods used to measure change in health status of the studies in this dissertation will be discussed.

Challenges in studies of older people

Recruitment strategies

A main concern in observational studies is the risk of a lower internal validity due to the likelihood of selective response, which in turn affects external validity.^{38,39} External validity in observational studies concerns the representativeness of the population.³⁸ The external validity of the Lifelines Cohort Study was explicitly examined⁴⁰, but not for the Seniorlines Cohort Study. Although the Seniorlines Cohort Study was performed in one hospital, the broad inclusion criteria for participation and the fact that the UMCG provides both primary and specialist care for patients from several regions of the Netherlands, we evaluated the sample as a representative reflection of older hospitalized patients in general. Recruitment was low in the Seniorlines 2.0 study and a certain degree of selective response is expected to have occurred because the extensive study protocol may have deterred the most vulnerable older patients from participation. We tried to overcome low recruitment rates by providing home-visits for follow-up assessments and using an opt-out recruitment strategy.^{41,42} Unfortunately, these efforts were to no avail, as shown by the small study sample and corresponding low statistical power in Chapter 6. When the more vulnerable patient would have participated, there may have been more subgroups identified within the sample, and the found differences between the best and worst identified subgroups may have been even larger. The results described in Chapter 6 showed that the differences between the mean fatigability score of the best and worst trajectories already exceeded the predefined smallest detectable change scores, indicating real subsamples. However, the number of trajectories is at odds with the number of patients within each trajectory⁴³, thus increasing the sample size is the main challenge for future research into trajectories of fatigability and mobility in older hospital patients.

Drop-out and attrition bias

Missing data and attrition bias in longitudinal studies are common and serious problems in longitudinal studies⁴⁴, especially in studies of older people.³⁷ Drop-out rates and percentage of deceased subjects of the studies used in this dissertation were 36% (†4%), 40% (†16%), and 22% (†17%) of the Lifelines, Seniorlines, and Seniorlines 2.0 cohort studies, respectively. To increase retention rates the research nurses proactively asked for the desire for a short break during the study-procedures (Seniorlines and Seniorlines 2.0) and advance notification and home visits were part of the follow-up measurements (Seniorlines 2.0). In addition, several analytical strategies were applied to reduce the negative effect of attrition and missing data on the results⁴⁴: 1. Missing data were imputed to increase statistical power and results were compared with sensitivity analysis including only complete cases (Chapters 2–4); 2. Maximum likelihood estimation of missing data for some time points in the group-based trajectory analysis were applied (Chapters 5 and 6); 3. A group-based trajectory modeling technique was applied in a sensitivity analysis, which takes non-random attrition into account⁴⁵ (Chapter 5). All of the sensitivity analyses yielded no different conclusions, indicating no serious effects of attrition or missing data on the results.

Measuring change in health status

Selecting the primary outcomes for trajectories of health status

The selection of primary outcome measures should be based on the purpose why an instrument is developed and sound psychometric properties. SRH was selected as the main outcome of Chapter 5 because SRH is recommended as a comprehensive indicator of health status within population studies.⁴⁶ However, SRH is a global measure of health status, consisting of one question, potentially limiting its responsiveness which may explain why four stable trajectories of SRH were identified in Chapter 5.⁴⁷ It is therefore suggested to use this outcome as a proxy health status over time, but preferably together with other instruments of health status, like generic multi-domain or disease specific instruments.⁴⁸ Unfortunately, no multi-domain or disease-specific instrument of health status was included in the Lifelines questionnaires. Despite this limitation, our results are of added value, as it has been shown that pre-frailty status in community-dwelling older people is associated with poor health status over time, which may inspire other researchers to elaborate on this.

Fatigability and mobility were the primary outcomes of trajectories of health status in Chapter 6. As explained above, fatigability was selected to elaborate on the assumed relationship with poor health outcomes. Mobility was selected as an objective and responsive proxy of health status due to its association with disabilities in ADL⁴⁹, self-report of functioning⁵⁰, and perceptions of health.⁵⁰ Moreover, mobility is conditional to independence and maintaining independence is one of the health-outcomes that older people prioritize, making mobility a meaningful outcome for older people.^{51,52} Based on excellent psychometric properties in older hospitalized patients and the absence of a floor effect in bedridden patients the preferred instrument when studying this population should be the De Morton Mobility Index (DEMMI).⁵³ For community-dwelling older people the Tinetti Scale or the Short Physical Performance Battery (SPPB) are recommended, yet when the study population concerns both community-dwelling and (non-bedridden) hospitalized older people one should consider using the Short Physical Performance Battery.⁵³

The use of anchor questions

Anchor questions are used to retrospectively assess perceived change in the construct of interest.⁵⁴ Although, the use of anchor questions is an established method to determine responsiveness of patient-reported outcome measures^{55,56}, its use is often criticized due to the risk of recall bias.^{57,58} Here, the perspective is important, as for individual patients in general the biggest gain is to 'feel better', regardless of whether it is caused by the treatment, a placebo, or one's ability to adapt.⁵⁹ Consequently, to assess perceived change in individuals, a single anchor question asking for the individual patients' perception of change, as in Chapter 2, can be applied without any problems. Recall bias may be of influence on population level, for instance when using an anchor question as 'the gold standard' for perceived change when studying the responsiveness.⁵⁴ In Chapter 3, recall bias could not be ruled out, as the responsiveness of the FI and FP was investigated by using two anchor questions referring to two frailty-related constructs (functional status and health status). However, especially in determining responsiveness, there is no preferred method to investigate, and combinations of distribution-based and anchor-based methods are recommended to gain an impression of the properties of an instrument to measure meaningful change on individual or group level.^{56,60} Therefore, with this study, we have taken a first step and more

will have to follow to get a clear picture of the responsiveness of existing frailty instruments in older people.

Trajectory analysis

Group-Based Trajectory Modeling (GBTM) provided the opportunity to identify clusters of individuals that followed certain health trajectories and determinants that were likely associated with belonging to the least favorable trajectories. In addition, GBTM provides a graphical representation of how an outcome changes over time for different subgroups within a population which make the results tangible and easier to understand.⁶¹ GBTM is distinguished from other methods that are used to analyze trajectories over time by restricting within-group variability.⁶² This means that all individuals that are allocated to a certain trajectory group are comparable concerning the direction (or slope) and level (or intercept) of their trajectory. Restricting within-group variability theoretically results in a true representation of the actual subgroups in a population.^{62,63} In addition, as predefined assumptions about the number of expected trajectories are not required, using GBTM can lead to new hypotheses or theories about the presence of subpopulations within a certain population. However, the disadvantage is that this method may lead to inexplicable or too small groups. Therefore, it is recommended not only to rely on statistical measures when selecting the optimal number of trajectories, but the identified trajectories should be clinically interpretable as well.⁶²⁻⁶⁴ Though GBTM, like any model, remains an estimation of reality, we believe that GBTM is a valuable method to identify those most at risk of, in this dissertation, poor health status, which may provide a base for future preventive interventions.

CLINICAL IMPLICATIONS AND FUTURE DIRECTIONS

Implications for the use of instruments to measure health status over time in older people

The common thread running through this dissertation is the attempt to use meaningful outcomes that were suitable to measure health status of older people over time. Table 1 presents an overview of the instruments that were investigated for their psychometric properties in the first three chapters of this dissertation. In summary, based on the results of the first three chapters of this dissertation, supplemented with existing literature, it can be concluded that: First, the FI is more responsive than the FP (Chapter 3). However, using an anchor-based method, the minimal detectable change (MDC) based on the stable group exceeded the mean change found in the deteriorated group (Chapter 3). Therefore, also taking into account that frailty is a concept that is not used by older people themselves⁶⁵, it is recommended to use this instrument at group-level only. Second, based on the poor agreement between self-reported change and measured change in ADL status in hospitalized older people (Chapter 2) and because the MDC exceeds the minimal important change in community-dwelling older people⁶⁶, it was recommended to use the Katz ADL Index at group-level only. In addition, a mean difference of at least three disabilities should be taken into account to identify change in ADL status beyond measurement error.⁶⁶ Third, based on the results of the test-retest data, the Dutch PFS can be applied to measure change over time on group-level, taking into account change scores of twelve and fourteen points for the physical and mental subscales, respectively (Chapter 4). These results and recommendations are summarized in Table 1.

