Developing and testing a patient-reported outcome measure for post-stroke fatigue – a mixed-methods approach

Doctoral Thesis

Ingrid Johansen Skogestad

Institute of Health and Society

Faculty of Medicine

University of Oslo

2023

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-348-0278-2

Cover: UiO.
Print production: Graphic center, University of Oslo.

Table of contents

Acknowledgement

Many people and institutions have contributed to making this project a reality. I am deeply grateful to Lovisenberg Diaconal Hospital (LDS) and the user organization LHL-stroke (Landsforeningen for Hjerte- og Lungesyke) for providing the opportunity and support to conduct this research. Their collaboration and resources have made it possible for me to work on this project between 2017 and 2023. I also want to extend my thanks to the National Association for Public Health, LDS research legacy, and Kirsten Rønnings legacy for funding this Ph.D. position.

Most importantly, I express my sincere gratitude to all the patients and health professionals who have participated in our study. Your time, experiences, and knowledge have been invaluable and have allowed me to expand my understanding of post-stroke fatigue. I am genuinely grateful for your contribution to this work.

I want to express my profound thanks to Professor Anners Lerdal, my main supervisor, who believed in me from the very beginning and for fostering an environment of intellectual freedom. Thank you for sharing your extensive network and knowledge of fatigue and for always bringing new ideas to the table. I would also like to thank my co-supervisor and head of the Muni-Health-Care research school, Professor Marit Kirkevold. Your extensive knowledge of qualitative research and enthusiasm for exploring our data has been a true inspiration. Thank you also for letting me in as a student at Muni-Health-Care, which enabled me to develop essential academic skills. I also thank my co-supervisor, Professor Bent Indredavik, for sharing valuable insight into the clinical stroke research field and inviting me to visit and observe clinicians at the stroke unit at St. Olavs Hospital. In addition to my team of supervisors, I am deeply grateful to have had the opportunity to work with a talented group of academics, including Christine, Anders, and Caryl. Thank you for all your contributions to this project.

Special thanks go to Therese and Hilde. You have been invaluable user consultants in this project. Your contributions have shaped this thesis's outcome and nurtured a professional collaboration that extends well beyond this work. I am grateful and honored for the opportunity to co-speech lectures with Hilde and to work alongside Therese, who has

been the principal founder, developing an innovative fatigue application on external funds in collaboration with our team at LDS.

I would also like to extend my sincere gratitude to Gro Mikkelsen, the head nurse of the stroke unit at LDS. Thank you for always encouraging me and providing the support and resources I needed to succeed in my role.

The day-to-day support of my fellow Ph.D. students, Unni and Petra, during our shared academic journey has been invaluable. Your motivation and comfort have helped me persevere through all the ups and downs, and you have significantly increased the quality of my life during these years. Furthermore, I would like to acknowledge my co-workers and friends at Unger Vetlesen Institute and Villa Viten. Thank you all for being supportive and friendly, and stimulating good academic discussions.

I also want to direct gratitude towards my friends and family. Knowing that I always had your support has given me tremendous strength and motivation to keep going.

Finally, I would like to express my deep appreciation for Jonas. My best friend, husband, and father to our children. Words will never be sufficient to describe your importance to me during this work. Not only have you provided everyday happiness and perspective, but you have also given me extensive academic support through countless discussions and in-depth scrutiny of all the details in my thesis. I will be forever grateful.
To my children, Sebastian, Selma Marie, and Ludwig – may your love of learning and curiosity lead you to greatness.

Thank you,
Oslo, February 2023
Ingrid Johansen Skogestad

Summary

Background

Post-stroke fatigue (PSF) is a highly prevalent and debilitating condition that negatively impacts stroke survivors' daily functioning, rehabilitation, and overall quality of life. Despite its high prevalence and significant impact on patients and society, there is a lack of evidence-based treatments for PSF. One major barrier to further progress is the lack of a clear definition and agreement on measuring and diagnosing PSF. Current research uses patient-reported outcome measures (PROMs) with several shortcomings, including not being developed following current PROM development guidelines.

Aims

The overall aims of this thesis were to explore PSF and fatigue PROMs used in stroke research and to use this knowledge to develop and evaluate measurement properties of a new PSF PROM adhering to current PROM development guidelines.

Methods

This thesis followed a sequential exploratory mixed-methods design. First, we reviewed 11 fatigue PROMs used in stroke research and assessed the 156 items with content analysis and descriptive statistics. Then, individual interviews with stroke survivors (n=9) and focus groups with health professionals (n=16) were conducted to explore PSF, and content analysis was applied to develop a definition and conceptual framework of PSF. Based on these findings, an expert panel (n=7) developed the Norwegian Fatigue Characteristics and Interference Measure (FCIM). The content validity of FCIM was tested in cognitive interviews (n=15) with stroke survivors. Finally, responses from a cross-sectional survey (n=169) of stroke survivors were subject to a Rasch analysis investigating the structural validity and internal consistency of FCIM.

Results

The content analysis of the 156 items in the eleven PROMs in this study revealed 83 *unique* items. The PROMs lacked items relevant to a stroke population, contained items irrelevant to a stroke population, were not comprehensive and had substantial in-between variations

and differences in item content. The analysis of individual interviews and focus group transcripts revealed four themes illustrating a conceptual framework of PSF: characteristics of PSF, interfering and aggravating factors, management, and awareness of PSF. The expert panel used the conceptual framework to develop an initial version of the Norwegian Fatigue Characteristics and Interference Measure (FCIM) version 1.0. This version comprised of a definition of PSF, a 10-item characteristics subscale, and a 20-item interference subscale with a seven-day recall period. In addition, FCIM included two pre-stroke fatigue items. Cognitive interviews assessed content validity, resulting in revising, "flagging", and removing some items. The participants reported that FCIM provided a comprehensive evaluation of PSF without any redundant items. The cognitive interviews resulted in FCIM version 2.0 with an updated PSF definition, ten characteristic items, 18 interference items, and two pre-stroke fatigue items. The following Rasch analysis of data from the cross-sectional survey resulted in the removal of additional ten items based on misfit analysis, local dependency, and issues with content validity. The final FCIM with a 6-item characteristics subscale and a 12-item interference subscale demonstrated good structural validity and internal consistency.

Conclusion

In this thesis, we found that existing fatigue PROMs used in stroke populations have significant shortcomings and do not comply with the best practice guidelines and high standards set by regulators. As a result, we developed a new PROM specifically targeted for measuring PSF. Based on findings from several qualitative studies, FCIM has a clear definition and conceptual framework. Further, FCIM has good content validity, structural validity, and internal consistency. Future studies need to evaluate the construct validity, responsiveness, and test-retest reliability of FCIM.

Summary in Norwegian

Bakgrunn

Utmattelse (fatigue) etter hjerneslag er et vanlig og plagsomt symptom som negativt påvirker evnen til å utføre aktiviteter i dagliglivet, rehabilitering og livskvalitet. Til tross for høy prevalens og betydelige konsekvenser for den enkelte, finnes det dessverre ingen effektiv behandling. Et viktig hinder er mangelen på en klar definisjon og enighet om hvordan man best kan kartlegge og diagnostisere utmattelse etter hjerneslag. Eksisterende pasientrapporterte utfallsmål (PROMs) som brukes i forskning og klinisk praksis har flere mangler, og de har ikke blitt utviklet i henhold til nyere retningslinjer for PROM-utvikling.

Hensikt

Den overordnede hensikten med denne studien var å utforske utmattelse etter hjerneslag og PROMs brukt i hjerneslagsforskning, og å bruke denne kunnskapen til å utvikle og evaluere måleegenskapene til et nytt PROM som kartlegger utmattelse hos hjerneslagpasienter. Dette instrumentet skal utvikles i henhold til gjeldende PROM-retningslinjer.

Metode

Denne avhandlingen benyttet både kvalitative og kvantitative metoder i et sekvensielt eksplorativt mikset metode-design. I en oversiktsartikkel analyserte vi 156 spørsmål hentet fra 11 PROMs som var blitt brukt til å måle utmattelse hos hjerneslagpasienter. Deretter gjennomførte vi individuelle intervjuer med slagpasienter (n=9) og fokusgrupper med helsepersonell (n=16). Innholdsanalyse ble benyttet for å utvikle en definisjon og et konseptuelt rammeverk for utmattelse etter hjerneslag. Basert på disse funnene utviklet et ekspertpanel (n=7) instrumentet: Fatigue Characteristics and Interference Measure (FCIM). Innholdsvaliditeten i FCIM ble testet med kognitive intervjuer (n=15) med hjerneslagpasienter. Data fra en tverrsnittstudie (n=169) med hjerneslagpasienter var deretter brukt til å evaluere strukturell validitet og intern konsistens ved FCIM gjennom en Rasch analyse.

Resultater

Innholdsanalysen av de 156 spørsmålene i de elleve instrumentene identifiserte 83 unike spørsmål. Våre analyser viste at instrumentene inneholdt irrelevante spørsmål, samt at de manglet både relevante spørsmål og en helhetlig kartlegging av utmattelse etter hjerneslag. Innholdsmessig var det liten grad av overlapp mellom PROMs. Innholdsanalyse av individuelle intervjuer og fokusgrupper førte til utvikling av et konseptuelt rammeverk hvor utmattelse etter hjerneslag besto av fire dimensjoner: 1) karakteristika ved utmattelse, 2) konsekvenser av og faktorer som trigget utmattelse, 3) mestring av utmattelse og 4) gradvis erkjennelse av utmattelse. Ekspertpanelet tok utgangspunkt i dette konseptuelle rammeverket ved utvikling av FCIM versjon 1.0. Denne versjonen bestod av en definisjon av utmattelse, 10-spørsmål i subskalaen *karakteristika* og 20 spørsmål i subskalaen *konsekvenser av utmattelse*, samt to spørsmål om utmattelse før hjerneslaget. Innholdsvaliditeten ble evaluert gjennom kognitive intervjuer med hjerneslagpasienter og resulterte i revidering, «flagging» og fjerning av noen spørsmål. Deltagerne beskrev FCIM som et helhetlig instrument uten overflødige spørsmål. Analyse av de kognitive intervjuene ledet til FCIM versjon 2.0. Denne versjonen ble testet i en tverrsnittstudie med påfølgende Rasch-analyse. Denne analysen avdekket metodiske svakheter ved noen av spørsmålene som deretter ble fjernet. Den oppdaterte versjonen av FCIM hadde totalt 20 spørsmål og begge subskalaene demonstrerte god strukturell validitet og intern konsistens.

Konklusjon

Denne avhandlingen har avdekket store svakheter med eksisterende pasientrapporterte utfallsmål benyttet for å kartlegge utmattelse hos hjerneslagpasienter. Disse instrumentene oppfyller ikke kravene til validitet og reliabilitet som man i dag stiller til denne typen pasientrapporterte utfallsmål. Gjennom flere kvalitative studier har vi derfor utviklet et nytt instrument som er tilpasset hjerneslagpasienter. FCIM er basert på en klar definisjon og konseptuelt rammeverk, og våre funn viser at FCIM har god innholdsvaliditet, strukturell validitet og intern konsistens. Før instrumentet kan tas i bruk, kreves flere studier for en ytterligere evaluering av instrumentets validitet, reliabilitet og responsivitet.

List of papers

1. Skogestad IJ, Kirkevold M, Indredavik B, Gay CL, Lerdal A; NORFAST (Norwegian Study of Fatigue After Stroke) Group. Lack of content overlap and essential dimensions - A review of measures used for post-stroke fatigue. J Psychosom Res. 2019 Sep;124:109759. doi: 10.1016/j.jpsychores.2019.109759. Epub 2019 Jul 3. PMID: 31443803.

2. Skogestad IJ, Kirkevold M, Larsson P, Borge CR, Indredavik B, Gay CL, Lerdal A. Post-stroke fatigue: an exploratory study with patients and health professionals to develop a patient-reported outcome measure. J Patient Rep Outcomes. 2021 Apr 21;5(1):35. doi: 10.1186/s41687-021-00307-z. PMID: 33881660; PMCID: PMC8060374.

3. Skogestad IJ, Kottorp A, Larsson P, Moen TM, Gay CL, Borge CR, Lerdal A. Development and testing of the Norwegian Fatigue Characteristics and Interference Measure (FCIM) for stroke survivors: cognitive interviews and Rasch analysis. [revised and re-submitted to Quality of Life Research]

Abbreviations

COSMIN	Consensus-based Standards for the Selection of Health Measurement Instrument
CTT	Classical Test Theory
FCIM	Fatigue Characteristics and Interference Measure
FDA	U.S. Food and Drug Administration
FSS	Fatigue Severity Scale
IRT	Item Response Theory
PRO	Patient-Reported Outcome
PROM	Patient-Reported Outcome Measure
PSF	Post-stroke fatigue

1.0 Introduction

Having an acute stroke changes life within seconds. The blood supply to a part of the brain is disrupted, which leads to damage and, eventually, death of brain cells. Immediate medical care is necessary to reduce the stroke impact, and optimized treatment and rehabilitation are vital to enhancing recovery [1]. After days or even months, with hospitalization and rehabilitation, most patients undergo significant functional improvement [2]. However, many patients experience life-changing post-stroke disabilities and symptoms, with post-stroke fatigue (PSF) being the most common symptom three months after stroke [3]. Stroke survivors rate fatigue as the second most crucial research priority [4].

PSF is a disabling condition that significantly affects stroke survivors' quality of life, rehabilitation, functioning, and ability to return to work [5-7]. The onset and trajectory vary, but patients often have to live with PSF as a permanent symptom [8, 9]. There are no evidence-based treatments for PSF, despite its high prevalence and tremendous impact on patients and society [10, 11]. A major obstacle to further research on interventions to prevent and treat PSF is the lack of consensus on defining, conceptualizing, measuring, and diagnosing fatigue [12]. To our knowledge, we lack a patient-reported outcome measure (PROM) that assesses the clinically relevant aspects of PSF and adheres to current PROM development guidelines.

Consequently, research currently uses various PROMs to assess PSF [13, 14]. Unfortunately, these instruments have significant disadvantages; most importantly, they lack content validity in stroke survivors [15]. Thus, the overall aims of this thesis were to explore PSF and fatigue PROMs used in research and to use this knowledge to develop and evaluate the measurement properties of a new PSF PROM.

2.0 Background

2.1 Stroke

2.1.1 Epidemiology and diagnosis

Every year, 12.2 million people worldwide have a first-time stroke, and more than 100 million people are living with a prior history of stroke [16]. In Norway, around 10.000 people are admitted to a hospital due to stroke each year [17]. Ischemic strokes cause 85% of all stroke cases, and intracerebral hemorrhage causes around 15% [17]. These Norwegian prevalence rates are similar to those in other high-income countries [18].

The International Classification of Diseases (ICD-11) has, in its newly updated version, defined stroke as a neurological disease [19]. The defining characteristic of a stroke is impaired perfusion through the blood vessels to the brain and reduced oxygen supply, which eventually lead to necrosis [1, 20]. Depending on the site and size of the stroke, loss of neuronal function can lead to many different symptoms and disabilities.

Symptoms of a stroke include sudden onset of numbness or weakness in the face, arm, or leg, trouble speaking, confusion, vision disturbances, dizziness, coordination problems, or loss of balance. These focal neurological symptoms can also occur with several other conditions. The key to differentiating stroke from stroke mimics (as well as identifying stroke subtypes) is brain imaging by CT scan or MRI [21]. When a correct diagnosis has been established, corresponding treatment can follow. Introducing thrombolysis and thrombectomy and implementing specialized stroke units have significantly improved the acute treatment for cerebral infarction. These advances in treatment have reduced mortality and disability rates [18]. For hemorrhagic stroke, there are no specific disease-modifying treatments. The acute phase treatments include lowering systemic blood pressure and intracranial pressure, unfortunately, with little impact on the prognostic variables [21]. However, the overall mortality rates for stroke are declining, and globally the population is growing and aging. Thus, the number of people living with functional impairments, disabilities, and reduced health due to stroke continuously increases, as stroke is the third-leading cause of lost disability-adjusted life years (DALYs) [18].

2.1.2 Post-stroke disabilities and functioning

Having a stroke can result in a wide range of disabilities and reduced functioning, having an impact on independence and quality of life [2]. The most significant improvement occurs in the first three to six months post-stroke. However, recovery can occur at any stage [2]. The timeline in stroke recovery is often divided into the hyperacute and acute phase (0-7 days), early subacute phase (7 days to 3 months), late subacute phase (3 to 6 months), and chronic phase (>6 months) [22]. After a stroke, biological, psychological, and socio-environmental factors influence recovery [23, 24]. A variety of impairments in body functions (e.g., orientation, attention, language, and physical functions) can have an impact on a patient's ability to perform daily activities and participate in social life (e.g., walking, doing housework, maintaining relationships with family and friends). Further, environmental, and personal factors can serve as facilitators or barriers to functioning and participation (e.g., help and recognition from next of kin) [23, 25].

At any point during stroke recovery, symptoms, such as fatigue, can be a significant threat to the overall rehabilitation process [26]. PSF may impede recovery, as PSF is associated with poorer physical outcomes and reduced quality of life [27, 28], which can lead to poor long-term outcomes [28]. Unfortunately, health professionals often neglect or overlook PSF, leaving the patients overwhelmed in their new situation [26]. The overall quality of quantitative PSF research is said to be poor [28], and significant research is needed to understand the etiology of PSF [29].

2.2 Post-stroke fatigue

Even though the first study on PSF was published as late as 1999 [30], fatigue-related symptoms in general have been reported in humans since the mid-19th century. However, the pathophysiology and associated factors of PSF remain unclear [31].

2.2.1 Historical perspectives

There have been long-lasting discussions concerning what constitutes and causes fatigue, and historical insight can shed light on the evolution of research. Fatigue is a recent term previously called *neurasthenia*. An American neurologist, George Beard, was one of the first

to describe neurasthenia as a distinct clinical entity that often occurred in upper-class women reporting symptoms of exhaustion, muscle weakness, and pain. It was a popular diagnosis that, for a while, was associated with neurological and other somatic disorders. However, during the 1900-century, neurasthenia as a diagnosis was reclassified as a mental disorder associated with anxiety, hysteria, and depression [32].

This view of fatigue as a mental disorder prevailed until relative recently. In the ICD-10, published in 1992 and used until 2022, the diagnosis of "Neurasthenia" was still classified as a mental and behavioral disorder [33]. However, the recently published ICD-11 reflects the contemporary view of fatigue and includes two essential changes from the ICD-10. Firstly, the term *neurasthenia* is replaced with *fatigue*. Secondly, fatigue is classified as a symptom unrelated to a specific body system. As such, fatigue is neither classified as a mental nor a physical condition [34].

2.2.2 Definition and multidimensionality

A major problem in PSF research is the lack of a clear definition. Due to the subjective nature of fatigue, the diagnosis and assessment of PSF are based on patients' self-reports. Different studies have defined PSF in various ways, and there is no consensus among researchers and clinicians on how to define PSF [28]. The only case definition for PSF is developed by Lynch et al. This case definition aims to detect clinically significant fatigue by interviewing patients with a pre-set list of questions. To fulfill the case definition, patients need to indicate that they 1) have experienced fatigue every day or nearly every day for the last 2 weeks in the past month and 2) the fatigue interferes with activities in their daily life. Based on this interview, PSF is determined to be either present or not present [35]. Whether a patient fulfills the case definition is based on subjective judgments by the interviewer, and it lacks an unambiguous definition of PSF.