Table 1. Summary and recommendations for the use of the investigated instruments as target for interventions (for individuals and groups) in older people.

	Frailty Phenotype	Frailty Index	Katz ADL Index	Pittsburgh Fatigability Scale
Original purpose of instrument	Screening of frailty status (robust, prefrail, frail). ⁸⁰	Identification of older people at risk for frailty. ^{27,81}	Quantifying functional status of older people based on self-report. ⁸²	Quantifying fatigability of older people based on self-report. ⁸³
Suitable for risk stratification of negative health outcomes*	Yes ⁸⁰	Yes ⁸¹	Yes ⁸⁴	Not known
Known relevant psychometric properties for responsiveness:				
MDC Distribution-based	Not known	Not known	In community-living older people: 1.18 - 2.82 points. ⁶⁶	In hospitalized older people: 12 points (physical subscale); 14 points (mental subscale). (Chapter 4)
MDC Anchor-based	In hospitalized older people using a 3-months interval: 1.58 - 1.84 points. (Chapter 2)	In hospitalized older people using a 3-months interval: 0.13 - 0.17 points. (Chapter 2)	In community-living older people using a 6-months interval: 1.99 - 1.81 points. ⁶⁶	Not known
Recommendations for use:				
Target of intervention studies	Possible, but the Frailty Index seems a better alternative. (Chapter 2)	Possible, taking into account a minimal change of ≥6 deficits. (Chapter 2)	Possible, taking into account a minimal change of ≥3 disabilities. ⁶⁶	Possible, taking into account a minimal change score of ≥12 (physical subscale) and ≥14 points (mental subscale). (Chapter 4)
Meaningful outcome for individuals	Inappropriate (Chapter 2)	Inappropriate (Chapter 2)	Inappropriate (Chapter 3)	Not known

Abbreviations: ADL, Activities of Daily Living; MDC, Minimal Detectable Change.
* Hospitalization, institutionalization, or death.

Future directions for preventing poor health status trajectories in older people

Having identified several trajectories of health status in older people, we identified presence of multiple chronic conditions and (pre-)frailty as the common determinants of the poorest health trajectories in both community-dwelling and hospitalized older people. As stated above, older people with multimorbidity or frailty may likely benefit from physical exercise intervention programs aiming at improving health behavior and health status.^{11,34–36} In addition, among older people with cancer, the benefit of physical exercise interventions on cancer-related fatigue has already been proven.⁶⁷ Given the association between fatigability and negative health outcomes^{18,68}, there is a strong demand for studies investigating which interventions are most suitable to improve fatigability in older people.^{69–71}

Future research on adverse health outcomes in older people

To optimize the benefits and costs of targeted interventions, it is important to identify people who need it most. Clinical symptoms of frailty, such as weight loss and exhaustion, are considered to play a key role in the identification of older hospitalized patients at risk for negative health outcomes.^{72,73} However, these are not stand-alone symptoms, but they are considered as an expression of the decreased physiological reserve due to underlying processes. Some suggest that when weight loss and exhaustion occur it is already too late to start interventions to delay or reverse the development of frailty.^{73,74} As previously demonstrated in community-dwelling older people, fatigability assessment could offer an alternative screening tool to distinguish apparently fit older people, who are susceptible to frailty, from their real fit counterparts.^{18,75} Future research should elaborate on the relationship between fatigability and frailty in older people. In addition, its use as a screening tool for determining populations who are still susceptible to delaying or reversing the development of frailty and improving resilience should be investigated as well.

Measuring change in health status in individual older people

This dissertation endorses the statement mentioned in the introduction: measuring health status and decline or improvement in health status is not straightforward. In the introduction, the clinical case description of

a fictive study participant illustrated the dynamic nature of health status by means of Mrs. White's adaptation to the changing circumstances as a consequence of the hospital admission. Consequently, simply counting the number of disabilities before and after hospitalization did not fully reflect the change in health status experienced by Mrs. White herself. In addition, the use of an outcome that is too generic, such as general self-rated health, may not be sensitive enough to serve as a realistic estimate of her perceived change in health status. It is therefore important to emphasize again that the choice of an instrument to measure health status depends on the purpose, the target group and the psychometric characteristics of the instrument of interest. Although the investigated instruments in this dissertation are suitable for measuring change in health status for research purposes (taking into account MDC), none of the investigated instruments should be used for measuring change at patient level. Fortunately, there are national and global initiatives that encourage and advise on the use of health-related outcomes that are meaningful to individual patients and can be used in the doctor's office.⁷⁶⁻⁷⁹

REFERENCES

1. Guyatt GH, Veldhuyzen Van Zanten SJO, Feeny DH, Patrick DL. Measuring quality of life in clinical trials: A taxonomy and review. *Cmaj*. 1989;140(12):1441-1448.
2. Barnes DE, Mehta KM, Boscardin WJ, et al. Prediction of recovery, dependence or death in elders who become disabled during hospitalization. *J Gen Intern Med*. 2013;28(2):261-268. doi:10.1007/s11606-012-2226-y
3. Boyd CM, Landefeld CS, Counsell SR, Robert M, Fortinsky RH, Kresevic D. Recovery in Activities of Daily Living Among Older Adults Following Hospitalization for Acute Medical Illness. *J Am Geriatr Soc*. 2008;56(12):2171-2179. doi:10.1111/j.1532-5415.2008.02023.x.Recovery
4. Dasgupta M, Brymer C. Poor functional recovery after delirium is associated with other geriatric syndromes and additional illnesses. *Int Psychogeriatrics*. 2015;27(5):793-802. doi:10.1017/S1041610214002658
5. Huang HT, Chang CM, Liu LF, Lin HS, Chen CH. Trajectories and predictors of functional decline of hospitalised older patients. *J Clin Nurs*. 2013;22(9-10):1322-1331. doi:10.1111/jocn.12055
6. Mcphail BS, Haines T. "Can You Handle the Truth?" The Response Shift Phenomenon in Clinical Trials. *J Clin Res Best Pract*. 2010;6(2):1-8.
7. Sprangers MAG, Schwartz CE. The challenge of response shift for quality-of-life-based clinical oncology research. *Ann Oncol*. 1999;10(7):747-749. doi:10.1023/A:1008305523548
8. Clegg A, Young J, Iliffe S, Olde Rikkert MGM, Rockwood K. Frailty in older people summary. *Lancet*. 2013;381(9868):752-762. doi:10.1016/S0140-6736(12)62167-9.Frailty
9. Buta BJ, Walston JD, Godino JG, et al. Frailty assessment instruments: Systematic characterization of the uses and contexts of highly-cited instruments. *Ageing Res Rev*. 2017;March(26):53-61. doi:10.1016/j.arr.2015.12.003.Frailty
10. Morley JE, Vellas B, Kan GA Van, et al. Frailty Consensus : A Call to Action. *J Am Med Dir Assoc*. 2013;14(6):392-397. doi:10.1016/j.jamda.2013.03.022
11. Apóstolo J, Cooke R, Bobrowicz-campos E, et al. Effectiveness of interventions to prevent pre-frailty and frailty progression in older adults : a systematic review. *JBISRIR-2017-003382*. 2018;16(1):140-232. doi:10.11124/JBISRIR-2017-003382
12. Dedeyne L, Deschodt M, Verschueren S, Tournoy J, Gielen E. Effects of multi-domain interventions in (pre)frail elderly on frailty, functional, and cognitive status: A systematic review. *Clin Interv Aging*. 2017;12:873-896. doi:10.2147/CIA.S130794
13. Apostolo J, Cooke R, Bobowicz-Campos E, et al. Predicting risk and outcomes for frail older adults: an umbrella review of frailty screening tools ´. *JBISRIR-2016-003018*. 2017;15(4):1154-1208. doi:10.11124/JBISRIR-2016-003018
14. De Vries NM, Staal JB, Van Ravensberg CD, Hobbelen JSM, Olde Rikkert MGM, Nijhuis-van der Sanden MWG. Outcome instruments to measure frailty: a systematic review. *Ageing Res Rev*. 2011;10:104-114. doi:10.1016/j.arr.2010.09.001
15. Drubbel I, Numans ME, Kranenburg G, Bleijenberg N, Wit NJ De. Screening for frailty in primary care : a systematic review of the psychometric properties of the frailty index in community-dwelling older people. *BMC Geriatr*. 2014;14(1):1-13. doi:10.1186/1471-2318-14-27
16. Henech I, Sawatzky R, Falk H, et al. Symptom distress profiles in hospitalized patients in sweden: A cross-sectional study. *Res Nurs Heal*. 2014;37(6):512-523. doi:10.1002/nur.21624
17. van Seben R, Reichardt LA, Aarden JJ, et al. The Course of Geriatric Syndromes in Acutely Hospitalized Older Adults: The Hospital-ADL Study. *J Am Med Dir Assoc*. 2019;20(2):152-158.e2. doi:10.1016/j.jamda.2018.08.003
18. Simonsick EM, Glynn NW, Jerome GJ, Shardell M, Schrack JA, Ferrucci L. Fatigued, but Not Frail: Perceived Fatigability as a Marker of Impending Decline in Mobility-Intact Older Adults. *J Am Geriatr Soc*. 2016;64(6):1287-1292. doi:10.1111/jgs.14138
19. Simonsick EM, Schrack JA, Santanasto AJ, Studenski SA, Ferrucci L, Glynn NW. Pittsburgh Fatigability Scale: One-Page Predictor of Mobility Decline in Mobility-Intact Older Adults. *J Am Geriatr Soc*.