From a measurement perspective, it is vital to clearly define the concept being measured [36]. One of the reasons why PSF is difficult to define is due to its multidimensionality and complexity [37]. PSF has been described as a symptom that is expressed in different ways and has different factors contributing to its experience [38]. Qualitative studies are the

preferred method for investigating symptom experience [39], yet only a few original qualitative studies have explored PSF as their main aim [40-48]. Other studies with different overall aims have reported fatigue as a secondary finding [49-53]. Central findings in these studies were the consequences of fatigue and different strategies for coping [42, 46-48, 51]. The feeling and experience of PSF need to be better studied, and none of these studies explored PSF with a later aim to develop a PROM.

A prerequisite for PROM development is to have a conceptual framework. A conceptual framework can be defined as "a model representing the relationships between the items and the construct to be measured" [36], and various aspects of the concept are preferably depicted in a diagram or model [36, 54]. A qualitative meta-synthesis developed a model displaying the three main themes of stroke survivors' experiences with PSF and how these themes were related [55]. The first theme was PSF's core characteristics, including lack of energy, abnormal need for sleep, being quickly tired, unpredictable feelings, and increased stress sensitivity. In addition, having PSF was described as invisible and distressing due to it being an undefined condition. The other theme was the acknowledgment of PSF from significant others, and the third theme was coping with PSF. However, the authors of this meta-synthesis highlight the need for more research, especially on the experience of PSF across the illness trajectory [55]. When measuring a complex symptom such as fatigue, the conceptual framework must be developed specifically for that PROM [36]. Without a clear definition and conceptual framework, it is challenging to adequately quantify a concept and develop items that genuinely reflect the concept it intends to measure [56].

2.2.3 Prevalence and trajectory

The lack of a clear definition of PSF is one reason for the considerable variation in estimates of its prevalence, which range between 25% and 85% [29]. Other explanations for this variation include the use of different diagnostic criteria and outcome instruments, and assessing PSF at different times since the stroke. Other methodological factors such as country/region or exclusion criteria (e.g., excluding patients with depression and cognitive impairment) might also add to the inconsistent results [57]. Meta-analyses have estimated the pooled prevalence rates for PSF between 48% (95% CI 42–53) [57] to 50% (95% CI 43–

57) [58]. However, both studies had substantial heterogeneity (I^2>90%), and results should be interpreted with caution. The geographical location could explain some of the variation across studies. Nevertheless, opposite results were found regarding depression and time since stroke as variables explaining the heterogeneity [57, 58]. The pooled prevalence rates were only based on studies assessing PSF with Fatigue Severity Scale (FSS). Thus, the reported heterogeneity between the included studies might be related to the lack of reliability of FSS itself [35].

The same variability is seen in studies investigating the trajectory of PSF. Some studies have reported that the proportion of fatigue decreases over time [59, 60]. Other studies have shown fatigue as stable [61] or increasing [62]. Despite some variable results in these specific details, fatigue seems to be a persistent and prevalent symptom reported up to 6 years post-stroke [8, 9]. Unfortunately, we lack effective treatment for this disabling condition [63].

2.2.4 Treatment and pathophysiology

Several reviews have scrutinized different pharmacological and non-pharmacological treatments tested in clinical intervention studies [37, 63-65]. Both antidepressants [66, 67] and Modafinil [68] (central nervous system stimulant, traditionally used to treat daytime sleepiness) have been tested in randomized clinical trials. Currently, there is insufficient evidence for the efficacy of any pharmacological treatment [63, 69]. A wide range of non-pharmacological interventions has also been tested, including, cognitive behavioral therapy [70], mindfulness [71], educational programs [72], and graded activity training [73]. Similar to the drug clinical trials, no significant differences have been found to support the effectiveness of any of these interventions [63-65]. One of the challenges in optimizing treatment is the need for more knowledge regarding PSF's pathophysiological mechanisms.

Several pathophysiological mechanisms have been proposed and are under study in stroke survivors. A well-investigated hypothesis is a causal link between PSF and stroke type, lesion location, or stroke severity. However, there is no robust evidence for such a relationship [13, 28]. Cortical excitability has been suggested as related to PSF, and one study proposed that

reduced response from cortical neurons contributed to high levels of fatigue [74]. Fatigue is prevalent in patients with autoimmune diseases and is often reported post-infection in otherwise healthy individuals. Hence, a causal link to inflammation has been suggested, but this hypothesis lacks substantial evidence and needs further investigation [28]. Despite the various proposed pathophysiological mechanisms, PSF may also be triggered and maintained by psychological, socio-environmental, and biological factors. Thus, a suggestion is to view the maintenance of PSF from a biopsychosocial perspective [13, 75].

2.3 Factors associated with PSF

Several symptoms and disorders are prevalent after a stroke [25, 76]. Some of these factors are reported to be associated with and have attributes overlapping with PSF [77, 78]. Under the evolution of a conceptual framework for PROM development, it is crucial to be aware of similar but distinct symptoms and disorders. In item development, these constructs should be differentiated from the symptom of interest [56]. The most important symptoms and disorders associated with PSF are depression, cognitive impairment, and physical impairment [69].

2.3.1 Depression

Depression is common after a stroke and strongly relates to PSF [13, 69]. According to ICD-11, *depressive disorders* can be defined as having a: "depressive mood (e.g., sad, irritable, empty) or loss of pleasure accompanied by other cognitive, behavioral, or neurovegetative symptoms that significantly affect the individual's ability to function". Such symptoms include difficulty concentrating, feelings of worthlessness, hopelessness, changes in appetite or sleep, and reduced energy or fatigue [34]. A *depressive disorder* is a clinical diagnosis that must meet strict diagnostic requirements. However, *symptoms of depression* can be assessed with various self-report instruments, which enable the monitoring of patients' symptoms and their impact on daily life. A meta-analysis reported a 29% (95% CI 25-32, I^2 = 93.9) pooled prevalence rate for post-stroke depressive disorder, which remained stable during the first ten years after the stroke [79]. Post-stroke depressive disorders are substantially higher than pre-stroke prevalence rates. A meta-analysis of stroke samples found a pooled prevalence rate of pre-stroke depressive disorders (collected from medical

records) of 9.4% (95% CI 6.2-14.0) [80]. In comparison, the prevalence rate for depressive disorders alone was 3.2%, and in patients with one chronic physical condition, it ranged from 9.3-18.1% depending on the specific comorbid disease [81].

Fatigue and lack of energy are general symptoms of depression, and it can be challenging to differentiate between them [69]. Several studies have shown a significant association between PSF and depressive disorders and symptoms. However, post-stroke fatigue is also present in stroke survivors without depression [69, 82-85]. Fatigue is often assessed in PROMs measuring symptoms of depression [86], and commonly-used fatigue PROMs contain items assessing aspects of depression such as reduced interest and motivation [87-89]. A recent review showed that several factors suggest that PSF and depression are two distinct clinical entities [28]. First, PSF is more prevalent than post-stroke depression. Second, antidepressants have been shown to effectively treat post-stroke depression but have no significant effect on fatigue symptoms [66, 67].

2.3.2 Cognitive impairment

Dementia and mild cognitive impairment are also commonly reported after a stroke, and fatigue is more frequently reported in people with dementia than cognitively non-impaired persons [90, 91]. According to ICD-11: “dementia is characterized by the presence of marked impairment in two or more cognitive domains relative to that expected given the individual’s age and general premorbid level of cognitive functioning, which represents a decline from the individual’s previous level of functioning.” Mild cognitive impairment is the transitional stage between normal aging and dementia. In contrast with dementia, only one domain needs to be affected in mild cognitive impairment, and mild cognitive impairment does not affect activities in daily living [34, 92]. Dementia is common in the elderly population and is also causally linked to stroke. A study found that the prevalence of post-stroke dementia was 41% and a study of older stroke patients revealed that up to 66% had mild cognitive impairment three months after stroke [93, 94].

Common cognitive symptoms of dementia and mild cognitive impairment are problems with concentration, orientation, attention, and memory. The most common psychological

symptoms are apathy, depression, and anxiety [95]. These symptoms of cognitive impairment are similar to some characterizations of PSF, described as reduced mental capacity, lack of interest, and poor motivation [13]. Additionally, these symptoms also overlap with symptoms of depression [34]. In developing a conceptual framework for a fatigue PROM, the boundaries between symptoms of fatigue and cognitive impairment must be carefully considered [56].

2.3.3 Physical impairments

Physical impairments are widespread consequences after a stroke and affect up to 80% of patients. Post-stroke physical impairments are characterized by limitations or complete loss of function in muscle control and movement [96]. This limitation in mobility can range from total paralysis to minor weakness or numbness and lead to activity limitations and constraints in participation [25, 96]. During the first 3-6 months, patients improve an average of 70% of their lost physical functions [2].

Several reviews show that PSF is associated with reduced physical functioning. However, the stroke size does not predict PSF, and PSF might be the only sequela in patients with fully recovered physical functions. In addition, not all patients with physical impairment experience PSF [28, 69]. PSF is often described as a lack of physical (or mental) energy that reduces the ability to perform physical activity [28]. Although PSF can interfere with physical activity, PSF and reduced physical functioning are two different post-stroke sequelae. Despite this, many existing fatigue PROMs include items referring to physical functions directly limited by the stroke. Such items can falsely inflate fatigue scores [12].

2.3.4 Pre-stroke fatigue

Fatigue is a prevalent symptom experienced across various diseases [32] and is also common in healthy populations [97]. Further investigation is needed to determine whether fatigue is a unified experience across diseases. One qualitative meta-synthesis investigated similarities and differences in the fatigue experience between PSF and cancer-related fatigue and found both similarities and differences in the fatigue experience [98]. Similar to PSF, the pathophysiology of fatigue in other diseases is not well known, and it is unclear if

the same underlying cause leads to fatigue in all conditions [32]. A variety of potential causal mechanisms are proposed for the different diseases. For example, fatigue in irritable bowel disease is hypothesized to result from nutritional deficiency, inflammation, or altered metabolism [99]. Cancer fatigue is suggested to be caused by cytokine dysregulation, circadian rhythm disruption, or anemia [100]. Fatigue is also common post-COVID and is suggested to be caused by pathological inflammation [101].

2.4 Fatigue PROMs

The knowledge base in PSF research is poor, and the results are often inconsistent, with a high risk of bias [29, 31, 63]. A potential, often overlooked, bias in PSF research is the validity of fatigue instruments in diagnosing, evaluating therapy, and predicting the future course of PSF [12].

Measuring health outcomes is essential to deciding how to best care for patients. Traditionally, researchers use measures of morbidity and mortality to track health outcomes. However, patient-reported outcomes (PROs) are increasingly recognized as an essential basis for patient-centered care [54] and provide the only reasonable strategy for evaluating the treatment impact of symptoms such as fatigue [102]. The definition of a PRO is: “any report of the status of a patient’s health condition that comes directly from the patient without interpretation of the patient’s response by a clinician or anyone else” [54]. Instruments or questionnaires, called patient-reported outcome measures (PROMs), are used to measure the PROs. In this thesis, the discussed PROMs are restricted to self-reported questionnaires for adults and interchangeably referred to as either instruments or PROMs.

More than 50 fatigue PROMs have been identified in the literature [103]. The items in these PROMs are often developed without apparent patient involvement and then tested in people with conditions such as multiple sclerosis, chronic fatigue syndrome, or cancer [12, 87]. Research on PSF has applied fatigue PROMs with the general assumption that these instruments can produce trustworthy findings on PSF [12]. However, several reviews have concluded that we lack a high-quality PROM to adequately assess fatigue in stroke survivors

[12, 37, 104, 105]. This lack of a superior PROM constitutes a barrier to progress and prevents the development of specific PSF treatments [105]. For example, in drug therapy development, the U.S. Food and Drug Administration (FDA) only supports labeling claims based on PROMs developed according to PROM development guidelines [54]. These guidelines represent the current state-of-the-art for PROM development, as presented in the next section.

2.4.1 Guidelines for PROM development

PROMs are often claimed in general terms to be "valid" and "reliable." *Validity* can be defined as "the degree to which an instrument truly measures the construct it purports to measure," and *reliability* can be defined as "the degree to which the measurement is free from measurement error" [106]. The validity and reliability aspects of a PROM are overall called measurement properties. Multiple measurement properties need to be evaluated before using a PROM in a study, and it is impossible to establish every form of validity and reliability in one test [36]. Assessing the validity of a PROMs is complicated, considering the subjective nature of PROs and the lack of a proper gold standard [36]. Several guidelines suggest different criteria to assist the development and critical appraisal of a PROM's measurement properties.

Widely used PROM development guidelines include the S*tandards for educational and psychological testing* (developed jointly by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education) [107] and the criteria developed by the Scientific Advisory Committee of the Medical Outcomes Trust [108]. Various academic disciplines, including psychology, medicine, and education, often concur with these standards [109]. However, these guidelines lack explicit requirements for what constitutes excellent measurement properties [110]. In addition, they are developed based on preference and expert opinion rather than evidence [36, 108].

With this in mind, Mokkink and colleagues performed an international Delphi study including around 50 experts, intending to achieve consensus on standards for evaluating the

methodological quality of studies on measurement properties for PROMs [111]. As a result, they developed *COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN)* [111]. The advantage of the COSMIN methodology is that it has a broad view of validity and reliability and is developed based on expert consensus in the field. In health care, it is advised to follow the COSMIN methodology [56]. Thus, this thesis adheres to the COSMIN methodology for instrument development.

In addition, this thesis also follows a guideline developed by the U.S. Food and Drug Administration (FDA) [36, 54]. A PROM used to measure treatment benefits in medical product clinical trials needs to be of exceptionally high quality if the aim is to support labeling claims approved by regulatory agencies such as the FDA or European Medicines Agency (EMA) [112]. To guide the process, the FDA has developed a guideline to support claims in medical product labeling ("*Guidance for Industry – Patient-reported outcomes measures: use in medical product development to support labeling claims")* [54]. FDA's guideline uses the same terminology and methodology as COSMIN and has equally high criteria for measurement properties of PROMs [54].

The COSMIN taxonomy divides the overall measurement properties into aspects of validity, reliability, responsiveness, and interpretability, where validity and reliability have several different sub-categories, as presented in Figure 1.

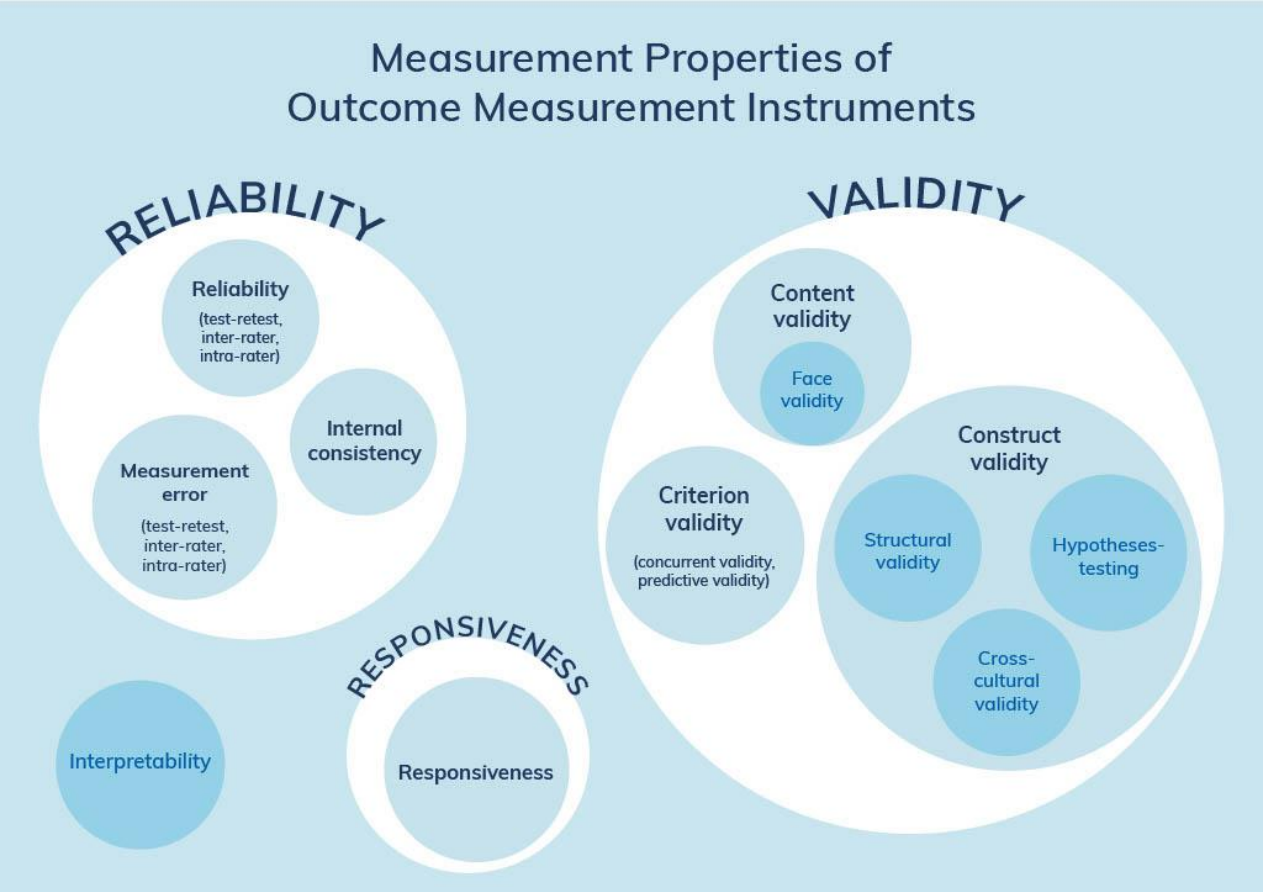


Figure 1. COSMIN taxonomy of measurement properties for health-related PROMs. Reprinted with permission from Elsevier [106].

2.4.2 Measurement properties of fatigue PROMs in stroke samples

Reviews of fatigue PROMs used in stroke populations conclude that no multidimensional instrument has been adequately tested in stroke survivors. In each of the currently used fatigue PROMs, only one or a few measurement properties have been assessed, which is insufficient in the views of COSMIN and FDA [12, 13, 38, 104, 105]. The current section aims to give a comprehensive view of the state of research concerning fatigue PROMs used in PSF research and display the knowledge gap at the start of our study. The following section will describe and define different types of PROMs measurement properties in detail and give examples that are relevant to PSF.

Content validity is viewed as the most crucial measurement property and can be defined as "the degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured" [106]. Qualitative studies with patients and other

stakeholders are the preferred method to establish content validity. First, a qualitative study can explore the concept and enable item generation. After developing the initial PROM, another qualitative study will further test the content validity by assessing whether all items are relevant and whether the concept is comprehensively covered by the items [36]. In existing PROMs, researchers and representatives from the target population can review content validity, assessing the relevance and comprehensiveness of the items [36]. The PROM must provide a clear definition and conceptual framework that is also reflected in the items. If a fatigue PROM lacks evidence of content validity, it may not be an accurate or reliable instrument for assessing the concept it intends to measure, which could lead to inaccurate results [36]. Content validity needs to be established before further assessment of validity and reliability, to avoid potentially assessing the reliability and validity aspects of a different concept [113].