- 2018;66(11):2092-2096. doi:10.1111/jgs.15531
20. Wanigatunga AA, Simonsick EM, Zipunnikov V, et al. Perceived Fatigability and Objective Physical Activity in Mid- to Late-Life. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2018;73(5):630-635. doi:10.1093/gerona/glx181
 21. DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality Prediction with a Single General Self-Rated Health Question A Meta-Analysis. *J Gen Intern Med*. 2005;20:267-275. doi:10.1111/j.1525-1497.2005.0291.x
 22. Idler EL, Benyamini Y. Self-Rated Health and Mortality : A Review of Twenty-Seven Community Studies. *J Health Soc Behav*. 1997;38(1):21-37. doi:10.1007/s10290-009-0045-y
 23. Fried LP, Xue QL, Cappola AR, et al. Nonlinear multisystem physiological dysregulation associated with frailty in older women: Implications for etiology and treatment. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2009;64(10):1049-1057. doi:10.1093/gerona/glp076
 24. Greenfield S, Apolone G, McNeil BJ, Cleary PD. The importance of co-existent disease in the occurrence of postoperative complications and one-year recovery in patients undergoing total hip replacement: Comorbidity and outcomes after hip replacement. *Med Care*. 1993;31(2):141-154. doi:10.1097/00005650-199302000-00005
 25. De Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity: A critical review of available methods. *J Clin Epidemiol*. 2003;56(3):221-229. doi:10.1016/S0895-4356(02)00585-1
 26. Weiss CO. Frailty and Chronic Diseases in Older Adults. *Clin Geriatr Med*. 2011;27(1):39-52. doi:10.1016/j.cger.2010.08.003
 27. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8(1):24. doi:10.1186/1471-2318-8-24
 28. Mackenbach JP, Van Den Bos J, Joung IMA, Van De Mheen H, Stronks K. The determinants of excellent health: Different from the determinants of ill-health? *Int J Epidemiol*. 1994;23(6):1273-1281. doi:10.1093/ije/23.6.1273
 29. von dem Knesebeck O, Verde PE, Dragano N. Education and health in 22 European countries. *Soc Sci Med*. 2006;63(5):1344-1351. doi:10.1016/j.socscimed.2006.03.043
 30. Verropoulou G. Determinants of change in self-rated health among older adults in Europe: A longitudinal perspective based on SHARE data. *Eur J Ageing*. 2012;9(4):305-318. doi:10.1007/s10433-012-0238-4
 31. Dalstra JAA, Kunst AE, Mackenbach JP. A comparative appraisal of the relationship of education, income and housing tenure with less than good health among the elderly in Europe. *Soc Sci Med*. 2006;62(8):2046-2060. doi:10.1016/j.socscimed.2005.09.001
 32. McCullough ME, Laurenceau JP. Gender and the natural history of self-rated health: A 59-year longitudinal study. *Heal Psychol*. 2004;23(6):651-655. doi:10.1037/0278-6133.23.6.651
 33. Leinonen R, Heikkinen E, Jylhä M. Self-rated health and self-assessed change in health in elderly men and women - A five-year longitudinal study. *Soc Sci Med*. 1998;46(4-5):591-597. doi:10.1016/S0277-9536(97)00205-0
 34. Brady TJ, Murphy L, O'Colmain BJ, et al. A meta-analysis of health status, health behaviors, and healthcare utilization outcomes of the Chronic Disease Self-Management Program. *Prev Chronic Dis*. 2013;10(1):120112.
 35. Theou O, Stathokostas L, Roland KP, et al. The effectiveness of exercise interventions for the management of frailty: A systematic review. *J Aging Res*. 2011;2011. doi:10.4061/2011/569194
 36. Frost R, Belk C, Jovicic A, et al. Health promotion interventions for community-dwelling older people with mild or pre-frailty: A systematic review and meta-analysis. *BMC Geriatr*. 2017;17(1):1-13. doi:10.1186/s12877-017-0547-8
 37. Chatfield MD, Brayne CE, Matthews FE. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *J Clin Epidemiol*. 2005;58(1):13-19. doi:10.1016/j.jclinepi.2004.05.006
 38. Carlson MDA, Morrison RS. Study design, precision, and validity in observational studies. *J Palliat Med*. 2009;12(1):77-82. doi:10.1089/jpm.2008.9690
 39. Gurwitz JH, Sykora K, Mamdani M, et al. Education and debate Reader's guide to critical appraisal of cohort studies: 1. Role and design. *Bmj*. 2005;330(April):895-897.

40. Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines cohort study. *PLoS One*. 2015;10(9):1–12. doi:10.1371/journal.pone.0137203
41. Lacey RJ, Wilkie R, Wynne-Jones G, Jordan JL, Wersocki E, Mcbeth J. Evidence for strategies that improve recruitment and retention of adults aged 65 years and over in randomised trials and observational studies: A systematic review. *Age Ageing*. 2017;46(6):895–903. doi:10.1093/ageing/afx057
42. Portegijs E, Karavirta L, Saajanaho M, Rantalainen T, Rantanen T. Assessing physical performance and physical activity in large population-based aging studies: home-based assessments or visits to the research center? *BMC Public Health*. 2019;19(1):1570. doi:10.1186/s12889-019-7869-8
43. Nagin DS. *Group Based Modeling of Development*. Cambridge, Massachusetts: Harvard University Press; 2005.
44. Graham JW. Missing data analysis: Making it work in the real world. *Annu Rev Psychol*. 2009;60:549–576. doi:10.1146/annurev.psych.58.110405.085530
45. Haviland AM, Jones BL, Nagin DS. Group-based Trajectory Modeling Extended to Account for Nonrandom Participant Attrition. *Sociol Methods Res*. 2011;40(2):367–390. doi:10.1177/0049124111400041
46. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Soc Sci Med*. 2009;69(3):307–316. doi:10.1016/j.socscimed.2009.05.013
47. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: A clarification. *J Clin Epidemiol*. 1989;42(5):403–408. doi:10.1016/0895-4356(89)90128-5
48. Galenkamp-van der Ploeg H. Feeling healthy versus being healthy: change and stability in older peoples self-rated health. 2013. doi:10.1039/tf9373300237
49. Wang DXM, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. *J Cachexia Sarcopenia Muscle*. 2020;11(1):3–25. doi:10.1002/jcsm.12502
50. Bravell ME, Zarit SH, Johansson B. Self-reported activities of daily living and performance-based functional ability: A study of congruence among the oldest old. *Eur J Ageing*. 2011;8(3):199–209. doi:10.1007/s10433-011-0192-6
51. Fried TR, Tinetti M, Agostini J, Iannone L, Towle V. Health outcome prioritization to elicit preferences of older persons with multiple health conditions. *Patient Educ Couns*. 2011;83(2):278–282. doi:10.1016/j.pec.2010.04.032
52. Ramer SJ, McCall NN, Robinson-Cohen C, et al. Health outcome priorities of older adults with advanced CKD and concordance with their nephrology providers' perceptions. *J Am Soc Nephrol*. 2018;29(12):2870–2878. doi:10.1681/ASN.2018060657
53. Soares Menezes KVR, Auger C, de Souza Menezes WR, Guerra RO. Instruments to evaluate mobility capacity of older adults during hospitalization: A systematic review. *Arch Gerontol Geriatr*. 2017;72(May):67–79. doi:10.1016/j.archger.2017.05.009
54. De Vet, Henrica C. W. Terwee CB, Knol DL, Mokkink LB. *Measurement in Medicine*. first. Cambridge: Cambridge University Press; 2011.
55. De Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care*. 2001;17(4):479–487.
56. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102–109. doi:10.1016/j.jclinepi.2007.03.012
57. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: A review of strengths and weaknesses and considerations for design. *J Man Manip Ther*. 2009;17(3):163–170. doi:10.1179/jmt.2009.17.3.163
58. Grøvlø L, Haugen AJ, Hasvik E, Natvig B, Brox JI, Grotle M. Patients' ratings of global perceived change during 2 years were strongly influenced by the current health status. *J Clin Epidemiol*. 2014;67(5):508–515. doi:10.1016/j.jclinepi.2013.12.001
59. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*. 2003;56:395–