Content validity, as well as any other measurement property, is disease-specific [36]. Many of the fatigue PROMs often used in stroke populations are generic (as opposed to disease-specific) PROMs developed for non-neurological conditions. These PROMs need an assessment of content validity before being used in a stroke population [36, 54, 104, 105]. The most commonly-used fatigue PROM in stroke research is the Fatigue Severity Scale (FSS) [87]. In addition, the Fatigue Assessment Scale (FAS) [88] and Short Form-36 Vitality (SF-36 Vitality) [114] are often used. However, these three instruments lacks an adequate assessment of content validity in stroke populations [12, 105]. These generic PROMs are developed based on theory or expert opinion, making them especially vulnerable to lacking relevance to the study population [36, 54]. The only fatigue PROMs developed with some input from stroke survivors are the Neurological Fatigue Index-Stroke [115] and Dutch Multifactorial Fatigue Scale [116]. The Neurological Fatigue Index (NFI) was initially developed for multiple sclerosis [117]. However, the content validity of the Neurological Fatigue Index in multiple sclerosis has been reviewed as limited [104]. In stroke survivors, the qualitative assessment of the Neurological Fatigue Index-Stroke was limited to an examination of the thematic similarities between six interviews with stroke survivors, compared to interviews with a multiple sclerosis sample. Five out of six interviewed stroke survivors were women, fatigue was not a criterion for inclusion, and assessment of the

items' relevance and comprehensiveness considering PSF was not assessed [115]. The Dutch Multifactorial Fatigue Scale consists of five sub-dimensions that were developed a-priori, and items within these dimensions were developed based on qualitative interviews with 14 patients having acquired brain injury (i.e., stroke or traumatic brain injury) [116]. However, content validity assessment and testing of the items after their development were not conducted as advised in PROM guidelines [36]. Even though content validity is rarely assessed, several studies evaluate other aspects of the validity and reliability of fatigue PROMs in stroke samples.

The most straightforward way to assess validity is with *criterion validity,* which concerns how well a measurement agrees with a gold standard. However, PROs, such as PSF, almost always lack a gold standard. In these situations, the assessment of *construct validity* is the preferred method [36]. Construct validity can be defined as "whether the instrument provides the expected scores, based on existing knowledge about the construct" [106]. There are three aspects of construct validity: structural validity, hypothesis testing, and cross-cultural validity. *Structural validity* can be defined as "the degree to which the scores of a measurement instrument are an adequate reflection of the dimensionality of the construct to be measured" [106]. Assessments investigate the dimensionality of an instrument and can also evaluate the quality of individual items, which is especially important in a development process [36]. Structural validity can be assessed with statistical analysis such as Rasch analysis (one type of item response theory) and factor analysis. Rasch analysis can provide a complex evaluation of the dimensionality and item characteristics in great detail [36]. Structural validity has only been assessed in a few fatigue PROMs in a stroke population: FSS [118], Neurological Fatigue Index-Stroke [115], Fatigue Assessment Scale [119], and Dutch Multifactorial Fatigue Scale [116].

If structural validity testing shows that a PROM contains several dimensions, this will guide the *hypothesis testing* of the instrument. Hypothesis testing assesses the instruments and their sub-dimensions' relationships to other instruments measuring similar or dissimilar constructs [36]. Hypothesis testing is equivalent to *convergent* and *discriminant validity* [39]. A prerequisite for hypothesis testing is the a-priory development of detailed and specific

hypotheses regarding the direction and size of each one of the correlations, together with sound justifications for these expectations [36]. Several studies have performed hypothesis testing of fatigue PROMs and their correlation to other fatigue PROMs or other constructs, such as symptoms of depression [103, 115, 116, 118-124]. However, a problem with the previously mentioned studies is the lack of explicitly reported hypotheses. For example, Mead et al. [103] and Visser Keizer et al. [116] only report the correlation with a general comment on whether it is low, moderate, or high.

The third important aspect of construct validity is *cross-cultural validity*, which is applicable whenever an instrument is translated from the original version and includes qualitative testing of the translation and the need for any cultural adaptations of the items. Brandal et al. have translated the Fatigue Assessment Scale into Swedish, reportedly with pre-testing in the target population, but did not provide additional information regarding this procedure [120]. Valko et al. validated the FSS in a Swiss cohort using the most common procedure of back-and-forth translation but without testing in the intended population [124]. Others have translated and tested the PROM in the general population [118], and some have translated the PROM without reporting the procedure [119, 122, 125].

The reliability of PROMs refers to three main aspects. First, *internal consistency* tests the degree of interrelatedness among the items in a PROM (or in a subscale) and is most commonly assessed with Cronbach's alpha [36]. Internal consistency calculation only requires one measurement; thus, Cronbach's alpha is often reported [103, 116, 118-121, 124]. Second, different *reliability parameters* can assess the relationship between a person's scores on the PROM over time; third, *measurement error parameters* assess the magnitude of the error [36]. Previous studies on the reliability of fatigue PROMs in a stroke population have assessed different aspects of reliability parameters. These results show limited reliability evidence [103, 115, 119, 120, 122-124] which are also confirmed by systematic reviews [12, 104, 105].

The overall ambition in medicine is to cure or reduce the symptom burden for patients. Thus, a PROM's ability to capture change over time is fundamental, especially in clinical

intervention studies and in medical product studies aiming to measure treatment benefits (or risks) to support labeling claims [36, 54, 105]. Instruments aiming to evaluate treatment must be responsive to change, called *responsiveness.* Responsiveness is an important measurement property in addition to validity and reliability. Like other measurement properties, responsiveness needs to be evaluated in the specific population and context of use [36]. Tseng et al. have reported responsiveness of the Visual Analog Fatigue Scale based on Effect Size and paired *t*-test [123]. However, these methods are not considered suitable parameters of responsiveness. Besides Tseng et al., responsiveness has generally not been reported in stroke populations, not even in the most commonly-used PROMs such as the FSS [12, 105].

2.5 Brief rationale for this thesis

PSF research routinely uses generic fatigue PROMs, which lack content validity evidence in general and may be less responsive to change [126]. Content validity is by far the most crucial measurement property [36, 54, 56, 127]. Evidence of other measurement properties will not overcome problems with content validity [54]. Just because a PROM is widely used, it does not guarantee that it has been appropriately developed or that all relevant measurement properties have been thoroughly assessed. The first initial step in instrument development is to gain knowledge of existing instruments, and it is advised to conduct a thorough review of relevant literature [56]. There is also a clear need for a specific definition and conceptual framework of PSF [28]. To provide valid and reliable results regarding diagnosis and treatment effects, PSF needs to be measured with a robust fatigue PROM developed explicitly for the stroke population, adhering to state-of-the-art guidelines.

3.0 Aims of the study

The overall aims of this thesis were to explore PSF and fatigue PROMs used in stroke research and to use this knowledge to develop and evaluate measurement properties of a new PSF PROM adhering to current PROM development guidelines.

Specific aims

Paper 1: To explore item relevance, missing items, and comprehensiveness in dimensions of PROMs used in PSF research and to identify similarities and differences through cross-comparison of PROMs.

Paper 2: Explore PSF as it is experienced by stroke survivors and described by health professionals to guide the future development of a new PSF-specific PROM. Further, the aim was to develop a definition and a conceptual framework of PSF.

Paper 3: Develop a PSF PROM and evaluate its content validity, structural validity, and internal consistency. Further, the aim was to develop an instrument that could measure fatigue characteristics and interference (in two subscales), have the potential to measure improvement (for use in intervention studies), assess the level and duration of fatigue experienced prior to stroke (i.e., pre-stroke fatigue) and be relatively short (feasible).

4.0 Methods

The main methods used in this thesis are briefly summarized in this section, and comprehensive descriptions of the methods are found in the attached manuscripts.

4.1 Study design

We used a sequential exploratory mixed-methods design as described by Creswell and Plano Clark [128] to address the overall aim, as well as the guidelines for the development of PROMs from COSMIN [36] and the U.S. Food and Drug Administration (FDA) [54].

This design, which is commonly used in instrument development, begins with qualitative exploration to ensure the content validity of the PROM. It then moves on to a quantitative stage, testing the PROM in a larger sample [128]. An overview of the studies in this thesis is shown in Table 1. Before developing the PROM, a comprehensive review of relevant literature is recommended to understand existing PROMs thoroughly [56]. COSMIN and FDA guidelines provide a clear and in-depth methodology to consider throughout the instrument development process, covering aspects such as creating a conceptual framework, generating items, and determining sample size [36].

Table 1. An overview of methods and data material included in this thesis.

	Study methods	**Data collection period**	**Analysis**
Paper 1	Review of PROMs (n=11)	Systematic literature search performed in May 2018	Content analysis and descriptive statistics
Paper 2	Individual interviews (n=9)	June 2017 – August 2018	Content analysis
	Focus groups (n=16)	November 2017 – May 2018	
Paper 3	Expert panel (n=7)	January – June 2020	Group discussions
	Cognitive interviews (n=15)	August – September 2020	Content analysis
	Cross-sectional survey (n=169)	January – March 2021	Rasch analysis

4.2 Review of fatigue PROMs used in PSF research (paper 1)

In paper 1, we conducted a review and content analysis of PROMs used to assess fatigue in PSF research. We systematically searched the literature for studies on PSF and found 24 different PROMs used across 78 studies. The most-used PROMs (n=8) and PROMs partly developed or tested in a stroke sample (n=3) were included for analysis. The literature search was performed following the PRISMA guidelines for reporting systematic reviews [129].

4.2.1 Data collection and analysis

In total, the 11 PROMs contained 156 items. These items were entered into NVivo software version 11, a tool for organizing and managing qualitative data [130]. We performed a qualitative content analysis to analyze the specific item content and compare individual items between instruments. Content analysis is a systematic approach suitable for categorizing words and phrases based on their content [131]. All items were compared to every other item, classifying them as either *similar* (identical or nearly identical wording) to one or more items or *single* and unique (no other items with substantially the same meaning). Two authors conducted the analysis independently and reached a consensus through discussion. The items were also categorized into dimensions and sub-dimensions. The findings from the content analysis were analyzed using descriptive statistics in SPSS version 26 to provide a summary of the data. The statistical overlap between the PROMs was calculated using the Jaccard similarity coefficient, which showed the degree of overlap between items in two PROMs [132].

4.3 Exploring PSF (paper 2)

Paper 2 used a qualitative, explorative design to gain an in-depth understanding of stroke survivors' experiences and health professionals' perspectives on PSF. The study was conducted using individual semi-structured interviews and focus groups. We followed the Consolidated Criteria for Reporting Qualitative Research (COREQ checklist), which ensured the proper reporting of all essential aspects of the research team, study design, and analysis [133].

4.3.1 Sample and setting

A total of 9 stroke survivors and 16 health professionals were included in this study. Participants were recruited using purposive sampling and aimed to recruit a heterogeneous sample of participants with diversity in age, gender, levels of physical impairment, communication disorders, and living accommodations. Inclusion criteria were the following: 1) had suffered a stroke within two years before enrollment, 2) met the diagnostic criteria for post-stroke fatigue defined by Lynch et al. [35], 3) had adequate mental and verbal ability, and 4) were at least 18 years of age. Five participants were recruited from a stroke user organization's Facebook page and four from a stroke outpatient clinic at a hospital in Oslo. Five interviews were conducted in the participants' homes and four at Lovisenberg Diaconal Hospital, Oslo, Norway.

Health professionals were purposively sampled from different levels of healthcare: a stroke unit at a hospital in Oslo, home healthcare services in Oslo municipality, and a rehabilitation hospital in eastern Norway. These sites represented specialized and general treatment settings for stroke patients. Only those with clinical experience treating stroke patients were eligible for inclusion, and a diverse range of health professionals was sought. Focus groups were held at the participants' workplaces during working hours.

4.3.2 Data collection and analysis

Interviews and focus groups lasted between 45 and 80 minutes, were audio-recorded, and followed a semi-structured interview guide. The health professionals were divided into three focus groups. All interviews were conducted and transcribed by the first author, with a co-moderator present in the focus groups. We performed an inductive content analysis of the data material, meaning that the codes, categories, and themes are derived from the data. The intent is to provide replicable and valid inferences from the material to develop a contextual framework of PSF [131]. Analysis was performed both simultaneously and after the completion of data collection. NVivo 12 was used to facilitate the analysis [130]. The first author independently coded all the material, and the second author coded half the material. Codes, sub-categories, categories, and themes were continuously discussed within the research team until agreement was reached.

4.4 Development and testing of FCIM (paper 3)

Paper 3 used a mixed-methods design with three iterative steps, following COSMIN guidelines [36]. The aim was to develop a PSF PROM and assess its content validity, structural validity, and internal consistency. The steps included item development, cognitive interviews, and a cross-sectional survey.

4.4.1 Samples and settings

Expert panel: The expert panel consisted of seven people invited to participate by the first author: three stroke researchers and clinicians, two researchers with expertise in PROMs, and two user consultants. The panel met regularly in face-to-face or online meetings and gave individual written comments on material provided by the first author.

Cognitive interviews: 15 stroke survivors were recruited from a stroke user organization in Norway (invitations were sent via text messages and e-mails and posted on the user organization's Facebook page). Participants were purposively sampled to ensure variability in age, gender, fatigue severity, and time since stroke. Inclusion and exclusion criteria were the same as in paper 2; however, to get participants with varying levels of fatigue, participants did not need to meet the diagnostic criteria for post-stroke fatigue defined by Lynch et al. [35]. Seven interviews were conducted at Lovisenberg Diaconal Hospital and eight in the homes of the participants.

Cross-sectional survey: A convenience sample of 169 stroke survivors completed a questionnaire that included our newly-developed PSF PROM. Participants was recruited through a stroke user organization (invitations posted on Facebook and Instagram and sent by e-mail). The inclusion criteria were self-reported stroke diagnosis and the ability to read Norwegian.

4.4.2 Data collection and analysis

Expert panel: Following COSMIN guidelines, the definition of PSF and conceptual framework established in our previous study served as the foundation for item development [36, 127].

Initially, a comprehensive list of 160 potential items was generated, covering all categories and sub-categories of the framework. Through detailed discussions, the list was reduced to a feasible size and selected based on relevance, comprehensibility, and comprehensiveness according to the conceptual framework [127]. As a result, we developed the Norwegian Fatigue Characteristics and Interference Measure (FCIM) version 1.0, consisting of 32 items and a definition of PSF. Five different fatigue levels were considered potentially clinically relevant and gave rationale for using a Likert-scale scoring with five response options [36].

Cognitive interviews: The content validity of FCIM was then tested in cognitive interviews. The audio-recorded interviews lasted between 24 to 92 minutes and followed a semi-structured interview guide based on the Three-Step Test Interview technique, developed to assess issues with new PROMs. First, the participants completed FCIM version 1.0 while thinking aloud and were observed by the interviewer. Then, specific follow-up questions were asked based on the observation. Finally, a debriefing was conducted to gather their experiences and opinions on the instrument and the corresponding PSF definition [134]. The first author conducted and transcribed all the interviews. In line with COSMINs and FDAs guidelines, we used content analysis to assess the content validity of the PROM [36, 54]. Transcripts were transferred to NVivo 12 [130], and data were analyzed using deductive content analysis, where data was categorized based on a theoretical framework [131]. Using Tourangeau's four-stage cognitive model (comprehension, retrieval, judgment, and response), each item was analyzed separately in a categorization matrix [135]. We also assessed item relevance and the overall comprehensiveness of the PROM [36]. After the cognitive interviews, some items were deleted, revised, and *flagged*. Flagging indicated issues with content validity, but the items were kept for further investigation in the Rasch analysis. The changes led to the creation of FCIM version 2.0.

Cross-sectional survey: A web-based survey was performed to test the measurement properties of FCIM further. Written consent was obtained electronically upon questionnaire entry, dependent on participants using their electronic bank-id. Participants then responded to FCIM version 2.0 in an online questionnaire suitable for use on the participant's phone, tablet, or computer. Responses were then transferred to Winsteps software (version

5.2.0.0) for Rasch analysis. Rasch analysis models the probability of a person's response to an item using logistic transformation to convert ordinal PROM data to interval-level data. This Rasch scale measures the underlying construct (e.g., fatigue), and both persons and items are placed on the same hierarchical scale for direct comparison [56]. The construct is measured in units called logits [136]. First, we analyzed how the rating scale functioned and followed pre-set steps and criteria. Structural validity was assessed with infit unstandardized mean square statistics (MNSQ), which indicates how well the responses to individual items fit the model assumptions. Then, unidimensionality was assessed with a Principal Components Analysis (PCA) of the residuals, aiming to detect unmodeled sources of variation in the data [136, 137]. It is essential to point out that the assessment of unidimensionality in Rasch analysis is statistical and not necessarily equivalent to conceptual multidimensionality explored in qualitative studies [36]. Local dependency was tested with Yen's Q3 statistics and assesses whether the response to an item is associated with responses to other items. Internal consistency was assessed with Kuder-Richardson Formula 20 (KR-20), an equivalent of Cronbach's Alphas [136, 138].

4.5 Study variables and PROMs in papers 2 and 3

To provide background characteristics of the samples, the stroke survivors participating in all three studies were asked to provide additional information regarding sociodemographic variables, medical history, symptoms, and fatigue. Sociodemographic and medical history questions are provided in Appendix 1. These variables are described in the following section, and Table 2 summarizes the data collected in each study.

4.5.1 Sociodemographic data and stroke history

Sociodemographic data included age, gender, marital status, living arrangements, education, pre- and post-stroke working status, and social support. *Stroke history* was measured by self-report, and respondents were asked whether MR or CT scan confirmed their diagnosis, aiming to provide data aligned with the prevailing ICD-10 classification of ischemic or hemorrhagic stroke [33]. Data was also collected on time since the stroke, acute stroke symptoms, previous history of stroke, and weekly rehabilitation services.

The level of physical disability and need for help in daily activities was measured using the *Modified Rankin Scale (mRS),* with scores ranging from 0 (no symptoms) to 6 (death). Self-reliant individuals have mRS scores of 0-2 [139, 140]. During the individual interviews (paper 2), participants were asked questions about their acute stroke symptoms to determine a retrospective acute phase mRS score.

Stroke-related speech disorders were assessed based on participants' self-reported medical history. Based on the self-report, the interviewer (paper 2) determined if aphasia was present in the acute phase. In the cross-sectional survey (paper 3), participants were asked to report any stroke-related speech disorder.

4.5.2 Fatigue

Clinically significant fatigue was assessed using the *case definition for PSF* developed by Lynch et al. [35]. In a structured case interview, participants must meet all four criteria to have clinically significant PSF: 1) report feelings that are consistent with experiences of fatigue (rather than boredom or lack of motivation), 2) fatigue needs to be present nearly every day for the past two weeks, 3) fatigue needs to be present more than 50% of waking hours, and 4) fatigue must be experienced as a problem affecting daily activities. The case definition has good face validity and convergent validity [35].

Fatigue interference was measured using the *Fatigue Severity Scale* (FSS-7). It measures fatigue interference in the past week through 7 items rated on a seven-point Likert scale (from strongly disagree to strongly agree) [87]. Scores are summarized, with a mean score ranging from 1 (minimum) to 7 (maximum). The FSS-7 is a short form with better measurement properties than the original FSS-9 and has been translated into Norwegian [97, 118].

In paper 3, participants also responded to our newly-developed PSF PROM: *Fatigue Characteristics and Interference Measure* (FCIM).

4.5.3 Depression, cognitive functioning, and quality of life

Symptoms of anxiety and depression were measured using the *Hospital Anxiety and Depression Scale* (HADS). It is a self-reported screening instrument consisting of two subscales, HADS-A for anxiety and HADS-D for depression, each with seven items answered on a four-point Likert scale. Total sum scores range from 0 to 42 points [141].

Cognitive function was assessed using the *Montreal Cognitive Assessment* (MoCA) tool, a screening instrument developed to discriminate between dementia, mild cognitive impairment, and healthy individuals [142]. The Montreal Cognitive Assessment screens several cognitive domains, including executive function, which is often impaired after a stroke. This screening tool has high sensitivity, specificity, and good predictive value for post-stroke cognitive impairment [143]. The first author had prior training in conducting the Montreal Cognitive Assessment and did the assessments.

Health-related quality of life was assessed using the *EuroQol-Visual Analog Scale* (EQ-VAS), a subscale of the EuroQol five dimension scale (EQ-5D). EQ-VAS scores ranges from 0 (worst possible) to 100 (best possible) health [144].

Table 2. List of variables included in the different studies.

	Paper 2		Paper 3	
Studies	**Interviews**	**Focus groups**	**Cognitive interviews**	**Cross-sectional survey**
Sociodemographic variables	x	x	x	x
Stroke history	x	n/a	x	x
Modified Rankin Scale	x	n/a		
Stroke-related speech disorders	x	n/a		x
Case definition for fatigue	x	n/a		
Fatigue Severity Scale	x	n/a	x	x
Fatigue Characteristics and Interference Measure		n/a	x	x
Hospital Anxiety and Depression Scale	x	n/a		
Montreal Cognitive Assessment	x	n/a		
EQ-VAS	x	n/a		

n/a: not applicable

4.6 User involvement

Two user consultants with fatigue and a history of stroke have participated throughout all the steps in this study. They were involved in proposal writing, development of the interview guides, and the pilot interviews, discussed findings in paper 2 and served on the expert panel in paper 3.

4.7 Ethics

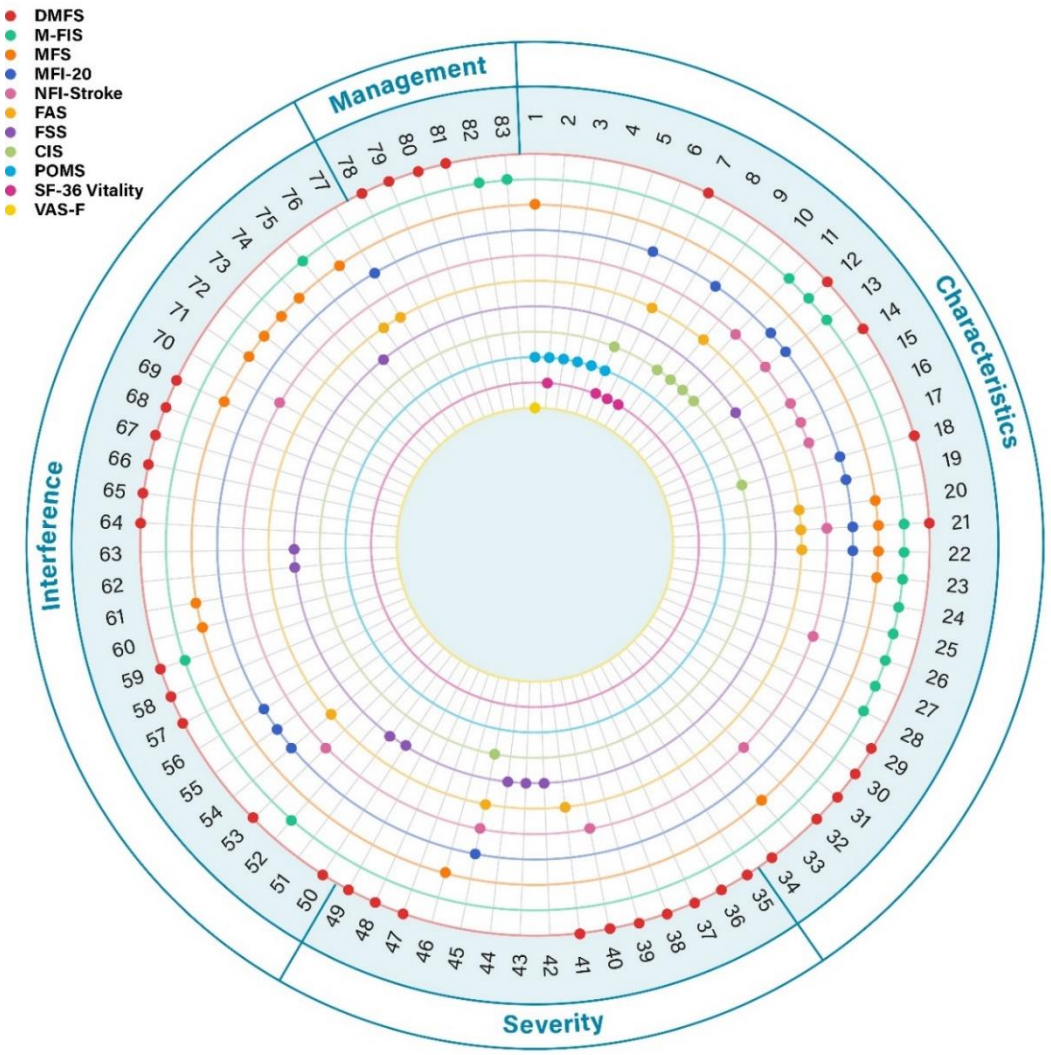
This project was conducted following the Helsinki declaration and approved by the Regional Committee for Medical Ethics, South-Eastern Norway (Reference 2016/1741). Informed written consent was obtained from all participants. Study data is stored on secure servers, at either Lovisenberg Diaconal Hospital (J: Forskning) or the University of Oslo, Services for Sensitive Data (TSD).

5.0 Main findings

5.1 Review of fatigue PROMs used in PSF research (paper 1)

A content analysis of the 156 items in the eleven PROMs in this study revealed 83 *unique* items, reflecting four dimensions of fatigue; *characteristics*, *severity*, *interference,* and *management*. These PROMs lacked items relevant to a stroke population, contained items irrelevant to a stroke population, were not comprehensive, and had substantial variation and differences in item content.

Only two PROMs measured the aspect of diurnal variation, with the Mental Fatigue Scale (MFS) and Dutch Multifactorial Fatigue Scale (DMFS) each assessing this aspect with a single item. None of the PROMs included items concerning pre-stroke fatigue, making it difficult to determine whether fatigue symptoms are pre-existing or related to the stroke. Several items were worded in a way that they did not directly relate to fatigue (e.g., “I have been forgetful” in the Modified Fatigue Impact Scale [145]) and did not include the word fatigue in the item to specify its content. Figure 2 shows that 75% of the 83 items measure fatigue interference or characteristics. Three instruments measure only fatigue quality: the Short Form-36 Vitality (SF-36 Vitality), the Profile of Mood States-Fatigue (POMS), and the Visual Analog Scale-Fatigue (VAS-F). Content overlap between PROMs is very low, and the three most commonly-used instruments, the FSS, VAS-F, and SF-36 Vitality, share none of the same items.



Characteristics

Quality

General Items

1. Fatigued
2. Worn out
3. Sluggish
4. Weary
5. Sleepy
6. Tired
7. Energy
8. Full of pep
9. Rested

Physical items

10. Exhausted
11. Feel weak
12. Physical condition
13. Physical capacity
14. Exercise causes fatigue
15. Body strength
16. Heavy limbs
17. Body unable to keep up
18. Feel fit
19. Feel active

Mental items

20. Mentally exhausted
21. Concentration
22. Thinking
23. Memory
24. Less alert
25. Organizing thoughts
26. Coordination
27. Attention
28. Making decisions
29. Conversations
30. Tired by emotions
31. Noise-induced fatigue
32. Thinking-induced fatigue

Diurnal variations

33. 24-hour variations
34. Fatigued in afternoon

Severity

Fatigue impact

35. Affects whole life
36. Serious complaint
37. Complaints get worse
38. Suffer terribly
39. Severe fatigue
40. Tired every day
41. Bothered by fatigue
42. Causes problems
43. Most disabling symptom

Onset speed

44. Easily fatigued
45. Easily tired

Recovery time

46. Long mental recovery
47. Recover easily
48. Fatigued next day
49. Wake up rested

Interference

Behavioral

50. Duties
51. Work, family, social life
52. Limited ability
53. Unable to go further
54. Think I do little
55. Get little done
56. Few plans
57. Say things I regret
58. Other people notice
59. Need to rest
60. Decreased sleep
61. Increased sleep

Physical

62. Interferes with physical function
63. Prevents sustained function
64. Induces headache
65. Body aches

Mental

66. Letting thoughts go
67. Unable to think
68. Makes mistakes
69. Emotional tendency
70. Speech problems
71. Stress sensitivity
72. Light sensitivity
73. Noise sensitivity
74. Irritability
75. Low motivation
76. Problems starting things
77. Feel no desire

Management

Coping

78. Let myself get tired out
79. Avoid overtiredness
80. Plan to rest
81. Finish tasks
82. Limit physical activity
83. Pace physical activity

Note. *Each colored circle represents a PROM and each dot represents one item. We have analyzed all the items and categorized them in dimensions (characteristics, severity, interference and management) and sub-dimensions.*

Figure 2. The eleven PROMs contain 83 different items used to measure fatigue in stroke survivors. The items are categorized into four dimensions. This figure is reprinted from Skogestad et al. 2019 [146] with permission from Elsevier.

5.2 Exploring PSF (paper 2)

The analysis of individual interviews (n=9) and focus groups (n=16) transcripts revealed four themes illustrating the experience and descriptions of PSF: characteristics of PSF, interfering and aggravating factors, management, and awareness of PSF (Figure 3). The characteristics of PSF were further divided into five categories: quality, intensity, diurnal variations, long restitution time, and invisibility. Participants reported that PSF was distinct from their pre-stroke experience of tiredness and fatigue and was mainly described as mental, physical, and general fatigue. PSF interfered with various aspects of daily life, such as emotions, cognitive performance, activities in daily life (including physical activity), and social participation. Furthermore, these activities could also aggravate or worsen fatigue symptoms. To manage PSF, individuals employed a range of emotional and behavioral strategies.

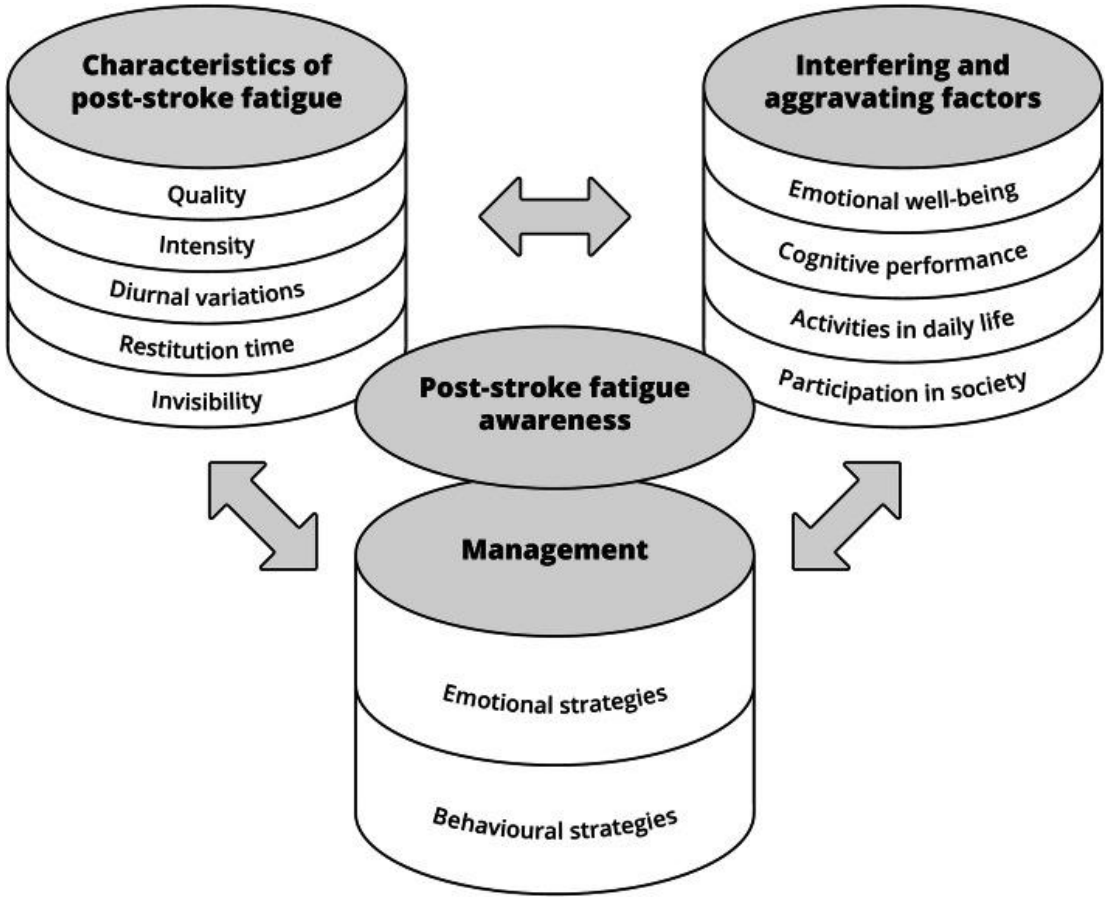


Figure 3. Conceptual framework of post-stroke fatigue displaying the four themes and corresponding categories, reprinted from Skogestad et al. 2021 [147] under the Creative Commons CC BY license terms.

The health professionals viewed acute phase fatigue after a stroke as a normal phenomenon and did not define it as PSF. The stroke survivors were not informed about PSF during hospitalization, possibly leading to a delay in their understanding of the condition. This late realization became known as PSF awareness. We also defined PSF as: "an experience of a mental, physical, or general feeling of exhaustion and tiredness, with a discrepancy between the level of activity and the level of fatigue. In individuals with a previous (pre-stroke) history of fatigue, PSF should only be considered when the feeling of fatigue is substantially different in its characteristics or severely increased in intensity" [147].

5.3 Development and testing of FCIM (paper 3)

5.3.1 Expert panel

The expert panel (n=7) created a list of 160 potential items that covered all dimensions and sub-dimensions in our conceptual framework. Conceptually, "aggravating" and "interference" were distinct sub-dimensions, and we developed several items to measure each one. However, we found that the aggravating and interfering items were too similar and unable to capture the conceptual difference between aggravating and interfering factors. As a result, we decided only to include the interference items in our PROM. The expert panel reduced the list of items and developed the initial version of the Norwegian Fatigue Characteristics and Interference Measure (FCIM). FCIM version 1.0 consisted of our PSF definition (from paper 2), a 10-item characteristics subscale, and a 20-item interference subscale with a seven-day recall period. In addition, the FCIM included two pre-stroke fatigue items. The characteristics subscale was developed for use at any time post-stroke, whereas the interference subscale is not intended for use in the acute phase. The pre-stroke fatigue items are only intended for the acute phase when recall is likely to be most accurate.

5.3.2 Cognitive interviews

The content validity of FCIM was tested in cognitive interviews (n=15). After the first ten interviews, we found comprehension problems with the PSF definition. As a result, we

simplified the wording to: “it is normal to feel tired during periods when you have a lot to do, but being fatigued means that you are more tired than you would expect considering what you have done.” In the subsequent five interviews, we did not detect any problems comprehending the revised definition.

In the characteristics subscale, all items were comprehensible and relevant, indicating high content validity. However, three items were flagged because they were similar to other items. We detected comprehension and judgment problems with nine items in the interference subscale. These items were revised before conducting the last five interviews. The main change was moving the word “fatigue” from the end to the start of the item. In addition, the pre-stroke fatigue items were moved from being the first two items of the FCIM to being at the end of the instrument. After all 15 interviews were completed, two interference items were removed because of lack of relevance, and two items in the interference subscale were flagged due to problems with comprehension and judgment. Overall, the instrument was described as providing a comprehensive assessment of PSF with no redundant items. The cognitive interviews resulted in FCIM version 2.0 with an updated PSF definition, ten characteristic items, 18 interference items, and two pre-stroke fatigue items.

5.3.3 Rasch analysis

Responses from the cross-sectional survey (n=169) were subject to a Rasch analysis. Two items in the characteristics subscale demonstrated poor fit to the Rasch model and were removed. Two additional items were removed due to local dependency (these items were previously flagged in the cognitive interviews). The Wright map shows that the subscale captures all fatigue levels but might work best for low and medium levels of fatigue (Figure 4). Investigation of the interference subscale showed misfit in items 15 and 16; these items were also locally dependent (Yens Q3 at 0.47) and removed. We removed two items due to local dependency and two more because they had previously been flagged in the cognitive interviews. The Wright map shows that the subscale works well across all fatigue levels (Figure 5). The final 6-item characteristics subscale and the 12-item interference subscale

were both unidimensional, and KR-20 was 0.86 and 0.94, respectively. Pearson correlation between the 6-item characteristic subscale and 12-item interference subscale was r=0.77.

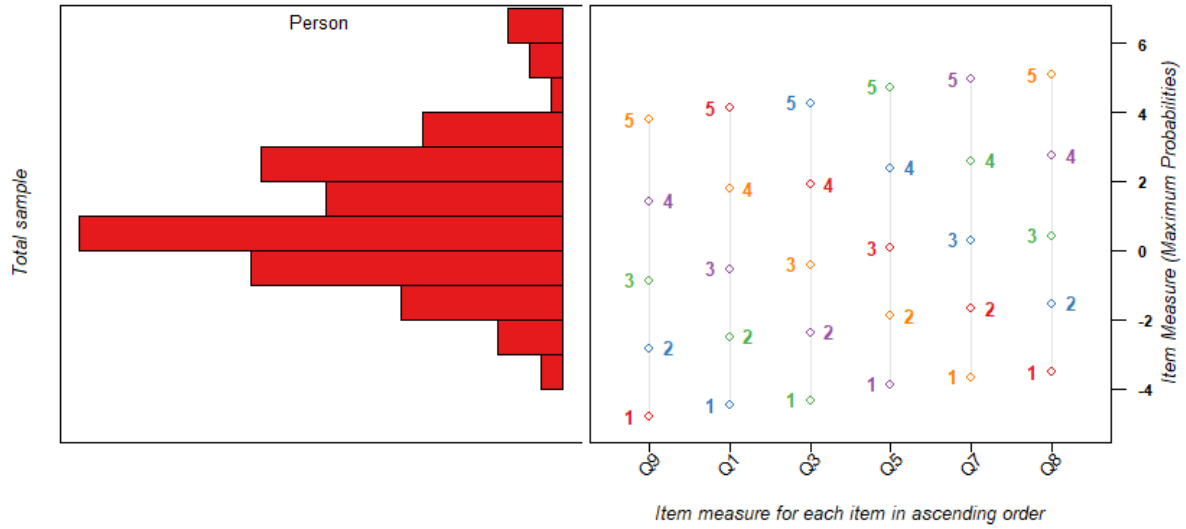


Figure 4. Wright map displaying the 6-item characteristics subscale. The left side shows the participants' ability measures (based on their mean logits), presented by the total sample. The right side shows the items' difficulty calibrations by the maximum probabilities of observing each category. The logits of the underlying Rasch model are presented on the right side, and both the participants' abilities (i.e., fatigue levels) and the item difficulties are spaced along this axis.