407. doi:10.1016/S0895-4356(03)00044-1
60. Guyatt GH, Osoba D, Wu AW, et al. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002;77(4):371-383. doi:10.4065/77.4.371
61. Nagin DS. Group-based trajectory modeling: An overview. *Ann Nutr Metab.* 2014;65:205-210. doi:10.1159/000360229
62. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol.* 2010;6:109-138. doi:10.1146/annurev.clinpsy.121208.131413
63. Nagin D. *Group-Based Modeling of Development.* Cambridge, Massachusetts: Harvard University Press; 2005.
64. Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: A comparison between methods. *J Clin Epidemiol.* 2012;65(10):1078-1087. doi:10.1016/j.jclinepi.2012.04.010
65. Durepos P, Sciences HH, Alsbury K, Hewston P, Borges J. Older Adults ' Perceptions of Frailty Language : A Scoping Review. *ResearchGate.* 2020;(January). doi:10.13140/RG.2.2.27686.06722
66. Suijker JJ, Rijn M van, Riet G ter, Charante EPM van, Rooij SE de, Buurman BM. Minimal Important Change and Minimal Detectable Change In Activities Of Daily Living In Community-living Older People. 2017;21(2):165-172.
67. Brown JC, Huedo-Medina TB, Pescatello LS, Pescatello SM, Ferrer RA, Johnson BT. Efficacy of exercise interventions in modulating cancer-related fatigue among adult cancer survivors: A meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2011;20(1):123-133. doi:10.1158/1055-9965.EPI-10-0988
68. Mänty M, Mendes De Leon CF, Rantanen T, et al. Mobility-related fatigue, walking speed, and muscle strength in older people. *Journals Gerontol - Ser A Biol Sci Med Sci.* 2012;67 A(5):523-529. doi:10.1093/gerona/glr183
69. Eldadah BA. Fatigue and Fatigability in Older Adults. *PM R.* 2010;2(5):406-413. doi:10.1016/j.pmrj.2010.03.022
70. Zengarini E, Ruggiero C, Pérez-Zepeda MU, et al. Fatigue: Relevance and implications in the aging population. *Exp Gerontol.* 2015;70:78-83. doi:10.1016/j.exger.2015.07.011
71. Kim I, Hacker E, Ferrans CE, Horswill C, Park C, Kapella M. Evaluation of fatigability measurement: Integrative review. *Geriatr Nurs (Minneapolis).* 2018;39(1):39-47. doi:10.1016/j.gerinurse.2017.05.014
72. Van Seben R, Reichardt LA, Essink DR, Van Munster BC, Bosch JA, Buurman BM. "I Feel Worn Out, as if I Neglected Myself": Older Patients' Perspectives on Post-hospital Symptoms after Acute Hospitalization. *Gerontologist.* 2019;59(2):315-326. doi:10.1093/geront/gnx192
73. Xue QL. The Frailty Syndrome: Definition and Natural History. *Clin Geriatr Med.* 2011;27(1):1-15. doi:10.1016/j.cger.2010.08.009
74. Avlund K. Fatigue in older adults: an early indicator of the aging process? *Aging Clin Exp Res.* 2010;22(2):100-115. <http://www.ncbi.nlm.nih.gov/pubmed/20440097>. Accessed January 20, 2017.
75. Avlund K, Rantanen T, Schroll M. Tiredness and subsequent disability in older adults: The role of walking limitations. *J Gerontol A Biol Sci Med Sci.* 2006;61(11):1201-1205. <http://www.ncbi.nlm.nih.gov/pubmed/17167163>. Accessed January 20, 2017.
76. De Boer D, Bos N, Zuidgeest M. *Ontwikkelingen in Het Meten En Gebuiken van Patiëntervaringen En Patiëntgerapporteerde Uitkomsten.* Utrecht; 2018. doi:ISBN: 978-94-6122-000-0
77. Santaguida P, Dolovich L, Oliver D, et al. Protocol for a Delphi consensus exercise to identify a core set of criteria for selecting health related outcome measures (HROM) to be used in primary health care. *BMC Fam Pract.* 2018;19(1):1-14. doi:10.1186/s12875-018-0831-5
78. Linnean Initiatief. <https://www.linnean.nl/default.aspx>. Published 2017. Accessed November 13, 2020.
79. Pel-Littel RE, Hofman CS, Yu L, et al. Recommendations of older adults on how to use the PROM 'TOPICS-MDS' in healthcare conversations: A Delphi study. *PLoS One.* 2019;14(11):1-17. doi:10.1371/journal.pone.0225344
80. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: Evidence for a phenotype. *Journals Gerontol Ser a-Biological Sci Med Sci.* 2001;56(3):M146-M156. doi:10.1093/gerona/56.3.M146
81. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *Cmaj.* 2005;173(5):489-495. doi:10.1503/cmaj.050051

82. Katz S, Akpom CA. 12. Index of ADL. *Med Care*. 1976;14(5 Suppl):116-118.
83. Glynn NW, Santanasto AJ, Simonsick EM, et al. The Pittsburgh fatigability scale for older adults: Development and validation. *J Am Geriatr Soc*. 2015;63(1):130-135. doi:10.1111/jgs.13191
84. Laan W, Zuithoff NPA, Drubbel I, et al. Validity and reliability of the Katz-15 scale to measure unfavorable health outcomes in community-dwelling older people. *J Nutr Health Aging*. 2014;18(9):848-854. doi:10.1007/s12603-014-0479-3

CHAPTER 8

Summary

Nederlandse samenvatting

About the author

Dankwoord (Acknowledgements)

Research Institute SHARE

SUMMARY

Societal context

Newspapers are full of the consequences of the so-called 'aging society', 'Gray Wave' or 'Silver Tsunami'. These metaphors mean that the number of older people in relation to the working population is increasing exponentially. Within the population aged 65 and older, the proportion of people aged 80 years and older will increase even more rapidly. This accelerated aging of society is not only predicted for the coming decades in the Netherlands, but also worldwide. The often negative connotation of the accelerated aging is referring to its associated expected problems. The increasing number of older people relative to a younger generation means there will be fewer people available to take care of the older generation. Consequently, this stresses the health care system and with that the burden on society. One solution to reduce this burden is to enable older people to live as independently and as healthy for as long as possible. On the one hand this meets the needs of the majority of the elderly themselves, on the other hand it ensures that care remains manageable (in terms of workforce) and affordable (in terms of costs) and thus reducing the burden on healthcare. One way to enable people to live as independently and healthy for as long as possible is to implement interventions to maintain or regain an optimal health status in older people. However, in order to do so, there is work to be done: First, persons who are at high risk to develop poor health status should be identified in order to target preventive interventions; Second, it should be investigated which determinants distinguish people with a high risk of poor health status from people with a low risk of poor health status in order to develop targeted interventions; Third, to evaluate the effectiveness of interventions requires valid, reliable, and responsive instruments. This dissertation aims to contribute to these questions by examining psychometric properties of multiple existing instruments used to measure change in health status of older people (Chapters 2, 3 and 4) and by identifying trajectories of health status of older people and to investigate the determinants of these different trajectories (Chapters 5 and 6).