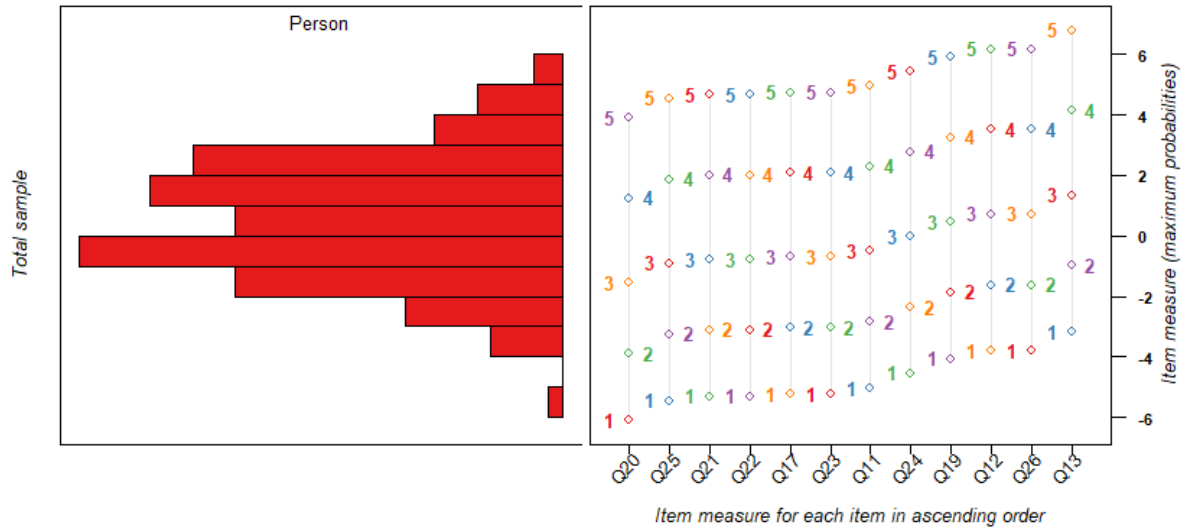


Figure 5. Wright map displaying the 12-item interference subscale. The left side shows the participants' ability measures (based on their mean logits), presented by the total sample. The right side shows the items' difficulty calibrations by the maximum probabilities of observing each category. The logits of the underlying Rasch model are presented on the right side, and both the participants' abilities (i.e., fatigue levels) and the item difficulties are spaced along this axis.

6.0 Discussion

6.1 Methodological considerations

In this thesis, we used an exploratory sequential mixed-methods design, intending to develop a PROM for PSF. PROM development guidelines clearly state the need for qualitative and quantitative methods, as using both methods is necessary to establish evidence for all measurement properties [36]. Using a mixed-methods design has the advantage that qualitative and quantitative findings can complement each other and capture essential details and different aspects of the phenomena under study [128]. A challenge is using the findings and integrating the qualitative data to build the quantitative measure and report this process in a trustworthy way [39, 128]. Trustworthiness is a crucial criterion for research quality. It involves the concepts of credibility, dependability, confirmability, and generalizability of findings [39]. The following section will discuss the study's trustworthiness in light of the main methodological considerations and how this may affect the collection and analysis of data presented in this thesis.

6.1.1 PROM selection

The COSMIN guidelines for reviewing PROMs emphasize the inclusion of all available instruments, not just the most widely-used ones [148]. Newly-developed PROMs, which may adhere to quality standards, may not yet have gained widespread use despite their potential benefits. In our review (paper 1), we included the most-used PROMs as well as infrequently-used instruments that were partly developed or tested in a stroke population. The Dutch Multifactor Fatigue Scale [116], Neurological Fatigue Index-Stroke [149], and Profile of Mood States-Fatigue [103, 150] were only used in one or two prior studies but were included because they potentially had high quality. The extensive systematic literature search followed PRISMA guidelines [129] and identified studies relevant to our review. We excluded 13 PROMs that were not tested in stroke survivors and were used in fewer than four studies. Including these 13 instruments could have yielded interesting and possibly differing results. However, the sample of 11 PROMs was considered adequate for providing a concise yet comprehensive review of fatigue instruments used in a stroke population.

6.1.2. Samples

Several decisions must be made regarding the sample when conducting a study. We drew the samples from the stroke population in Norway, and the intent was to make some inferences about the population based on the study results. Therefore, evaluating the sampling strategy and sample sizes is crucial to assess the study's credibility [39].

COSMIN guidelines recommend maximum variation sampling for qualitative studies to obtain rich, differing, and potentially contrasting perspectives [113, 127]. In both interview studies (papers 2 and 3), we purposively sampled stroke survivors and included participants with diverse demographic and clinical variables, including diverse experiences with PSF. Despite the samples' diversity, some common patterns emerged in the data that supported the credibility and generalizability of the findings. PROM development guidelines also address the importance of obtaining "buy-in" from all stakeholders (i.e., patients, clinicians, and researchers) [27]. The advantage of including data from health professionals is that it can provide a broader range of descriptions beyond individual experiences [39]. Thus, we also conducted focus groups with health professionals (paper 2) and included researchers, clinicians, and stroke survivors in our expert panel (paper 3). Collecting data from different types of people is called person triangulation and aims to validate the data through multiple perspectives [39]. Including different stakeholders can also affect the sample size needed to achieve data saturation. Overall, including different perspectives adds to the credibility and dependability of findings [39].

In qualitative studies, sample size is influenced by factors such as data quality and richness, interview length, and research scope [39]. According to COSMIN guidelines, gaining sample diversity and data saturation is of higher significance than the number of participants, and a specific sample size is not required for conceptual framework development [127]. In paper 2, the participants had high fatigue levels and provided in-depth reflections about their experiences. The interviews also covered all the topics in the interview guide and participants gave detailed descriptions of their experiences. Both interview studies (papers 2 and 3) achieved data saturation through simultaneous data analysis and data collection stopped when no new categories emerged. In cognitive interviews assessing the content

validity of a PROM, a sample size of at least 7 participants is rated as very good [127]. Based on these overall considerations, the sample sizes were considered sufficient for the two interview studies [39].

In the cross-sectional survey (paper 3), we used a convenience sampling method, which requires a few considerations. Nonprobability samples are less likely to represent the population as compared to probability sampling [39]. Inviting participants via the internet could indicate that participants had some specific characteristics. For example, participants needed to sign the consent form using an “electronic bank-id,” and some contacted the research team asking for help to complete this. As a result, people with low digital competence were unlikely to be recruited. Digital competence is also inversely correlated with age, which could at least partially explain why our digitally-competent sample was also younger (median 52.4 years) than the stroke population in Norway (74.2 years) [17]. As a result, the findings might be less generalizable to the elderly stroke population, and future studies should aim to validate the FCIM in a more representative sample. Next, data from the cross-sectional survey were subject to a Rasch analysis. A Rasch analysis converts the ordinal data from the Likert scale responses to interval data (logits) presented on a hierarchical scale [136]. Both persons and items are presented on the same scale. Rasch analysis is a probability model, and the characteristics of the items are, in theory, independent of the people responding to them [56]. These methodological underpinnings make Rasch analysis less vulnerable to sampling biases than other statistical analyses [56].

The sample size for a Rasch analysis would ideally exceed 200 participants. However, a sample of 169 respondents is considered *adequate* (100-199) for performing Rasch analysis in general [148], and this sample size also surpasses the required minimum of 100 to analyze internal consistency [148].

6.1.3 Data collection

Data collection in qualitative studies involves several important considerations. Reflexivity is crucial and involves researchers reflecting critically on their competence and role in knowledge development [39]. It is impossible to avoid personal "bias" altogether in

qualitative research. As a result, researchers should provide detailed information regarding personal factors and characteristics that might influence the data collection and analysis [133]. A novice qualitative researcher is particularly vulnerable to unintentionally influencing participants or not detecting and following up on vital perspectives during the interviews [151]. To counter this, we conducted pilot interviews, and the interview transcripts were thoroughly reviewed by experienced qualitative researchers focusing on interview skills to enhance the credibility and confirmability of findings.

In all three studies involving stroke survivors (papers 2 and 3), we collected data on sociodemographic variables, medical history, and symptoms assessed with different PROMs. Describing sample characteristics is considered necessary in light of interpreting the findings and the analytic generalizability of the study [39]. Despite substantial criticism of its measurement properties, we assessed fatigue with the Fatigue Severity Scale (FSS). In our qualitative studies, the FSS was only reported as a descriptive characteristic of the included sample. In contrast, in the cross-sectional survey, the FSS was used to analyze differential item functioning (DIF) in the responses to the FCIM. Not only is the FSS one of the few fatigue measures available in Norwegian [97], it also has established scoring cut-offs for different levels of fatigue [152], and dividing fatigue levels into groups was needed to analyze DIFs [136]. This use of FSS, despite its deficiencies, presents the current paradox in PSF research. The limitations of FSS highlight the need for new and improved PROMs, but validating new instruments requires relying on old, low-quality PROMs.

6.1.4 Data analysis

When conducting reviews of existing PROMs, COSMIN guidelines advise that content validity be rated based on prior studies reporting the PROMs' development and assessment of their measurement properties. In addition to the evidence from the literature, the reviewers can establish an expert panel with patients to perform ratings of the PROMs' content. Thus, COSMIN's content validity assessment for existing PROMs has two different analytical approaches: review of existing literature AND ratings by an expert panel [127]. In our review (paper 1), we did not evaluate existing studies on content validity. Content validity in these PROMs is generally not assessed in a stroke population [104, 105], and there is scarce

evidence of content validity studies to include in a review. In addition, a review of the few existing studies has been conducted elsewhere [12, 105]. COSMIN guidelines suggests that the expert panel (ideally including patients) assesses each PROM individually [36]. In our review, we aimed to analyze the similarities and differences in item content across the 156 items in the 11 included PROMs. Performing content analysis on the items can be considered a more structured approach than usual content validity assessment [36] and required extensive time and methodological competence. Thus including stroke survivors in this analysis was not considered feasible.

A content analysis was also applied to the individual interviews and focus group transcripts (paper 2) [131]. However, qualitative data can be analyzed following numerous approaches [39]. Content analysis aims to give broad and condensed descriptions of the phenomenon under study, and inductive content analysis is especially useful for developing a conceptual framework [131]. Thus, this approach was well suited for analyzing the individual interviews and focus groups. Content analysis can also be used in deductive analyses [131], which was the rationale for applying content analysis to the cognitive interview transcripts (paper 3) where the text segments were categorized into pre-defined categories. Qualitative analysis has no universal rules or standard procedures, making it challenging to report the analytical process [39]. To enhance the trustworthiness of our findings, both interview studies (papers 2 and 3) followed the widely-endorsed consolidated criteria for reporting qualitative research (COREQ checklist) [133].

In the development of PROMs, different statistical analyses can be conducted to assess structural validity and internal consistency [36]. Two widely-used approaches are classical test theory (CTT) and item response theory (IRT), with Rasch analysis being one type of analysis within IRT [39]. IRT is viewed as a more sophisticated and complex analysis than CTT [56]. CTT assumes that each individual has a true score and that the observed score is a reflection of this true score combined with random error. The analysis then focuses on diminishing this error [153]. Within CTT, factor analysis is used to evaluate the statistical dimensionality of a PROM. Factor analysis identifies clusters of items that can be grouped and measure one underlying dimension of the construct [39]. In contrast, IRT analysis

estimates the probability of responses and items placed on a specific point on the same underlying hierarchical scale, as shown in Figures 4 and 5 (main findings, 5.3.3). The Rasch model converts the ordinal data to interval-level data, and the scale for the FCIM ranges from -4 to 6. The aim is to include items with different "difficulties" spread out over the scale capturing all levels of fatigue from -4 to 6. Having items with various positions on the scale enables a detailed estimation of a person's difficulty level [136]. Rasch analysis evaluates the fit of the data to the model one item at a time. Other IRT models evaluate the global fit of all the items as one group and then accept or reject this model [56]. In our study, it was most suitable to apply a Rasch analysis since the aim was item reduction and selection of the items with the best fit to the model. To enhance the trustworthiness of findings, we followed the Rasch reporting guideline for rehabilitation research [154].

6.1.5 Item selection with mixed-methods

In the PROM development process, we first established an initial list of 160 items which, step by step, were reduced to a 20-item PROM. Supported by qualitative and quantitative findings, we aimed to select items from the initial list with the highest methodological quality. A critical discussion in mixed-methods research is whether one strand of findings (i.e., qualitative or quantitative) should be prioritized over the other and how findings from the two strands can be integrated [39].

In exploratory sequential mixed-methods designs, qualitative data is the starting point and is commonly prioritized over quantitative data. The rationale is that little information about the topic is known in advance [155]. This priority is also acknowledged in PROM development guidelines such as COSMIN and FDA, where extensive documentation on patient input based on qualitative data is vital to provide a comprehensive understanding of the phenomena under study [54]. In this thesis, the qualitative findings were emphasized. The development of the initial list of items and the reduction of this list to a PROM containing 30 items (FCIM version 2.0) were based on qualitative findings. First, during the analysis of the cross-sectional survey data, we used both qualitative and quantitative analysis to support the decisions of which items to retain or remove from the final version of the FCIM. The qualitative and quantitative findings were *integrated* during this process,

meaning that findings from both strands were used to support the analytical decisions. This integration is often viewed as an essential part of mixed-methods research. A challenge can occur if the findings are contradictory. Thus *how* the qualitative and quantitative findings are used and integrated is important to discuss [39].

In the cross-sectional survey, 169 stroke survivors completed the FCIM. This data was first investigated quantitatively, where we removed items based on findings from the Rasch analysis. In cases where the quantitative findings alone did not provide conclusive results, we based our decisions on qualitative findings from the cognitive interviews. In detail, items 1 and 2, 3 and 4, and 5 and 6, represented three item pairs containing the same content with slightly different wording. Thus, we never intended to keep both items in these three item pairs. In agreement with our theoretical understanding, the Rasch analysis showed that these item pairs were locally dependent. We then used results from the cognitive interviews to guide the decision about which items should be removed. In the cognitive interviews, items 2, 4, and 6 had been flagged. Thus we decided to delete these items and keep items 1, 3, and 5. We also removed the two flagged items in the interference subscale (27 and 28). Removing these items decreased the unexplained variance in the residual contrasts to below the pre-set criteria for unidimensionality [136]. In sum, removing these five items increased the overall measurement properties of the FCIM. This analytical process makes explicit that inferences based on both qualitative and quantitative findings add to our study's credibility, confirmability, and generalizability and is an example of the advantages of a mixed-methods design [39, 128].

6.1.6 Analytic generalizability

The overall aim of this thesis was to develop a PROM that can be used to measure fatigue in stroke survivors. Following PROM development guidelines, PSF was explored with qualitative studies, and findings were used to inform later stages of item generation [36]. Inherent in this design lay the assumption that it is possible to generalize qualitative findings and draw broad inferences based on a few observations [127]. However, generalizability in qualitative research is complicated and controversial [39]. One strand argues that qualitative research aims not to generalize but to explore multiple realities and provide

detailed and in-depth descriptions of contextualized experiences. The other side argues that qualitative research is well suited for extrapolation because it can develop higher-level theories and reveal concepts relevant to most people [156]. We aimed to use the qualitative findings to develop a conceptual framework of PSF, called *analytic generalization* [156]. To assess the quality of our analytic generalization, several aspects of the analysis and reporting of findings conducted in this thesis are relevant to address [39].

The analysis identified four overarching dimensions with corresponding categories that were interpreted as not unique to the individual and thus suitable for use as a conceptual framework for PROM development. Participants' actual experiences within each category could be different and contrasting, but all participants described, for example, diurnal variation in their fatigue and that it interfered with activities in daily life. All participants also described different management strategies. However, the specific strategies they used were very different. Several aspects support the trustworthiness of this analysis and its analytic generalizability: two researchers independently coded and analyzed the data, one of whom was highly experienced, and including both stroke survivors and health professionals' ensured validity by person triangulation. Data collection was conducted simultaneously with data analysis, and data collection continued until no new categories emerged in the additional data. Our findings from the explorative study were also in line with previous qualitative studies of PSF, which adds to the credibility and trustworthiness of our study. In addition, we followed reporting guidelines to ensure sufficient information in the paper to enable the reader to judge the analytic generalizability of the findings.

Other strengths of this study were that we conducted the interviews and the cross-sectional survey in different samples and that the findings from all three studies contributed to the development and testing of the FCIM. This triangulation of methods also adds to the trustworthiness of our analytic generalizability [39]. The FCIM was developed and tested in a Norwegian-speaking sample. For the FCIM to be translated into other languages and used in more culturally diverse samples, it needs cross-cultural validation following proper guidelines [36]. Thus, the assessment of FCIMs measurement properties should be considered a continuous process.

6.1.7 Ethical considerations

Care must be taken to protect human rights when they are the source of knowledge development. The Belmont Report divides ethical considerations in research into three essential principles: beneficence, respect for human dignity, and justice [157].

The principle of beneficence in research imposes an obligation on researchers to minimize harm and discomfort, even temporary discomfort [39]. In this study's interviews with stroke survivors, particularly attention was paid to the stress and fatigue the interview experience might cause for the participants. The interviewer discussed this with participants in advance and assured them that they could take a break or end the interview at any time. Several participants requested a break, indicating that the interview was stressful, but also that they understood their right to take a break. The interviewer also discussed in advance that participating in this study would not be of specific help to the participant, but rather, would produce benefits for future patients. However, afterward, several participants expressed gratitude for and comfort with being able to talk about their situation and satisfaction that their participation might help others in a similar situation.

The second important ethical principle is respect for human dignity. It involves treating participants as autonomous people who can make voluntary decisions about participating in a study and fully informing participants about the nature of the study and study participation [157, 158]. To comply with these principles, we obtained all participants' written informed consent (electronic or on paper).

Justice, the third principle in research ethics, involves the right to privacy [157]. Privacy is vital and potentially challenging in qualitative research, where findings are reported by presenting rich, detailed, and unique segments of interview transcripts. Despite the removal of name, age, and other personal identifiers, contextual quotes can potentially identify participants. Reporting findings with quotes requires careful consideration [159]. Thus, we thoroughly considered each quote and did not include examples where detailed information could lead to deductive disclosure of the participants.

6.2 Discussion of main findings

In patient-centered healthcare, clinicians and regulators are increasingly interested in using PROMs as an important outcome for healthcare decision-making on both individual and population levels [160]. PROMs can even serve as primary outcomes in clinical drug trials assessing symptom treatment or as secondary outcomes helping to differentiate drugs with similar survival rates [54, 161]. To rely on PROMs in decision-making, it is crucial to ensure that they have strong measurement properties. The first step is to establish a clear definition of the concept, a well-described conceptual framework, and evidence of content validity. Only then can other measurement properties be evaluated [54, 112].

6.2.1 Conceptual definition of PSF

The understanding of fatigue can vary considerably among stakeholders (e.g., researchers, clinicians, and stroke survivors). For this reason, it is vital to include a clear and precise definition in the PROM in line with the guidelines of COSMIN and the FDA [36, 54]. Including a definition will ensure that all users have a consistent understanding of the symptom and can distinguish fatigue from related concepts, such as normal tiredness, and differentiate between acute and chronic fatigue and pre- and post-stroke fatigue, which is vital for the development of high-quality PROMs [36, 54].

We defined PSF as: “an experience of a mental, physical or a general feeling of exhaustion and tiredness, with a discrepancy between the level of activity and the level of fatigue” (paper 2). An important aspect of this PSF definition is the differentiation between 1) normal tiredness after activity and 2) fatigue without preceding activity, which aligns with the ICD-11 definition of fatigue [34]. During PROM development, conceptual definitions and frameworks typically evolve, supported by subsequent qualitative data collection from patients and other relevant stakeholders [36]. We evaluated our initial definition of PSF through cognitive interviews (paper 3) and found that the initial version was long and

challenging for the participants to understand. In contrast, the participants described the revised and simpler version[1] as being easily understandable.

Commonly-used fatigue PROMs, such as the FSS [87], Fatigue Assessment Scale (FAS) [88], and Checklist Individual Strength (CIS) [162], do not include a definition of fatigue in their instrument. In paper 1, we found that these PROMs include several items that do not contain the word fatigue, such as "I have enough energy for everyday life" [88] and "Do you feel sleepy" [150]. Without clearly stating what the PROM intends to measure and why, it can be difficult for users to understand that the intent is to measure fatigue and to differentiate between concepts such as fatigue, lack of energy, and sleep disturbances. Such response biases are a common issue in PROM responses [56]. The COSMIN methodology emphasizes explicitly defining the measured symptom and related concepts to mitigate this issue [36].

In the acute phase after a stroke, it is considered normal to be extremely exhausted (paper 2). The health professionals viewed tiredness during the acute phase after a stroke as natural and temporary and did not define it as PSF. Several studies distinguish between acute and chronic PSF, where PSF is defined as chronic when experienced more than six months after a stroke [105, 163]. The stroke survivors in our study (paper 2) interpreted acute phase fatigue as a normal reaction to being acutely ill. First, when fatigue persisted after the initial phase, they became aware of PSF as a bothersome symptom. This awareness of PSF is similar to previous qualitative studies, which found that most participants experienced fatigue during hospitalization. However, the extent of fatigue and the impact on role loss was mostly recognized sometime after discharge [43, 47]. To date, there is not enough evidence to conclude whether acute phase fatigue (in the first 7 days after stroke) should be defined as PSF or whether PSF only should be considered and measured when it has lasted for a while and interferes with daily functions. However, it is necessary to note that the impact of fatigue on daily activities is typically recognized only after the acute

[1] Revised PSF definition included in the FCIM: "It is very normal to feel tired during periods when you have a lot to do, but being fatigued means that you are more tired than you would expect considering what you have done."

phase of a stroke. Therefore, measurement of fatigue interference should be focused on later stages of post-stroke recovery.

The FSS, the most widely used fatigue PROM in stroke samples, mainly measures interference. The FSS includes items such as “Fatigue interferes with my work, family or social life” and “Fatigue interferes with carrying out certain duties and responsibilities” [87]. Some studies, such as Wang et al. [164, 165], have used the FSS in the acute phase. The results of these studies need to be interpreted with caution since interference items might be difficult to assess during the acute phase. An advantage of the FCIM is that it separates characteristics and interference into two sub-dimensions (paper 3). Thus, we suggest that the FCIM can be used both in the acute phase and in later stages of recovery, and that the characteristics subscale specifically can be used at all stages, whereas the interference subscale might only be relevant after the acute phase.

Differentiating PSF from pre-stroke fatigue is also a pertinent aim. Fatigue occurs in a wide variety of diseases [32]. After a stroke, people may report fatigue that was also present preceding their stroke, and it is unclear whether PSF and fatigue in other conditions share similar underlying mechanisms [32]. In line with a qualitative meta-synthesis where stroke survivors described PSF as a different type than fatigue occurring pre-stroke [98], we found that PSF was described as a different feeling compared to prior experiences. Moreover, we suggested that a PSF diagnosis should only be considered if the feeling of fatigue after a stroke was substantially different or increased in intensity compared to before the stroke (paper 2). Different causal mechanisms can potentially respond to different treatment initiatives [38], and research might benefit from differentiating PSF from pre-stroke fatigue. Based on prior knowledge, the expert panel decided to include two pre-stroke fatigue items in the FCIM (paper 3). We recommend that these pre-stroke items only be measured in the acute phase after a stroke, considering that a more extended recall period increases the probability of recall bias [56].

6.2.2 Conceptual framework of PSF

There is a consensus that PSF is a multidimensional concept [13, 28]. However, there is no universal agreement on what constitutes the main dimensions of PSF or which dimensions should be considered the most important. As a result, current fatigue PROMs measure different fatigue dimensions and have substantial differences in item content, as reported in paper 1.

In particular, there was little overlap in item content across the three most widely used fatigue PROMs: the FSS, VAS-F, and SF-36 Vitality. Convergent validity studies (i.e., the degree to which different instruments measuring the same concept produce similar results [39]) generally show weak to moderate correlations between fatigue PROMs in a stroke sample. If two instruments measure similar dimensions of a concept, a correlation coefficient of 0.7 or higher would be expected [36]. However, correlations between FSS and SF-36 Vitality [166], FSS and Lynch interview [61], FSS and Neurological Fatigue Index-stroke subscales [115], and SF-36 Vitality and Fatigue Assessment Scale have been reported to be between r= 0.41 – 0.62 [103]. In comparison, we found that the two subscales of the FCIM were highly correlated, with r= 0.77 (paper 3).

These moderate correlations in existing fatigue PROMs and the lack of content overlap found in paper 1 suggests, at least, that the different PROMs measure different dimensions of fatigue but also indicates that they might be measuring different concepts altogether. A recent study explored PSF using three different fatigue instruments and found considerable variability in identifying fatigue cases among stroke patients. This variability was attributed to the three PROMs addressing different aspects of PSF [61]. Existing PROMs split fatigue into dimensions in multiple different ways. For example, the Multidimensional Fatigue Inventory-20 separates fatigue into mental fatigue, general fatigue, physical fatigue, reduced activity, and reduced motivation [89]. In contrast, the FCIM separates the items into characteristics and interference subscales. From a measurement perspective, it is crucial to clearly articulate the conceptual framework that informs the dimensions and related items rather than solely focusing on how items are grouped into different dimensions [36].

Several previous qualitative studies support our findings of PSF as a multidimensional phenomenon. We found that PSF consisted of four themes; characteristics, aggravating and interfering factors, management, and PSF awareness (paper 2). In agreement with our findings, a qualitative meta-synthesis reported characteristics of PSF as a central theme, where the participants described a lack of energy to perform activities and becoming quickly tired, unpredictable feelings of fatigue, and that it was a complicated invisibility [55]. Previous studies have also described mental, physical, and general feelings of fatigue, as well as diurnal variations [41, 53]. Other pervasive findings are the impacts of PSF on cognitive functions, emotions, activities, and social life [40, 43, 44, 50, 52, 167], and a recent study also described "trigger activities" in line with our findings of factors that could aggravate the feeling of PSF [46].

The expert panel (paper 3) developed items based on the conceptual framework developed in paper 2. COSMIN guidelines suggest incorporating aspects and dimensions in the conceptual stage rather than finding out post hoc via factor analysis or other statistical tests [36]. Thus, the expert panel divided the instrument into two dimensions: characteristics and interference (paper 3). Assessment of interference has been seen as an important dimension in other patient groups [168, 169], providing rationale for retaining the interference items in our PROM. In hemodialysis patients, symptoms that were rated high in intensity were not always accompanied by high interference ratings [168], illustrating the importance of distinguishing between characteristics/intensity and interference.

We further identified that fatigue management involved a wide range of strategies, consistent with previous research that addresses the need for different, flexible, and individualized approaches to coping [40-43, 45, 46, 48, 55]. FDA points out the importance of ensuring that all items are relevant to most of the patients in a study [54], and considering that fatigue management is highly individual, the expert panel concluded that it was challenging to develop an instrument including a wide range of different coping strategies (paper 3). In addition, the aim was to develop a PROM suited for clinical research, with the potential to measure improvement. Thus, management of fatigue was not

considered the most relevant factor. This was supported by the content validity study, where participants in the cognitive interviews (paper 3) judged the FCIM to be comprehensive and inclusive of the most relevant items. In addition, a study investigating five different subscales in a fatigue PROM found that "coping with fatigue" had very low correlations with the other subscales (r= 0.08-0.15), indicating that coping might be a distinct and different concept [116].

6.2.3 Content validity of fatigue PROMs

New and existing PROMs need to have high content validity by including items that reflect the construct they intend to measure. Evaluation of content validity includes assessing item relevance, comprehensibility, and comprehensiveness of the PROM [36]. The FDA addresses the importance of only including relevant items, as irrelevant items might result in a failure to detect treatment effects (even if the treatment was effective) or problems with missing responses [54]. In paper 1, we found that the 11 included fatigue PROMs lacked relevant items, contained irrelevant or unspecific items, and needed to be more comprehensive. Building on this knowledge, in addition to qualitative studies, we developed the FCIM, aiming to meet higher standards for a PROM. Overall, FCIM version 2.0 had high content validity (paper 3). To illustrate the importance of content validity, a discussion yielding examples of similarities and differences between items in the FCIM and PROMs in the review (paper 1) will follow in the next paragraphs.

The fatigue PROMs included in our review lacked crucial items, such as the assessment of diurnal variations (paper 1). Changes in fatigue intensity during the day are essential to assess for several reasons. First, in line with other qualitative studies [41, 53], we found that PSF fluctuates throughout the day (paper 2). These diurnal variations have recently been confirmed by a quantitative study comparing real-time fatigue assessments for six consecutive days with retrospective FSS scores [170]. Second, this quantitative study also found that individuals with similar total FSS scores demonstrated very different fatigue patterns from day to day [170]. In a sample of HIV/AIDS patients, morning and evening fatigue were found to occur in distinct daily patterns associated with different symptoms of anxiety and depression [171]. In the FCIM, we included several items measuring diurnal

variations that all had high content validity in the cognitive interviews (paper 3). The Wright map (Figure 4) showed that fatigue in the morning (item 7) and noon (item 8) were the most difficult items in the characteristics subscale, meaning that only stroke survivors with the highest levels of fatigue were likely to select the highest response category, “almost always” (category 5), on these items. Interestingly, fatigue in the afternoon (item 9) was the “easiest” item, meaning that lower overall fatigue levels were needed to select the highest response category on this item. Our findings indicate that afternoon fatigue might be more prevalent than morning and noon fatigue, particularly in respondents with lower overall fatigue levels. Cognitive interviews confirmed the content validity of these items, and the ability to measure diurnal variations provided a detailed and comprehensive understanding of stroke survivors fatigue levels.

Including irrelevant or unspecific items for measuring PSF could also negatively affect the validity of results. Physical impairments are common in stroke survivors [20], and seven of the 11 included PROMs contained items that coincide with this prevalent stroke sequela (paper 1). These items can lead to measurement error in a stroke sample, and content validity studies aim to minimize such biases [127]. For example, the Dutch Multifactorial Fatigue Scale, developed to assess fatigue in patients after acquired brain injuries, includes items such as: “I have a good physical condition” and “I feel physically fit” [116]. Using physical condition as an indicator of fatigue in stroke survivors with physical impairments can lead to inaccurate results, as the link between physical condition and fatigue may not be clear. This example highlights the importance of conducting content validity studies to ensure that the respondents understand the intended meaning of an item [172]. However, it should be noted that fatigue seems to have a physical component. We found that PSF was described as a physical sensation in the body and that fatigue often interfered with plans of physical activity (paper 2), which is why we included items in the FCIM that specifically address this aspect of fatigue. The precise wording of items is crucial, and even minor variations in wording can impact the interpretation of an item and subsequent responses [173]. Aiming to ensure a clear linkage to fatigue, we included the following items in the FCIM: “did you feel physically fatigued” and “did you avoid physical activity because of

fatigue". Both items were found to be relevant and comprehensible from a content validity perspective (paper 3).

Other problematic items we found in the review were related to mental characteristics in six of the 11 PROMs (paper 1). Most of these items asked only about one specific cognitive function, such as Neurological Fatigue Index-stroke, which included: "sometimes I really have to concentrate on what are usually simple things" [115]. Considering mild cognitive impairment and dementia are common after a stroke, such items are also ambiguous. However, we found that fatigue can interfere with concentration, and deep concentration can aggravate fatigue (paper 2). Thus, concentration and cognitive performance seem related to the fatigue experience. In the FCIM, we included the item: "how often were you so fatigued that you had problems concentrating". In the first version of the FCIM, this item had the word "fatigue" at the last end of the question. Moving the word "fatigue" from the end to the start of the question seemed to more clearly link the item to fatigue and not only to concentration difficulties. This finding implicates the importance of carefully wording the items and assessing comprehension and relevance of the items in a content validity study [36, 54]. Only after establishing content validity can other measurement properties, such as structural validity and internal consistency, be assessed [36, 54].

6.2.4 Structural validity

Assessing the structural validity of a PROM using Rasch analysis involves evaluating the fit of the item responses to the underlying theoretical model of the assessed construct. This assessment helps ensure that the items in the PROM function as intended and measure a single concept, known as unidimensionality. Evaluating the statistical unidimensionality of the subscales also provides information regarding the quality of individual items. The knowledge regarding the functioning of individual items can guide item reduction in a PROM development process [36]. In paper 3, we tested the structural validity of each of the FCIM's two subscales and removed several items that did not fit with the Rasch model. We verified the statistical unidimensionality of the two subscales (characteristics and interference), which is a prerequisite for interpreting internal consistency analysis [148].

The quality of individual items in a Rasch analysis can be evaluated with infit MNSQ values. A high infit MNSQ value means that the individual scores deviate from the Rasch model's expectation, which indicates randomness in the data and that the item is not functioning as intended. Randomness is a serious threat to validity, and it is advised to delete items with high MNSQ values [136, 137]. For example, item 15 ("fatigue interferes with taking a bath or shower") and item 16 ("fatigue interferes with getting dressed/undressed") in the interference subscale were removed due to a high MNSQ value exceeding our pre-set criterion of <1.3 [174]. High MNSQ values suggests that people's responses are different than predicted, which can imply that these items do not add to the same dimension [136]. After the removal of misfitting, flagged, and locally-dependent items, the two dimensions were found to be unidimensional and displayed good evidence of structural validity.

Despite the importance of testing structural validity, only a few studies have investigated this in fatigue PROMs used in stroke research [115, 116, 119]. A study investigating the FSS found that two items in the FSS-9 had unacceptably high MNSQ values. The remaining seven items had better measurement properties than the original FSS-9, and therefore, it is recommended to use the 7-item version of the FSS. [118]. Despite this critical finding, several PSF studies conducted after 2011 use the 9-item FSS-version [164, 175, 176]. The Neurological Fatigue Index, initially developed for multiple sclerosis, had four subscales (physical, cognitive, relief by sleep, and abnormal sleep). However, the two sleep scales showed misfits and were omitted after evaluations in a stroke sample [115]. The Fatigue Assessment Scale has 10 items on a scale meant to be unidimensional. However, a factor analysis conducted in a stroke sample found that four of the items had low factor loadings on the Fatigue Assessment Scale and instead loaded high on a corresponding depression scale, indicating that these items were measuring depression instead of fatigue [119]. Structural validity assessment evaluates whether the scores are a good reflection of the concept it intends to measure. It is fundamental to establish this underlying structure before assessing internal consistency (i.e., the level of correlation between items) [36].

6.2.5 Internal consistency

High internal consistency means that the answers to individual items in a PROM correlate highly with every other item. Internal consistency of a PROM is commonly reported with Cronbach's alpha. It is essential to note that Cronbach's alpha does not test whether the individual items are influenced by one or several concepts, i.e., it is not a measure of the dimensionality of a PROM. Instruments consisting of items representing several concepts can obtain a high Cronbach's alpha. Thus, statistical unidimensionality is a prerequisite for correctly interpreting internal consistency analysis [36, 110]. In the analysis of the FCIM, we found that KR-20 (Cronbach's alpha equivalent) was 0.86 for the characteristics subscale and 0.94 for the interference subscale. Since unidimensionality is a criterion for providing a meaningful Cronbach's alpha [36], only Cronbach's alphas for each scale were presented. A Cronbach's alpha between 0.7-0.8 is considered acceptable, and 0.8-0.9 is good [56]. Very high values, generally considered at 0.95 or higher, might indicate redundant items due to the high correlations between items. However, Cronbach's alpha depends on the number of items in a scale, and very high values are often found in PROMs with many items [110]. The Cronbach's alpha for the characteristic subscale is good. However, the Cronbach's alpha for the interference subscale is very close to the highest acceptable level, indicating possibly redundant items. On the other hand, findings from the cognitive interviews did not indicate item redundancy, which provides a good rationale for keeping all items in the final 12-item interference subscale. In sum, both FCIM subscales were considered to have good internal consistency.

6.2.6 Additional measurement properties

It is necessary to conduct several studies on the FCIM's measurement properties before using it in a clinical study. Ideally, this thesis would include such subsequent studies. However, one of the challenges of a mixed-methods design is its resource implications. It requires considerable time to plan, conduct and analyze several sequential data sets [155]. We have followed COSMIN's guidelines which advise evaluating measurement properties in a specific order. First is content validity establishment, followed by an assessment of structural validity and internal consistency [148]. The next step is to document further evidence of construct validity, responsiveness, and test-retest reliability [36].

7.0 Conclusions

In this thesis, we found that existing fatigue PROMs used in stroke populations have significant shortcomings and do not comply with the best practice guidelines and high standards set by regulators. As a result, we developed a new PROM specifically targeted for measuring PSF. The FCIM was developed based on a clear definition and a conceptual framework of PSF. Further, we found that the FCIM has good content validity, structural validity, and internal consistency.

7.1 Implications for clinical practice

The development of FCIM and findings from this thesis has the potential to impact clinical practice and improve patient outcomes. The FCIM can provide a more accurate and comprehensive assessment of PSF compared to existing measures. The use of PROMs in patient care can help to empower patients, assist clinical decision-making and track patient progress [177]. By allowing patients to provide direct feedback, FCIM can help health professionals to better understand the patient's needs and tailor their treatment and rehabilitation plans, taking into account their specific symptoms and experiences with PSF. After proper translation into English, we would also advocate for FCIM's international recognition as a standard of best practice.

We discovered low content validity and limited overlap in items among existing fatigue PROMs. These drawbacks highlights the need for caution when selecting a fatigue instrument. It is crucial to carefully consider the intended use of the PROM and choose one that is appropriate for the specific purpose. For example, if the goal is to monitor changes in an individual over time, it is important to choose an instrument that is responsive to change. On the other hand, if the aim is to assess management strategies, an instrument including such items must be used. Widely used PROMs, such as the FSS, may be helpful if the aim is to compare results between populations or samples. Our review offers a thorough examination of item content and dimensions in existing PROMs, and this can enable clinicians to make well-informed choices about the appropriate instrument for their requirements.