Measuring change in health status of older people

In geriatrics, functional status and frailty status are two important outcomes that are frequently used to measure the health status in older people. Functional status is often measured by counting the number of activities of daily living (ADL) a person can perform without assistance. The Katz ADL Index is the most commonly used instrument among older hospitalized patients. To measure frailty status, two instruments are widely used: 1. The 'Frailty Phenotype', a screening instrument to measure physical frailty, and 2. The 'Frailty Index', which measures the frailty status of older people across multiple domains, such as physical, social, and mental functioning, symptoms or diseases. The Katz ADL Index, the Frailty Phenotype and the Frailty Index are already used in scientific research to measure change in functional or frailty status over time. However, literature on the responsiveness of these instruments is sparse. Without information about the responsiveness, it is difficult to determine whether a measured difference actually represents a change in functional or frailty status. Therefore, in this dissertation two studies were devoted to investigate the validity and responsiveness of these common used instruments of functional and frailty status.

In Chapter 2 the agreement between the change in functional status experienced by the patient and the change in the number of ADL limitations, measured with the Katz ADL Index from pre-hospital admission to three and twelve months post discharge was investigated. Agreement was low (<50%) over both time frames. These results give reason to doubt the usefulness of this instrument to measure change in functional status over time that is relevant to the individual patient.

In Chapter 3 the reproducibility and responsiveness were investigated of the Frailty Index and the Frailty Phenotype. Again, the time frames from pre-hospital admission up to three and twelve months after hospitalization have been included. The Frailty Index showed good reproducibility and responsiveness over both three and twelve months timeframes. The Frailty Phenotype performed less well, as indicated by poor to moderate reproducibility and a sufficient responsiveness over a timeframe of 12 months only.

Chapter 4 was devoted to the Dutch translation and validation of the Pittsburgh Fatigability Scale (PFS) in a sample of older hospitalized patients. The PFS aims to measure mental and physical fatigue while performing physical activities, which is defined as fatigability. The Dutch version of the PFS appeared a valid and reliable instrument for measuring mental and physical fatigability in hospitalized patients aged 70 years and older. Future research should prove whether the Dutch version of the PFS is also suitable for other groups, such as community-dwelling older people, and to what extent a change measured with the PFS represents meaningful change in fatigability to older people themselves.

Based on the results of the first three chapters of this dissertation, supplemented by existing literature, in Chapter 7, the general discussion, recommendations were formulated for the use of the investigated instruments to measure change in health status over time. First, based on the results of Chapter 2, supported by previous research that demonstrated that the minimal important change according to the patient is lower than the smallest detectable change in community-dwelling older people, it was recommended that the Katz ADL Index can only be used at group level. Second, we recommend to prefer the Frailty Index over the Frailty Phenotype to measure change in frailty status over time among older hospitalized people at group level. Third, we recommend the Dutch version of the PFS to measure physical and mental fatigability of hospitalized older people, including change in fatigability over time at group level.

Trajectories of health status and their determinants in older people

In order to support individual people with a high risk of poor health status, for example by providing behavioral interventions, we need to identify the subgroup of individuals having the highest risk of developing poor health status. Trajectory modeling provides the opportunity to identify clusters of individuals within a given population. Once these different subpopulations have been identified, characteristics of individuals within a subpopulation can be compared with characteristics of individuals belonging to another subpopulation. In this way, determinants of persons with a high versus low risk of poor health status can be distinguished. In this dissertation, determinants of trajectories of generic health

status (general self-rated health, SRH) (Chapter 5) and trajectories of fatigability and mobility were investigated (Chapter 6).

In Chapter 5, four stable trajectories of SRH over a five-year period were identified. Persons belonging to the least favorable SRH trajectory were more often female and lower educated and they more often had two or more chronic conditions, an unhealthy lifestyle, and abnormalities in three or more physiological markers, of which the latter is indicative for an early stage of frailty.

In Chapter 6 three different trajectory combinations of fatigability and mobility from hospital admission up to six months after discharge were identified. Individuals belonging to the least favorable trajectory, the high fatigability and low mobility group, were more likely to have a lower comorbidity score and a higher Frailty Index before hospital admission. However, the results of Chapter 6 should be interpreted with caution due to the small sample size.

Based on the results of Chapters 5 and 6, in Chapter 7, the general discussion, it was concluded that certain determinants that can be traced back to frailty have a negative impact on health status trajectories of older people. The presence of multiple chronic conditions and (an early stage of) frailty have been identified as determinants of the least favorable health status trajectories in both community-dwelling and hospitalized older people. If these determinants could be modifiable and slow down the deterioration of health status, lifestyle behaviors appeared a potential starting point for the design of interventions. Previous research demonstrated that older people with multimorbidity or frailty may benefit from interventions aimed at improving lifestyle and with that improving health status. In order to further optimize the costs and benefits of such targeted interventions, it is important to identify even more precisely which individuals can benefit most from such interventions. Previous research has shown an association between increased fatigability and negative health outcomes in older people, such as functional decline. Chapter 6 also demonstrated that the course of fatigability was associated with the course of mobility of older hospitalized patients. Consequently, we recommend to design future studies to investigate the added value of fatigability as a screening tool

for the identification of individuals who are most likely to benefit from interventions to delay or improve their frailty status. In addition, the relationship between fatigability and frailty should be investigated in more detail, as well as how to design interventions aimed to improving fatigability in older people in general.

NEDERLANDSE SAMENVATTING

Maatschappelijke context

Het zal u niet ontgaan zijn: kranten staan vol van de gevolgen van de zogenaamde grijze golf of, mooier verpakt, de zilveren tsunami. Met deze metafoor wordt bedoeld dat het aantal ouderen ten opzichte van de werkende bevolking toeneemt, wat niet alleen de verwachting is voor de komende decennia in Nederland, maar ook wereldwijd. De veelal negatieve connotatie van deze termen komt door de daarmee gepaard gaande verwachte problemen. Ons sociale stelsel is zo ingericht dat de werkende bevolking de lasten draagt van de niet werkende bevolking, waaronder ouderen vallen. Met een grotere niet werkenden / werkenden ratio neemt de lastendruk op de werkende bevolking toe. Een van de maatregelen om deze lastendruk te verlagen is door mensen in staat te stellen zo lang mogelijk zelfstandig en in goede gezondheid te leven. Enerzijds komt dit tegemoet aan de behoeften van de meerderheid van de ouderen zelf, anderzijds zorgt het ervoor dat de zorg beheersbaar (qua personeelsbestand) en betaalbaar blijft en draagt op die manier bij aan verlaging van de lastendruk op onze werkende bevolking. Een manier om mensen in staat te stellen zo lang mogelijk zelfstandig en in goede gezondheid te leven is door interventies te ontwerpen en te implementeren om de optimale gezondheidstoestand bij ouderen te behouden of te herwinnen. Echter om hiertoe over te kunnen gaan is het nodige voorwerk vereist: ten eerste dient onderzocht te worden welke personen tot een hoog risico groep behoren om gericht interventies te kunnen toepassen; ten tweede dient onderzocht te worden wat mensen met een hoog risico op een slechte gezondheidstoestand onderscheidt van mensen met een laag risico op een slechte gezondheidstoestand om aanknopingspunten te verkrijgen voor het ontwikkelen van gerichte interventies; ten derde is het noodzakelijk om valide, betrouwbare en responsieve instrumenten te gebruiken om de effectiviteit van interventies te beoordelen. Met dit proefschrift is beoogd om een deel van dit voorwerk te doen, door psychometrische eigenschappen te onderzoeken van verschillende bestaande instrumenten die worden gebruikt om verandering in de gezondheidstoestand van ouderen te meten (Hoofdstukken 2, 3 en 4) en door trajecten van gezondheidstoestand van ouderen in kaart te brengen en de determinanten van deze verschillende trajecten te onderzoeken (Hoofdstukken 5 en 6).

Het meten van verandering in de gezondheidstoestand van ouderen

In de ouderenzorg zijn functionele status en kwetsbaarheidsstatus (frailty status) twee belangrijke uitkomsten die gebruikt worden om gezondheidstoestand van ouderen te meten. Functionele status wordt veelal gemeten door te tellen hoeveel activiteiten van het dagelijkse leven (ADL) iemand zonder hulp kan uitvoeren. De Katz ADL Index is daarbij het meest gebruikte instrument bij oudere ziekenhuispatiënten. Om de kwetsbaarheidsstatus van ouderen in kaart te brengen worden in de wetenschap twee instrumenten veel gebruikt: 1. Het 'Frailty Phenotype' instrument, een screeningsinstrument om fysieke kwetsbaarheid van ouderen te meten, en 2. De 'Frailty Index', die de kwetsbaarheid van ouderen in kaart gemeten over meerdere domeinen, zoals fysiek functioneren, sociaal functioneren, mentaal functioneren of symptomen of ziektes. Zowel de Katz ADL Index, het Frailty Phenotype instrument als de Frailty Index worden reeds gebruikt bij wetenschappelijk onderzoek om verandering in functionele dan wel kwetsbaarheidsstatus over de tijd te meten. In de literatuur is echter nog weinig bekend over de zogeheten responsieve eigenschappen van deze instrumenten, dat wil zeggen het vermogen van een instrument om verandering over de tijd te meten. Daarom zijn in dit proefschrift twee studies uitgevoerd om de validiteit en responsiviteit van deze instrumenten te onderzoeken.

In Hoofdstuk 2 is onderzocht of de door de patiënt ervaren verandering in functionele status overeen komt met de gemeten verandering van het aantal ADL beperkingen, gemeten met de Katz ADL Index. Hierbij werden twee tijdsbestekken meegenomen: de functionele status vóór de ziekenhuisopname tot drie maanden na ziekenhuisopname en de functionele status vóór de ziekenhuisopname tot twaalf maanden na ziekenhuisopname. Over beide tijdsbestekken was de overeenkomst laag (<50%). Deze resultaten geven reden tot twijfel over de bruikbaarheid van dit instrument om verandering in ervaren functionele status van een patiënt over de tijd te meten. In Hoofdstuk 3 zijn de reproduceerbaarheid en responsiviteit onderzocht van de Frailty Index en het Frailty Phenotype instrument. Wederom zijn de tijdsbestekken van de situatie vóór de ziekenhuisopname tot drie en twaalf maanden na de ziekenhuisopname meegenomen. De Frailty Index bleek over een goede reproduceerbaarheid en responsiviteit te beschikken over beide tijdsbestekken. De Frailty Phenotype instrument presteerde minder goed: een slechte tot matige

reproduceerbaarheid en een voldoende responsiviteit over een tijdsbestek van 12 maanden.

Hoofdstuk 4 was gewijd aan een (tot nog toe) minder gebruikt instrument in de ouderenzorg: de 'Pittsburgh Fatigability Scale (PFS)'. Deze vragenlijst heeft als doel om vermoeidheid tijdens het uitvoeren van fysieke activiteiten, zogeheten vermoeibaarheid, te meten en is door ons in het Nederlands vertaald en gevalideerd bij een groep oudere ziekenhuispatiënten. De Nederlandse versie van de PFS bleek een valide en betrouwbaar instrument om mentale en fysieke vermoeibaarheid bij ziekenhuispatiënten van 70 jaar en ouder te meten. Of deze Nederlandse versie van de PFS ook geschikt is voor andere groepen, zoals thuiswonende ouderen en in hoeverre een verandering gemeten met de PFS ook betekenisvol is voor de oudere zelf zal toekomstig onderzoek moeten uitwijzen.

Op basis van de resultaten van de eerste drie hoofdstukken van dit proefschrift, aangevuld met bestaande literatuur, zijn, in Hoofdstuk 7, de algemene discussie, aanbevelingen geformuleerd voor het gebruik van de onderzochte instrumenten om verandering in gezondheid van ouderen over de tijd te meten. Ten eerste, op basis van de slechte overeenkomst tussen ervaren verandering en gemeten verandering in functionele status bij oudere ziekenhuispatiënten en omdat uit eerder onderzoek is gebleken het kleinste betekenisvolle effect voor de patiënt binnen de kleinst meetbare verandering valt bij thuiswonende ouderen, werd aanbevolen om de Katz ADL index alleen op groepsniveau te gebruiken. Ten tweede is aanbevolen om de Frailty Index boven het Frailty Phenotype instrument te verkiezen om kwetsbaarheidsstatus over de tijd te meten bij ouderen. Ook voor deze instrumenten geldt de aanbeveling ze alleen op groepsniveau te gebruiken. Ten derde hebben we geconcludeerd op basis van de resultaten van de test-hertest gegevens dat de Nederlandse PFS kan worden gebruikt om verandering in vermoeibaarheid over de tijd op groepsniveau te meten.

Trajecten en determinanten van de gezondheidstoestand van ouderen

Om mensen met een hoog risico op gezondheidsproblemen te ondersteunen om hun risico te verlagen door bijvoorbeeld gedrags interventies toe te passen, dienen we te onderzoeken welke mensen tot de hoog risico

groep behoren. Traject analyse is hiervoor een geschikte methode omdat het binnen een gegeven groep subpopulaties in kaart brengt. Zodra verschillende subpopulaties binnen een groep gemodelleerd zijn, kunnen kenmerken van individuen binnen een subpopulatie worden vergeleken met kenmerken van individuen die tot een andere subpopulatie behoren. Op deze manier worden determinanten van bijvoorbeeld personen met een 'hoog risico op gezondheidsproblemen' zichtbaar. In dit proefschrift zijn determinanten van trajecten van generieke ervaren gezondheid (general self-rated health, SRH) (Hoofdstuk 5) en trajecten van vermoeibaarheid en mobiliteit onderzocht (Hoofdstuk 6).

In Hoofdstuk 5 werden vier stabiele trajecten van SRH over een tijdsbestek van vijf jaar geïdentificeerd. Personen die toebehoorden aan het minst gunstige SRH-traject waren vaker vrouw en hadden vaker een lagere opleidingsniveau, twee of meer chronische aandoeningen, een ongezonde leefstijl en afwijkingen in drie of meer fysiologische markers, wat een indicatie is voor een voorstadium van kwetsbaarheid.

In Hoofdstuk 6 werden drie verschillende trajectcombinaties van vermoeibaarheid en mobiliteit vanaf ziekenhuisopname tot zes maanden na ontslag geïdentificeerd. Personen die toebehoorden aan de minst gunstige traject, namelijk hoge mate van vermoeibaarheid en een lage mobiliteit, hadden vaker een lagere comorbiditeitscore en een hogere kwetsbaarheidsstatus vóór de ziekenhuisopname. De resultaten van Hoofdstuk 6 dienen echter met voorzichtigheid geïnterpreteerd te worden vanwege het kleine aantal deelnemers aan deze studie.