Health professionals in this study considered acute phase fatigue to be a typical response to stroke and did not classify it as PSF. However, stroke survivors reported experiencing fatigue from the onset of their stroke, but their awareness of PSF only became apparent after discharge. The pooled prevalence rate of PSF is around 50% [57, 58]. Despite this, the participants in our study were not provided with information about PSF during the acute stage. Based on the findings from our study, stroke survivors should be informed of PSF during hospitalization, along with potential management strategies and follow-up routines. Knowledge of PSF could enable stroke survivors to be better prepared to cope with PSF and potentially develop more effective management strategies from the outset.

7.2 Implications for future studies

Our review of PROMs can help interpret research about PSF. By highlighting the different dimensions and aspects of PSF being assessed by various PROMs, it provides a framework for understanding the findings of different studies. This framework can provide insight into inconsistencies in the literature and give a clearer picture of the current state of knowledge of PSF. Additionally, by providing insight into the strengths and limitations of different PROMs, this review can inform future research and help ensure that studies are designed to lead to more accurate and meaningful results.

The next step is to evaluate the FCIM's construct validity, responsiveness, and test-retest reliability. In addition, we need to validate the instrument cross-culturally in an English-speaking sample. Our findings also open several new avenues for future research. Using FCIM in longitudinal studies can examine changes in fatigue characteristics (intensity) and interference over time and their relationship to recovery and rehabilitation outcomes. Another interesting area to explore is the impact of pre-stroke fatigue on PSF and recovery. Moreover, identifying subtypes of PSF, such as physical versus mental fatigue and different diurnal patterns, and examining the factors contributing to these subtypes can provide essential insights. In addition, if the FCIM proves to be responsive to change, it could be used in targeted intervention studies to reduce PSF and potentially support labeling claims in drug therapy developments.

References

1. Kuriakose D, Xiao Z. Pathophysiology and Treatment of Stroke: Present Status and Future Perspectives. *International Journal of Molecular Sciences.* 2020;21(20). doi: 10.3390/ijms21207609.
2. Grefkes C, Fink GR. Recovery from stroke: current concepts and future perspectives. *Neurological Research and Practice.* 2020;2:17. doi: 10.1186/s42466-020-00060-6.
3. Skaner Y, Nilsson GH, Sundquist K, Hassler E, Krakau I. Self-rated health, symptoms of depression and general symptoms at 3 and 12 months after a first-ever stroke: a municipality-based study in Sweden. *BMC Family Practice.* 2007;8:61. doi: 10.1186/1471-2296-8-61.
4. Rudberg AS, Berge E, Laska AC, Jutterstrom S, Nasman P, Sunnerhagen KS, Lundstrom E. Stroke survivors' priorities for research related to life after stroke. *Topics in Stroke Rehabilitation.* 2021;28(2):153-8. doi: 10.1080/10749357.2020.1789829.
5. Andersen G, Christensen D, Kirkevold M, Johnsen SP. Post-stroke fatigue and return to work: a 2-year follow-up. *Acta Neurologica Scandinavica.* 2012;125(4):248-53. doi: 10.1111/j.1600-0404.2011.01557.x.
6. Naess H, Lunde L, Brogger J. The effects of fatigue, pain, and depression on quality of life in ischemic stroke patients: the Bergen Stroke Study. *Vascular Health and Risk Management.* 2012;8:407-13. doi: 10.2147/VHRM.S32780.
7. MacIntosh BJ, Edwards JD, Kang M, Cogo-Moreira H, Chen JL, Mochizuki G, Herrmann N, Swardfager W. Post-stroke Fatigue and Depressive Symptoms Are Differentially Related to Mobility and Cognitive Performance. *Frontiers in aging neuroscience.* 2017;9:343. doi: 10.3389/fnagi.2017.00343.
8. Elf M, Eriksson G, Johansson S, von Koch L, Ytterberg C. Self-Reported Fatigue and Associated Factors Six Years after Stroke. *PLoS ONE [Electronic Resource].* 2016;11(8):e0161942. doi: 10.1371/journal.pone.0161942.
9. Duncan F, Wu S, Mead GE. Frequency and natural history of fatigue after stroke: a systematic review of longitudinal studies. *Journal of Psychosomatic Research.* 2012;73(1):18-27. doi: 10.1016/j.jpsychores.2012.04.001.
10. Wu S, Mead G, Macleod M, Chalder T. Model of understanding fatigue after stroke. *Stroke.* 2015;46(3):893-8. doi: 10.1161/STROKEAHA.114.006647.
11. Kutlubaev MA, Mead GE, Lerdal A. Fatigue after stroke--perspectives and future directions. *International Journal of Stroke.* 2015;10(3):280-1. doi: 10.1111/ijs.12428.
12. Tyson SF, Brown P. How to measure fatigue in neurological conditions? A systematic review of psychometric properties and clinical utility of measures used so far. *Clinical Rehabilitation.* 2014;28(8):804-16. doi: 10.1177/0269215514521043.
13. Acciarresi M, Bogousslavsky J, Paciaroni M. Post-stroke fatigue: epidemiology, clinical characteristics and treatment. *European Neurology.* 2014;72(5-6):255-61. doi: 10.1159/000363763.
14. Cumming TB, Mead G. Classifying post-stroke fatigue: Optimal cut-off on the Fatigue Assessment Scale. *Journal of Psychosomatic Research.* 2017;103:147-9. doi: 10.1016/j.jpsychores.2017.10.016.
15. Kruithof N, Van Cleef MH, Rasquin SM, Bovend'Eerdt TJ. Screening Poststroke Fatigue; Feasibility and Validation of an Instrument for the Screening of Poststroke

Fatigue throughout the Rehabilitation Process. *Journal of Stroke & Cerebrovascular Diseases.* 2016;25(1):188-96.
16. Feigin VL, Brainin M, Norrving B, Martins S, Sacco RL, Hacke W, Fisher M, Pandian J, Lindsay P. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. *Int J Stroke.* 2022;17(1):18-29. doi: 10.1177/17474930211065917.
17. Norwegian Stroke Registry. Annual report for 2021. 2022. https://stolav.no/Documents/%c3%85rsrapport%20Norsk%20hjerneslagregister%202021.pdf.
18. Feigin VL, Stark BA, Johnson CO, Roth GA, Bisignano C, Abady GG, Abbasifard M, Abbasi-Kangevari M, Abd-Allah F, Abedi V, Abualhasan A, Abu-Rmeileh NME, Abushouk AI, Adebayo OM, Agarwal G, Agasthi P, Ahinkorah BO, Ahmad S, Ahmadi S, et al. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Neurology.* 2021;20(10):795-820. doi: 10.1016/S1474-4422(21)00252-0.
19. Shakir R, Norrving B. Stroke in ICD-11: the end of a long exile. *The Lancet.* 2017;389(10087):2373. doi: 10.1016/S0140-6736(17)31567-2.
20. Chen S, Zeng L, Hu Z. Progressing haemorrhagic stroke: categories, causes, mechanisms and managements. *Journal of Neurology.* 2014;261(11):2061-78. doi: 10.1007/s00415-014-7291-1.
21. Campbell BCV, Khatri P. Stroke. *The Lancet.* 2020;396(10244):129-42. doi: 10.1016/S0140-6736(20)31179-X.
22. Bernhardt J, Hayward KS, Kwakkel G, Ward NS, Wolf SL, Borschmann K, Krakauer JW, Boyd LA, Carmichael ST, Corbett D, Cramer SC. Agreed definitions and a shared vision for new standards in stroke recovery research: The Stroke Recovery and Rehabilitation Roundtable taskforce. *International Journal of Stroke.* 2017;12(5):444-50. doi: 10.1177/1747493017711816.
23. World Health Organization (WHO). How to use the ICF: A practical manual for using the International Classification of Functioning, Disability and Health (ICF). 2013.
24. Engel GL. The need for a new medical model: a challenge for biomedicine. *Science.* 1977;196(4286):129-36.
25. Geyh S, Cieza A, Schouten J, Dickson H, Frommelt P, Omar Z, Kostanjsek N, Ring H, Stucki G. ICF Core Sets for stroke. *Journal of Rehabilitation Medicine.* 2004(44 Suppl):135-41. doi: 10.1080/16501960410016776.
26. Lanctot KL, Lindsay MP, Smith EE, Sahlas DJ, Foley N, Gubitz G, Austin M, Ball K, Bhogal S, Blake T, Herrmann N, Hogan D, Khan A, Longman S, King A, Leonard C, Shoniker T, Taylor T, Teed M, et al. Canadian Stroke Best Practice Recommendations: Mood, Cognition and Fatigue following Stroke, 6th edition update 2019. *International Journal of Stroke.* 2020;15(6):668-88. doi: 10.1177/1747493019847334.
27. Lerdal A, Gay CL. Fatigue in the acute phase after first stroke predicts poorer physical health 18 months later. *Neurology.* 2013;81(18):1581-7. doi: 10.1212/WNL.0b013e3182a9f471.
28. Hinkle JL, Becker KJ, Kim JS, Choi-Kwon S, Saban KL, McNair N, Mead GE, American Heart Association Council on Cardiovascular and Stroke Nursing and Stroke Council. Poststroke Fatigue: Emerging Evidence and Approaches to Management: A Scientific Statement for Healthcare Professionals From the American Heart Association. *Stroke.* 2017;48(7):e159-e70. doi: 10.1161/STR.0000000000000132.

29. Cumming TB, Yeo AB, Marquez J, Churilov L, Annoni JM, Badaru U, Ghotbi N, Harbison J, Kwakkel G, Lerdal A, Mills R, Naess H, Nyland H, Schmid A, Tang WK, Tseng B, van de Port I, Mead G, English C. Investigating post-stroke fatigue: An individual participant data meta-analysis. *Journal of Psychosomatic Research.* 2018;113:107-12. doi: 10.1016/j.jpsychores.2018.08.006.
30. Ingles JL, Eskes GA, Phillips SJ. Fatigue after stroke. *Archives of Physical Medicine & Rehabilitation.* 1999;80(2):173-8.
31. De Doncker W, Dantzer R, Ormstad H, Kuppuswamy A. Mechanisms of poststroke fatigue. *Journal of Neurology, Neurosurgery & Psychiatry.* 2018;89(3):287-93. doi: 10.1136/jnnp-2017-316007.
32. Torres-Harding S, Jason LA. What is fatigue? History and epidemiology. In: Deluca J, editor. Fatigue as a window to the brain. Cambridge: MIT; 2005.
33. World Health Organization (WHO). International Classification of Diseases, Tenth Revision (ICD-10). 1992.
34. World Health Organization (WHO). International Classification of Diseases, Eleventh Revision (ICD-11). 2019/2021. https://icd.who.int/browse11.
35. Lynch J, Mead G, Greig C, Young A, Lewis S, Sharpe M. Fatigue after stroke: the development and evaluation of a case definition. *Journal of Psychosomatic Research.* 2007;63(5):539-44.
36. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide. Practical Guides to Biostatistics and Epidemiology. Cambridge: Cambridge University Press; 2011.
37. Choi-Kwon S, Mitchell PH, Kim JS. Nursing Interventions for Poststroke Fatigue. *Stroke.* 2015;46(10):e224-7. doi: 10.1161/STROKEAHA.115.009534.
38. Lerdal A, Bakken LN, Kouwenhoven SE, Pedersen G, Kirkevold M, Finset A, Kim HS. Poststroke fatigue--a review. *Journal of Pain & Symptom Management.* 2009;38(6):928-49. doi: 10.1016/j.jpainsymman.2009.04.028.
39. Polit DF, Beck CT. Nursing Research: Generating and Assessing Evidence for Nursing Practice. Ninth editio ed. Wolters Kluwer Health/Lippincott Williams & Wilkins; 2012.
40. Flinn NA, Stube JE. Post-stroke fatigue: qualitative study of three focus groups. *Occupational Therapy International.* 2010;17(2):81-91. doi: 10.1002/oti.286.
41. Barbour VL, Mead GE. Fatigue after Stroke: The Patient's Perspective. *Stroke Research and Treatment.* 2012;2012:863031. doi: 10.1155/2012/863031.
42. Kirkevold M, Christensen D, Andersen G, Johansen SP, Harder I. Fatigue after stroke: manifestations and strategies. *Disability & Rehabilitation.* 2012;34(8):665-70. doi: 10.3109/09638288.2011.615373.
43. White JH, Gray KR, Magin P, Attia J, Sturm J, Carter G, Pollack M. Exploring the experience of post-stroke fatigue in community dwelling stroke survivors: a prospective qualitative study. *Disability & Rehabilitation.* 2012;34(16):1376-84. doi: 10.3109/09638288.2011.645111.
44. Young CA, Mills RJ, Gibbons C, Thornton EW. Poststroke fatigue: the patient perspective. *Topics in Stroke Rehabilitation.* 2013;20(6):478-84.doi: 0.1310/tsr2006-478.
45. Thomas K, Gamlin C, De Simoni A, Mullis R, Mant J. How is poststroke fatigue understood by stroke survivors and carers? A thematic analysis of an online discussion forum. *BMJ Open.* 2019;9(7):e028958. doi: 10.1136/bmjopen-2019-028958.

46. Ablewhite J, Nouri F, Whisker A, Thomas S, Jones F, das Nair R, Condon L, Jones A, Sprigg N, Drummond A. How do stroke survivors and their caregivers manage post-stroke fatigue? A qualitative study. *Clinical Rehabilitation.* 2022: doi: 10.1177/02692155221107738.
47. Bicknell ED, Said CM, Haines KJ, Kuys S. "I Give It Everything for an Hour Then I Sleep for Four." The Experience of Post-stroke Fatigue During Outpatient Rehabilitation Including the Perspectives of Carers: A Qualitative Study. *Frontiers in neurology.* 2022;13:900198. doi: 10.3389/fneur.2022.900198.
48. Eriksson G, Larsson I, Guidetti S, Johansson U. Handling fatigue in everyday activities at five years after stroke: A long and demanding process. *Scandinavian Journal of Occupational Therapy.* 2022:1-11. doi: 10.1080/11038128.2022.2089230.
49. Carlsson GE, Moller A, Blomstrand C. Consequences of mild stroke in persons <75 years -- a 1-year follow-up. *Cerebrovascular Diseases.* 2003;16(4):383-8. doi: 10.1159/000072561.
50. Roding J, Lindstrom B, Malm J, Ohman A. Frustrated and invisible--younger stroke patients' experiences of the rehabilitation process. *Disability & Rehabilitation.* 2003;25(15):867-74. doi: 10.1080/0963828031000122276.
51. Robison J, Wiles R, Ellis-Hill C, McPherson K, Hyndman D, Ashburn A. Resuming previously valued activities post-stroke: who or what helps? *Disability & Rehabilitation.* 2009;31(19):1555-66. doi: 10.1080/09638280802639327.
52. Thompson HS, Ryan A. The impact of stroke consequences on spousal relationships from the perspective of the person with stroke. *Journal of Clinical Nursing.* 2009;18(12):1803-11. doi: 10.1111/j.1365-2702.2008.02694.x.
53. Liljehult J, Molsted S, Christensen T, Moller T, Overgaard D. Mastering Health Following Minor Stroke-A Qualitative Explorative Study. *Journal of Stroke & Cerebrovascular Diseases.* 2022;31(8):106607. doi: 10.1016/j.jstrokecerebrovasdis.2022.106607.
54. U.S. Department of Health and Human Services Food and Drug Administration (FDA). Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. https://www.fda.gov/media/77832/download.
55. Eilertsen G, Ormstad H, Kirkevold M. Experiences of poststroke fatigue: qualitative meta-synthesis.*Journal of Advanced Nursing.*2013;69(3)514-25. doi:10.1111/jan.12002
56. Polit DF, Yang FM. Measurement and the Measurement of Change: A Primer for the Health Professions. 2014.
57. Alghamdi I, Ariti C, Williams A, Wood E, Hewitt J. Prevalence of fatigue after stroke: A systematic review and meta-analysis. *European Stroke Journal.* 2021;6(4):319-32. doi: 10.1177/23969873211047681.
58. Cumming TB, Packer M, Kramer SF, English C. The prevalence of fatigue after stroke: A systematic review and meta-analysis. *International Journal of Stroke.* 2016;11(9):968-77.
59. Christensen D, Johnsen SP, Watt T, Harder I, Kirkevold M, Andersen G. Dimensions of post-stroke fatigue: a two-year follow-up study. *Cerebrovascular Diseases.* 2008;26(2):134-41. doi: 10.1159/000139660.

60. Duncan F, Greig C, Lewis S, Dennis M, MacLullich A, Sharpe M, Mead G. Clinically significant fatigue after stroke: a longitudinal cohort study. *Journal of Psychosomatic Research.* 2014;77(5):368-73. doi: 10.1016/j.jpsychores.2014.06.013.
61. Kjeverud A, Andersson S, Lerdal A, Schanke AK, Ostlie K. A cross-sectional study exploring overlap in post-stroke fatigue caseness using three fatigue instruments: Fatigue Severity Scale, Fatigue Questionnaire and the Lynch's Clinical Interview. *Journal of Psychosomatic Research.* 2021;150:110605. doi: 10.1016/j.jpsychores.2021.110605.
62. Schepers VP, Visser-Meily AM, Ketelaar M, Lindeman E. Poststroke fatigue: course and its relation to personal and stroke-related factors. *Archives of Physical Medicine & Rehabilitation.* 2006;87(2):184-8.
63. Wu S, Kutlubaev MA, Chun HY, Cowey E, Pollock A, Macleod MR, Dennis M, Keane E, Sharpe M, Mead GE. Interventions for post-stroke fatigue. *Cochrane Database of Systematic Reviews.* 2015(7):CD007030. doi: 10.1002/14651858.CD007030.pub3.
64. Su Y, Yuki M, Otsuki M. Non-Pharmacological Interventions for Post-Stroke Fatigue: Systematic Review and Network Meta-Analysis. *Journal of Clinical Medicine.* 2020;9(3):25. doi: 10.3390/jcm9030621.
65. Smetheram T, Amiama ME, Hebert D, Law G, Dawson DR. Occupational Therapy Interventions for Poststroke Fatigue: A Scoping Review. *OTJR: Occupation, Participation, & Health.* 2022;42(3):182-8. doi: 10.1177/15394492221087968.
66. Karaiskos D, Tzavellas E, Spengos K, Vassilopoulou S, Paparrigopoulos T. Duloxetine versus citalopram and sertraline in the treatment of poststroke depression, anxiety, and fatigue. *The Journal of neuropsychiatry and clinical neurosciences.* 2012;24(3):349-53.
67. Choi-Kwon S, Choi J, Kwon SU, Kang DW, Kim JS. Fluoxetine is not effective in the treatment of post-stroke fatigue: a double-blind, placebo-controlled study. *Cerebrovascular Diseases.* 2007;23(2-3):103-8.
68. Bivard A, Lillicrap T, Krishnamurthy V, Holliday E, Attia J, Pagram H, Nilsson M, Parsons M, Levi CR. MIDAS (Modafinil in Debilitating Fatigue After Stroke): A Randomized, Double-Blind, Placebo-Controlled, Cross-Over Trial. *Stroke.* 2017;48(5):1293-8. doi: 10.1161/STROKEAHA.116.016293.
69. Paciaroni M, Acciarresi M. Poststroke Fatigue. *Stroke.* 2019;50(7):1927-33. doi: 10.1161/STROKEAHA.119.023552.
70. Nguyen S, Wong D, McKay A, Rajaratnam SMW, Spitz G, Williams G, Mansfield D, Ponsford JL. Cognitive behavioural therapy for post-stroke fatigue and sleep disturbance: a pilot randomised controlled trial with blind assessment. *Neuropsychological Rehabilitation.* 2019;29(5):723-38. doi: 10.1080/09602011.2017.1326945.
71. Johansson B, Bjuhr H, Ronnback L. Mindfulness-based stress reduction (MBSR) improves long-term mental fatigue after stroke or traumatic brain injury. *Brain Injury.* 2012;26(13-14):1621-8. doi: 10.3109/02699052.2012.700082.
72. Clarke A, Barker-Collo SL, Feigin VL. Poststroke fatigue: does group education make a difference? A randomized pilot trial. *Topics in Stroke Rehabilitation.* 2012;19(1):32-9. doi: 10.1310/tsr1901-32.
73. Zedlitz AM, Rietveld TC, Geurts AC, Fasotti L. Cognitive and graded activity training can alleviate persistent fatigue after stroke: a randomized, controlled trial. *Stroke.* 2012;43(4):1046-51. doi: 10.1161/STROKEAHA.111.632117.