Gebaseerd op de resultaten van Hoofdstukken 5 en 6 is in Hoofdstuk 7, de algemene discussie, geconcludeerd dat bepaalde determinanten die terug te voeren zijn op kwetsbaarheid een negatieve invloed op gezondheidstrajecten bij ouderen hebben. De aanwezigheid van meerdere chronische aandoeningen en (een voorstadium van) kwetsbaarheid zijn geïdentificeerd als de meest voorkomende determinanten van de minst gunstige gezondheids trajecten bij zowel thuiswonende ouderen als oudere ziekenhuispatiënten. Leefstijl kan een mogelijk aanknopingspunt vormen voor de inrichting van interventies. Uit eerder onderzoek blijkt dat ouderen met multimorbiditeit of kwetsbaarheid mogelijk kunnen profiteren van interventies die gericht zijn op het verbeteren van

leefstijl en daarmee de gezondheidstoestand. Om de baten en kosten van dergelijke gerichte interventies verder te optimaliseren, is het belangrijk om nog nauwkeuriger de juiste groep mensen te identificeren die het meest kunnen profiteren van dergelijke interventies. Eerder onderzoek laat een verband zien tussen vermoeibaarheid en negatieve gezondheidsuitkomsten van ouderen, zoals functionele achteruitgang. Ook uit Hoofdstuk 6 blijkt dat de mate van vermoeibaarheid geassocieerd is met het verloop van mobiliteit van oudere ziekenhuispatiënten. Op basis hiervan is aanbevolen dat toekomstig onderzoek gedaan dient te worden naar de meerwaarde van vermoeibaarheid als screeningsinstrument voor de identificatie van ouderen die mogelijk kunnen profiteren van interventies om hun kwetsbaarheidsstatus te verbeteren. Daarnaast pleiten we voor toekomstige studies die de relatie tussen vermoeibaarheid en kwetsbaarheid nader onderzoeken alsmede welke interventies het meest geschikt zijn om vermoeibaarheid bij ouderen te verbeteren.

ABOUT THE AUTHOR

Marlies Feenstra was born on March 21, 1985 in Roden. In 2005, she started her study in Human Movement Sciences at the University of Groningen. As part of the bachelor's degree, she investigated the effects of observing walking movements on executive functions in dementia, supervised by prof. dr. Erik Scherder. In 2009 she continued her study with the master Movement, Aging, and Health. During the master's program, she set up an experimental study on modulations of circadian rhythms in Alzheimer's disease, supervised by dr. Mathieu de Greef and dr. Marijke Gordijn. In the meantime, she expanded the master's program with a teacher training course, for which she did her internships at the School of Nursing and Medical Imaging and Radiation Therapy at the Hanze University of Applied Science in Groningen.

After graduating, she started working at the 'Instituut voor Zorgprofessionals' as a trainer, scientist, and project officer in several research and educational projects in the post-secondary education field. In December 2015, she started as a PhD candidate at the department of Geriatrics of the University Medical Center Groningen. During her PhD, she co-supervised several students, assisted in practical statistical courses, participated at several national and international conferences, and she started to obtain her registration as Epidemiologist B. As part of the training program of the Graduate School of Medical Sciences, she completed several courses on epidemiology and statistics. She also went to the University of Massachusetts in Amherst for the course Group-Based Trajectory Modeling for the Medical and Social Sciences taught by Daniel Nagin. This resulted in the research methods used in chapters five and six of this dissertation.

She currently works as sr. data analyst in the Value Based Health Care (VBHC) team of the Medisch Spectrum Twente in Enschede. VBHC is a management strategy behind the Santeon Together Better Program. To this end, for multiple medical disorders, data are structurally collected and analyzed with the aim to continuously learn and improve healthcare by informing and empowering patients. Data are shared and compared between the seven Santeon hospitals, after which improvements are initiated and the effects are monitored in the local hospitals. She is also involved in strengthening networking and cooperation between science, innovation and VBHC within the Medisch Spectrum Twente.

DANKWOORD (ACKNOWLEDGEMENTS)

Dit dankwoord heeft nogal wat teweeg gebracht in de afgelopen jaren. In zekere heeft dit dankwoord, ver voordat er een letter van op papier stond, samen met mijn eigen motivatie en toewijding om dit project tot een goed einde te brengen, bergen verzet. Met het schrijven van dit dankwoord leg ik de laatste hand aan mijn proefschrift wat mij de gelegenheid biedt om de mensen te bedanken die me in de afgelopen vijf jaren gesteund en geholpen hebben bij de totstandkoming van dit proefschrift en alles wat daarbij kwam kijken.

Ten eerste mijn promotieteam. Sophia, door een team te creëren van onderzoekers en onderzoeksmedewerkers met diverse achtergronden en persoonlijkheden heb je een mooie basis gecreëerd om mijzelf als jonge onderzoeker verder te ontwikkelen. Je leerde me onder andere dat een dergelijk team met verschillende persoonlijkheden verder komt. Daarnaast heb ik me door jou gestimuleerd gevoeld om zelf richting te geven aan mijn professionele en persoonlijke ontwikkeling. Je staat altijd open voor nieuwe ideeën en bent bereid ze te faciliteren als er een gedegen plan achter zit. Zo maakte je het mogelijk dat ik naar Amerika kon voor een cursus, deel kon nemen aan internationale congressen en de cursus spirometrie kon volgen. Door jouw overstijgende denken wist je de onderdelen van mijn stukken vaak weer te verbinden waardoor ik de rode draad weer zag en verder kon. Dank voor je vertrouwen. Nynke, onze samenwerking laat zich enigszins kenmerken door de uitdrukking 'zonder weerstand geen groei'. Ondanks de wrijving die er soms was heb je me gestimuleerd om het beste uit mezelf te halen, voor de Epidemiologie registratie te gaan en om door te zetten. Jouw uitgebreide feedback en vragen hebben mijn stukken inhoudelijk versterkt. Dank voor je lessen op persoonlijk en epidemiologisch vlak. Barbara, naarmate mijn traject vorderde hebben we meer contact gekregen. Mijn stukken zijn door jouw klinische blik en ervaring een stuk leesbaarder geworden, omdat je me stimuleerde om na te denken over de toepassing van mijn onderzoeksresultaten in de kliniek. De vraag 'Wat kunnen we hier als dokters mee?' zou elke medisch wetenschappelijk onderzoeker zich moeten stellen bij het schrijven van een artikel. Dank voor deze onmisbare aanvulling.

Bianca Buurman, dank dat je me hebt gewezen op de vacature voor dit promotietraject in Groningen. Ik weet niet zeker of je ook nog een goed woordje voor me hebt gedaan bij Sophia, maar je hebt zeker de eerste aanzet gegeven, waarvoor ik je erg dankbaar ben.

Dank ook aan de leden van de beoordelingscommissie, te weten Tibor Hortobágyi, Job van der Palen en Marielle Emmelot-Vonk voor het lezen en beoordelen van dit manuscript.

Dear co-authors, Janet, Nancy, Frederike en Carolien, thank you for the fruitful and inspiring collaboration, I have found your contributions to some chapters in this thesis very informative and appreciate your effort.

Collega's van Z3.13, ons toevluchtsoord: Ruth, Leroy, Maaïke, Maartje, Carolien, Kiki en Fiona. Wat hebben we het getroffen met ons team van betrokken collega's met goede baking skills! Ruth, jij ontfermde je over de jonge onderzoekers als een moeder en fungeerde tegelijkertijd als levend anticonceptiemiddel door je vreselijke, maar geweldige verhalen over je vier fantastische kinderen. Dank dat je me zo zorgvuldig hebt ingewerkt voor het Seniorlines onderzoek. Carolien, Kiki en Fiona, jullie wil ik ook graag apart bedanken. Ten eerste voor de vriendschap die tussen ons is ontstaan in de afgelopen jaren en voor jullie bijdrage op de inhoud van mijn proefschrift: Carolien, wat heb je een werk verzet om al onze studies in Castor te krijgen. Ik vond het erg leuk om samen met jou de analyses van het artikel in hoofdstuk 3 te doen en heb bewondering voor je gestructureerde, consciëntieuze manier van werken en je doorzettingsvermogen. Kiki, heel fijn dat we samen op kon trekken tijdens onze promotietijd door onder andere elkaars introducties en discussies te lezen. Ik waardeer je erg als collega en nu als vriendin, waarbij je altijd klaar staat om als klankbord te dienen of om een grappige anekdote te vertellen om problemen te relativeren. Fijn dat we elkaars paranimfen kunnen zijn! Fiona, dank voor je hulp bij de vertaling van de Pittsburgh Fatigability Scale uit hoofdstuk 4 en voor het meedenken over de casus uit de introductie. Jij brengt vaak reuring, maar neemt ook tijd voor goede, diepgaande gesprekken. Fijn dat je mijn paranimf wilt zijn.

Dank aan alle vaste en tijdelijke medewerkers die Seniorlines draaiende hebben gehouden en in het bijzonder de mensen die voor Seniorlines 2.0

zich de benen uit het lijf hebben gelopen: Adriëlla, Luka, Marloes, Saskia en Maud, dank voor de gezellige tijd en jullie harde werken om de data op alle meetmomenten, door heel Nederland, op tijd te verzamelen.

Lieve vrienden, dank voor jullie betrokkenheid, interesse en steun tijdens mijn promotietraject. Lieve Ank, vriendin, klankbord, wat ben ik blij dat we elkaar getroffen hebben op de opleiding bewegingswetenschappen. Als vriendin waardeer ik het enorm dat je altijd in bent voor leuke nieuwe dingen, dat je altijd een luisterend oor hebt en interesse toont in wat mij bezig houdt. Nu je met Eduard bent doen we de leuke uitjes en vakanties gewoon met zijn drieën. Ank en Eduard, dank jullie wel voor alle waardevolle momenten van ontspanning en gezelligheid! Lieve Grietje, onze samenwerking onder 'Margriet Lifestyle' hebben we helaas moeten stoppen omdat ik aan mijn promotietraject begon. Gelukkig is de vriendschap die we hebben opgebouwd wel gebleven. Ik vond het fijn dat ik je kon helpen met jouw opleiding en ik ben je heel dankbaar voor je hulp bij het structureren van de algemene introductie. Het is altijd heerlijk om bij jou en René thuis in het bos te komen en lange wandelingen te maken of samen een dagje cultuur te snuiven en ik hoop dat we dat nog lang blijven doen. Ronald, dank dat ik stevast alle zaterdagen en vaak ook ten minste één avond lekker kon ontspannen en uitwaaien met jouw paarden.

Zuster en broeders van Al Anon: Helena, Janneke, Lammert en Gea in het bijzonder. Van jullie leerde ik het onveranderlijke te aanvaarden, het veranderlijke (waar wenselijk) te veranderen en deze twee uitersten van elkaar te onderscheiden. Dank voor jullie steun en dit mooie inzicht dat me op menig dieptepunt helpt relativeren en nieuw perspectief biedt.

Lieve pap en mam, wat voel ik me gezegend met jullie onvoorwaardelijke liefde en steun en de diepe band die we hebben. Mam, wat fijn dat je altijd naar me wilt luisteren als ik een theorie of een abstract idee aan je uitleg om vervolgens door jouw vragen zelf beter te begrijpen waar ik naar toe wil. Ik vind het heerlijk om samen met jou op pad te gaan: op vakantie, naar de sauna, wandelen, winkelen, klussen; het maakt niet uit wat, met jou is het altijd gezellig en vertrouwd. Pap, wat fijn dat we elke zondagochtend met de MTB de bossen van met name Gasselte en Appelscha onveilig maakten. Mooi dat we deze vader-dochter momentjes hadden in de afgelopen jaren, ze hebben me wekelijks veel energie gegeven. Maar ook

nu het samen fietsen niet meer gaat vinden we ongetwijfeld weer een nieuwe gezamenlijke interesse die we samen kunnen doen. Lieve pap en mam, ik hou van jullie!

Yvonne, lieve tante, al sinds mijn jonge jaren hebben wij een bijzondere band. Je bent er altijd om een helpende hand, een luisterend oor of een oplossing te bieden. Dank je wel dat ik een tijd lang in het huisje aan het meer mocht wonen, ik kon me geen fijnere plek wensen om te zijn op dat moment. Lieve Yvonne, jouw vermogen om overal kansen en mogelijkheden te zien zijn inspirerend voor me. Ik hoop dat je straks volop van het omaschap kan genieten en dat we samen, net als we met oma deden, altijd gekke dingen blijven doen.

Lieve opa en oma, ondanks dat jullie mijn promotie niet meer mee kunnen maken, horen jullie in dit dankwoord thuis. Wat ben ik blij met de mooie herinneringen die we de afgelopen jaren nog hebben kunnen maken: de vakantie naar Ibiza, met oma de slappe lach hebben, de vertrouwensband die we hadden en de goede gesprekken die daar uit voort kwamen zelfs tot aan jullie sterfbed toe. Jullie zijn altijd een inspiratiebron voor me geweest. Wat voel ik me gezegend met jullie als voorouders.

Ten slotte wil ik graag mijn dank uiten aan alle Seniorlines deelnemers: dank voor het vertrouwen dat wij van Seniorlines tijdens een voor u vaak ingrijpende periode een stukje met u op mochten lopen, voor de gastvrijheid tijdens de huisbezoeken en voor de inspirerende levenslessen die ik van u mee kreeg. Dankzij u is dit proefschrift met daarin een nieuw puzzelstukje aan kennis als co-creatie ontstaan waardoor we de wereld (van de ouderenzorg) een klein beetje beter hebben gemaakt.

RESEARCH INSTITUTE SHARE

This thesis is published within the **Research Institute SHARE** (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen. Further information regarding the institute and its research can be obtained from our internet site: <http://www.share.umcg.nl/>

More recent theses can be found in the list below.

(supervisors are between brackets)

2021

Nanninga CS

Home & place making after stroke; exploring the gap between rehabilitation and living environment

(prof K Postema, prof LB Meijering, dr AT Lettinga, dr MC Schönherr)

06.01.2021 EXPAND

2020

Wassink-Vossen S

Physical activity and functional recovery in late-life depression

(prof RC Oude Voshaar, dr P Naarding, dr R Collard)

Hovenkamp-Hermelink A

The long-term course of anxiety disorders; an epidemiological perspective

(prof RA Schoevers, dr H Riese, dr B Jeronimus)

Moazzen S

Nutrients and diet quality in gastrointestinal cancers

(prof GH de Bock, dr BZ Alizadeh)

Poonsiri J

Exploring cycling and sports in people with a lower limb amputation: prosthetic aspects

(prof JHB Geertzen, prof PU Dijkstra, prof R Dekker, dr JM Hijmans)

Rausch CP

Geriatric syndromes prevalence: associated factors and outcomes

(prof U Bültmann, prof SEJA de Rooij, prof L Laflamme, dr J Möller)

Shahabeddin Parizi A

Self-reported health status after solid-organ transplantation

(prof PFM Krabbe, prof SJL Bakker, prof E Buskens, dr KM Vermeulen)

Vervoort D

Adaptability of gait and balance across the adult lifespan

(dr CJC Lamoth, prof T Hortobagyi, dr N Vuillerme, dr AR den Otter)

Munck L de

Breast cancer: screening, stage and outcome; studies based on the Netherlands Cancer Registry

(prof GH de Bock, prof S Siesling)

Wijnen A

Rehabilitation policies following total hip arthroplasty; across borders

(prof SK Bulstra, prof D Lazovic, dr M Stevens)

Spinder N

Maternal occupational exposure and congenital anomalies

(prof HM Boezen, prof H Kromhout, dr HEK de Walle, dr JEH van Kammen-Bergman)

Driel-de Jong TJW van

Factors associated with the persistence of medically unexplained symptoms in later life

(prof RC Oude Voshaar, prof JGM Rosmalen, Dr PH Hilderink, dr DJC Hanssen)

Timkova V

Self-reported health outcomes in patients with obstructive sleep apnoea; unraveling the role of bio-psycho-social factors

(prof U Bültmann, prof R Tkacova, dr JP van Dijk, dr I Nagyova)

Alma HJ

Discovering the dynamics of the minimal clinically important difference of health status instruments in patients with chronic obstructive pulmonary disease

(prof T van der Molen, prof R Sanderman, dr C de Jong)

Hylkema TH

Total knee arthroplasty among working-age patients

(prof S Brouwer, prof SK Bulstra, dr M Stevens, dr PPFM Kuijer)

Donk LJ van der

Cancer survivors' experience with depressive symptoms and their (low) need for psychological care; Lessons learned from a multi-center randomized controlled trial

(dr MJ Schroevers, dr J Fleeer, prof R Sanderman)

Hovenkamp-Hermelink A

The long-term course of anxiety disorders; an epidemiological perspective

(prof RA Schoevers, dr H Riese, dr B Jeronimus)

Blikman T

Neuropathic-like symptoms in hip and knee osteoarthritis

(prof SK Bulstra, dr M Stevens, dr I van den Akker-Scheek)

For earlier theses visit our website.