74. Kuppuswamy A, Clark EV, Turner IF, Rothwell JC, Ward NS. Post-stroke fatigue: a deficit in corticomotor excitability? *Brain.* 2015;138(Pt 1):136-48. doi: 10.1093/brain/awu306.
75. Penner I-K, Paul F. Fatigue as a symptom or comorbidity of neurological diseases. *Nature Reviews Neurology.* 2017;13(11):662-75. doi: 10.1038/nrneurol.2017.117.
76. Kim JS. Post-stroke Mood and Emotional Disturbances: Pharmacological Therapy Based on Mechanisms. *Journal of Stroke.* 2016;18(3):244-55.
77. Zhang S, Cheng S, Zhang Z, Wang C, Wang A, Zhu W. Related risk factors associated with post-stroke fatigue: a systematic review and meta-analysis. *Neurological Sciences.* 2021;42(4):1463-71. doi: 10.1007/s10072-020-04633-w.
78. Aarnes R, Stubberud J, Lerdal A. A literature review of factors associated with fatigue after stroke and a proposal for a framework for clinical utility. *Neuropsychological Rehabilitation.* 2020;30(8):1449-76. doi: 10.1080/09602011.2019.1589530.
79. Ayerbe L, Ayis S, Wolfe CD, Rudd AG. Natural history, predictors and outcomes of depression after stroke: systematic review and meta-analysis. *British Journal of Psychiatry.* 2013;202(1):14-21. doi: 10.1192/bjp.bp.111.107664.
80. Taylor-Rowan M, Momoh O, Ayerbe L, Evans JJ, Stott DJ, Quinn TJ. Prevalence of pre-stroke depression and its association with post-stroke depression: a systematic review and meta-analysis. *Psychological Medicine.* 2019;49(4):685-96. doi: 10.1017/S0033291718002003.
81. Moussavi SMPH, Chatterji SD, Verdes EP, Tandon AP, Patel VP, Ustun BMD. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet.* 2007;370(9590):851-8. doi:10.1016/S0140-6736(07)61415-9.
82. Choi-Kwon S, Han SW, Kwon SU, Kim JS. Poststroke fatigue: characteristics and related factors. *Cerebrovascular Diseases.* 2005;19(2):84-90.
83. van der Werf SP, van den Broek HL, Anten HW, Bleijenberg G. Experience of severe fatigue long after stroke and its relation to depressive symptoms and disease characteristics. *European Neurology.* 2001;45(1):28-33. doi: 10.1159/000052085.
84. Winward C, Sackley C, Metha Z, Rothwell PM. A population-based study of the prevalence of fatigue after transient ischemic attack and minor stroke. *Stroke.* 2009;40(3):757-61. doi: 10.1161/strokeaha.108.527101.
85. Lerdal A, Bakken LN, Rasmussen EF, Beiermann C, Ryen S, Pynten S, Drefvelin AS, Dahl AM, Rognstad G, Finset A, Lee KA, Kim HS. Physical impairment, depressive symptoms and pre-stroke fatigue are related to fatigue in the acute phase after stroke. *Disability & Rehabilitation.* 2011;33(4):334-42. doi: 10.3109/09638288.2010.490867.
86. Fried EI. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders.* 2017;208:191-7.
87. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives Neurology.* 1989;46(10):1121-3. doi: 10.1001/archneur.1989.00520460115022.
88. Michielsen HJ, De Vries J, Van Heck GL. Psychometric qualities of a brief self-rated fatigue measure: The Fatigue Assessment Scale. *Journal of Psychosomatic Research.* 2003;54(4):345-52. doi: 10.1016/s0022-3999(02)00392-6.

89. Smets EMA, Garssen B, Bonke B, De Haes JCJM. The multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research.* 1995;39(3):315-25. doi: 10.1016/0022-3999(94)00125-O.
90. Pihlaja R, Uimonen J, Mustanoja S, Tatlisumak T, Poutiainen E. Post-stroke fatigue is associated with impaired processing speed and memory functions in first-ever stroke patients. *Journal of Psychosomatic Research.* 2014;77(5):380-4. doi: 10.1016/j.jpsychores.2014.08.011.
91. Panagiotakis SH, Simos P, Zaganas I, Basta M, Perysinaki GS, Fountoulakis N, Giaka M, Kapetanaki S, Koutentaki I, Bertsias A, Duijker G, Scarmeas N, Tziraki C, Lionis C, Vgontzas A, Boumpas DT. Self-reported fatigue as a risk index for dementia diagnosis. *European geriatric medicine.* 2018;9(2):211-7. doi: 10.1007/s41999-017-0020-4.
92. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Kokmen E, Tangelos EG. Aging, memory, and mild cognitive impairment. *International Psychogeriatrics.* 1997;9 Suppl 1:65-9. doi: 10.1017/s1041610297004717.
93. Pendlebury ST, Rothwell PM. Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis. *Lancet Neurology.* 2009;8(11):1006-18. doi: 10.1016/S1474-4422(09)70236-4.
94. Ballard C, Rowan E, Stephens S, Kalaria R, Kenny RA. Prospective Follow-Up Study Between 3 and 15 Months After Stroke. *Stroke.* 2003;34(10):2440-4. doi: doi:10.1161/01.STR.0000089923.29724.CE.
95. Kales HC, Gitlin LN, Lyketsos CG. Assessment and management of behavioral and psychological symptoms of dementia. *BMJ.* 2015;350.
96. Langhorne PF, Coupar FB, Pollock AP. Motor recovery after stroke: a systematic review. *Lancet Neurology.* 2009;8(8):741-54. doi: 10.1016/S1474-4422(09)70150-4.
97. Lerdal A, Wahl A, Rustoen T, Hanestad BR, Moum T. Fatigue in the general population: a translation and test of the psychometric properties of the Norwegian version of the fatigue severity scale. *Scandinavian Journal of Public Health.* 2005;33(2):123-30. doi: 10.1080/14034940410028406.
98. Whitehead LC, Unahi K, Burrell B, Crowe MT. The Experience of Fatigue Across Long-Term Conditions: A Qualitative Meta-Synthesis. *Journal of Pain & Symptom Management.* 2016;52(1):131-43.e1. doi: 10.1016/j.jpainsymman.2016.02.013.
99. Borren NZ, van der Woude CJ, Ananthakrishnan AN. Fatigue in IBD: epidemiology, pathophysiology and management. *Nature Reviews Gastroenterology & Hepatology.* 2019;16(4):247-59.
100. O'Higgins C, Brady B, O'Connor B, Walsh D, Reilly R. The pathophysiology of cancer-related fatigue: current controversies. *Supportive Care in Cancer.*2018;26(10):3353-64.
101. Yong SJ. Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. *Infectious diseases.* 2021;53(10):737-54.
102. Johnston BC, Patrick D, Devji T, Maxwell L, Bingham IC, Beaton D, Boers M, Briel M, Busse J, Carrasco-Labra A, Christensen R, da Costa B, el Dib R, Lyddiatt A, Ostelo R, Shea B, Singh J, Terwee C, Williamson P, et al. Chapter 18: Patient-reported outcomes. In: Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 63 Cochrane. 2022. www.training.cochrane.org/handbook.

103. Mead G, Lynch J, Greig C, Young A, Lewis S, Sharpe M. Evaluation of fatigue scales in stroke patients. *Stroke.* 2007;38(7):2090-5.
104. Elbers RG, Rietberg MB, van Wegen EE, Verhoef J, Kramer SF, Terwee CB, Kwakkel G. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Quality of Life Research.* 2012;21(6):925-44. doi: 10.1007/s11136-011-0009-2.
105. Nadarajah M, Goh HT. Post-stroke fatigue: a review on prevalence, correlates, measurement, and management. *Topics in Stroke Rehabilitation.* 2015;22(3):208-20. doi: 10.1179/1074935714Z.0000000015.
106. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology.* 2010;63(7):737-45. doi: 10.1016/j.jclinepi.2010.02.006.
107. National Council on Measurement in Education, American Psychological Association. Standards for educational and psychological testing. Washington, D.C: American Educational Research Association; 2014.
108. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, Stein RE. Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research.* 2002;11(3):193-205. doi: 10.1023/a:1015291021312.
109. Chan EKH. Standards and Guidelines for Validation Practices: Development and Evaluation of Measurement Instruments. In: Zumbo BD, Chan EKH, editors. Validity and Validation in Social, Behavioral, and Health Sciences. Cham: Springer International Publishing; 2014. p. 9-24.
110. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology.* 2007;60(1):34-42. doi: 10.1016/j.jclinepi.2006.03.012.
111. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research.* 2010;19(4):539-49. doi: 10.1007/s11136-010-9606-8.
112. EMA. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. The use of patient-reported outcome (PRO) measures in oncology studies. London: United Kingdom2016.
113. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Quality of Life Research.* 2009;18(9):1263-78. doi: 10.1007/s11136-009-9540-9.
114. Ware JE, Jr., Sherbourne CD. The MOS 36-ltem Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Medical Care.* 1992;30(6).
115. Mills RJ, Pallant JF, Koufali M, Sharma A, Day S, Tennant A, Young CA. Validation of the Neurological Fatigue Index for stroke (NFI-Stroke). *Health & Quality of Life Outcomes.* 2012;10:51. doi: 10.1186/1477-7525-10-51.
116. Visser-Keizer AC, Hogenkamp A, Westerhof-Evers HJ, Egberink IJ, Spikman JM. Dutch multifactor fatigue scale: a new scale to measure the different aspects of fatigue

after acquired brain injury. *Archives of Physical Medicine & Rehabilitation.* 2015;96(6):1056-63. doi: 10.1016/j.apmr.2014.12.010.
117. Mills RJ, Young CA, Pallant JF, Tennant A. Development of a patient reported outcome scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS).*Health and Quality of Life Outcomes.* 2010;8(1):22.doi:10.1186/1477-7525-8-22.
118. Lerdal A, Kottorp A. Psychometric properties of the Fatigue Severity Scale-Rasch analyses of individual responses in a Norwegian stroke cohort. *International Journal of Nursing Studies.* 2011;48(10):1258-65. doi: 10.1016/j.ijnurstu.2011.02.019.
119. Smith OR, van den Broek KC, Renkens M, Denollet J. Comparison of fatigue levels in patients with stroke and patients with end-stage heart failure: application of the fatigue assessment scale. *Journal of the American Geriatrics Society.* 2008;56(10):1915-9. doi: 10.1111/j.1532-5415.2008.01925.x.
120. Brandal A, Eriksson M, Wester P, Lundin-Olsson L. Reliability and validity of the Swedish Fatigue Assessment Scale when self-administrered by persons with mild to moderate stroke. *Topics in Stroke Rehabilitation.* 2016;23(2):90-7. doi: 10.1080/10749357.2015.1112057.
121. Butt Z, Lai JS, Rao D, Heinemann AW, Bill A, Cella D. Measurement of fatigue in cancer, stroke, and HIV using the Functional Assessment of Chronic Illness Therapy - Fatigue (FACIT-F) scale. *Journal of Psychosomatic Research.* 2013;74(1):64-8. doi: 10.1016/j.jpsychores.2012.10.011.
122. Chuang LL, Lin KC, Hsu AL, Wu CY, Chang KC, Li YC, Chen YL. Reliability and validity of a vertical numerical rating scale supplemented with a faces rating scale in measuring fatigue after stroke. *Health & Quality of Life Outcomes.* 2015;13:91. doi: 10.1186/s12955-015-0290-9.
123. Tseng BY, Gajewski BJ, Kluding PM. Reliability, responsiveness, and validity of the visual analog fatigue scale to measure exertion fatigue in people with chronic stroke: a preliminary study. *Stroke Research and Treatment.* 2010. doi: 10.4061/2010/412964.
124. Valko PO, Bassetti CL, Bloch KE, Held U, Baumann CR. Validation of the fatigue severity scale in a Swiss cohort. *Sleep.* 2008;31(11):1601-7. doi: 10.1093/sleep/31.11.1601.
125. Hubacher M, Calabrese P, Bassetti C, Carota A, Stocklin M, Penner IK. Assessment of post-stroke fatigue: the fatigue scale for motor and cognitive functions. *European Neurology.* 2012;67(6):377-84. doi: 10.1159/000336736.
126. McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Medicine.* 2011;9:86. doi:10.1186/1741-7015-9-86.
127. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Bouter LM, de Vet HCW, Mokkink LB. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research.* 2018;27(5):1159-70. doi: 10.1007/s11136-018-1829-0.
128. Creswell JW, Plano Clark VL. Designing and conducting mixed methods research. 3rd edition. International student edition. ed. Los Angeles: Sage; 2018.
129. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology.* 2009;62(10):1006-12. doi: 10.1016/j.jclinepi.2009.06.005.

130. QSR International Pty Ltd (2020). NVivo 12 (released in March 2020) https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home
131. Elo S, Kyngas H. The qualitative content analysis process. *Journal of Advanced Nursing.* 2008;62(1):107-15. doi: 10.1111/j.1365-2648.2007.04569.x.
132. Levandowsky M, Winter D. Distance between Sets. *Nature.* 1971;234(5323):34-5. doi: 10.1038/234034a0.
133. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care.* 2007;19(6):349-57. doi: 10.1093/intqhc/mzm042.
134. Hak T, Veer Kvd, Jansen H. The Three-Step Test-Interview (TSTI): An observation-based method for pretesting self-completion questionnaires. *Survey Research Methods.* 2008;2(3). doi: 10.18148/srm/2008.v2i3.1669.
135. Tourangeau R. Cognitive Science and Survey Methods. In: Jabine TB, Tanur JM, Tourangeau R, editors. Cognitive Aspects of Survey design: Building a Bridge between Disciplines. Washington, DC: National Academy Press; 1984. p. 73-100.
136. Linacre J. A User's Guide to Winsteps Ministeps Rasch-Model Computer Programs. Program Manual. 2022.
137. Bond TG, Fox. CM. Applying the Rasch model: Fundamental measurement in the human sciences. New Jersey: Psychology Press; 2013.
138. Linacre JM. Optimizing rating scale category effectiveness. *Journal of Applied Measurement.* 2002;3(1):85-106.
139. Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. *Scottish Medical Journal.* 1957;2(5):200-15. doi: 10.1177/003693305700200504.
140. Swieten JCv, Koudstaal PJ, Visser MC, Schouten HJ, Gijn Jv. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke.* 1988;19(5):604-7. doi: doi:10.1161/01.STR.19.5.604.
141. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica.* 1983;67(6):361-70. doi: 10.1111/j.1600-0447.1983.tb09716.x.
142. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society.* 2005;53(4):695-9. doi: 10.1111/j.1532-5415.2005.53221.x.
143. Chiti G, Pantoni L. Use of Montreal Cognitive Assessment in patients with stroke. *Stroke.* 2014;45(10):3135-40. doi: 10.1161/strokeaha.114.004590.
144. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research.* 2011;20(10):1727-36. doi: 10.1007/s11136-011-9903-x.
145. Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the Functional Impact of Fatigue: Initial Validation of the Fatigue Impact Scale. *Clinical Infectious Diseases.* 1994;18:S79-S83. doi: 10.1093/clinids/18.Supplement_1.S79.
146. Skogestad IJ, Kirkevold M, Indredavik B, Gay CL, Lerdal A, Group N. Lack of content overlap and essential dimensions - A review of measures used for post-stroke fatigue. *Journal of Psychosomatic Research.* 2019;124:109759. doi: 10.1016/j.jpsychores.2019.109759.
147. Skogestad IJ, Kirkevold M, Larsson P, Borge CR, Indredavik B, Gay CL, Lerdal A. Post-stroke fatigue: an exploratory study with patients and health professionals to

develop a patient-reported outcome measure. *Journal of Patientreported Outcomes.* 2021;5(1):35. doi: 10.1186/s41687-021-00307-z.
148. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research.* 2018;27(5):1147-57. doi: 10.1007/s11136-018-1798-3.
149. Mills RJ, Pallant JF, Koufali M, Sharma A, Day S, Tennant A, Young CA. Validation of the Neurological Fatigue Index for stroke (NFI-Stroke). *Health & Quality of Life Outcomes.* 2012;10:51. doi: 10.1186/1477-7525-10-51.
150. McNair DM, Lorr M. An analysis of mood in neurotics. *Journal of Abnormal Psychology.* 1964;69:620-7. doi: 10.1037/h0040902.
151. Thomas E, Magilvy JK. Qualitative Rigor or Research Validity in Qualitative Research. *Journal for Specialists in Pediatric Nursing.* 2011;16(2):151-5. doi: 10.1111/j.1744-6155.2011.00283.x.
152. Lerdal A. Curvilinear Relationship Between Age and Post-Stroke Fatigue among Patients in the Acute Phase following First-Ever Stroke. *International journal of physical medicine & rehabilitation.* 2013;1(5). doi: 10.4172/2329-9096.1000141.
153. Cappelleri JC, Lundy JJ, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical therapeutics.* 2014;36(5):648-62.
154. Van de Winckel A, Kozlowski AJ, Johnston MV, Weaver J, Grampurohit N, Terhorst L, Juengst S, Ehrlich-Jones L, Heinemann AW, Melvin J, Sood P, Mallinson T. Reporting Guideline for RULER: Rasch Reporting Guideline for Rehabilitation Research: Explanation and Elaboration. *Archives of Physical Medicine and Rehabilitation.* 2022;103(7):1487-98. doi: 10.1016/j.apmr.2022.03.019.
155. Halcomb EJ. Mixed methods research: The issues beyond combining methods. *Journal of Advanced Nursing.* 2019;75(3):499-501. doi: 10.1111/jan.13877.
156. Polit DF, Beck CT. Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies.* 2010;47(11):1451-8. doi: 10.1016/j.ijnurstu.2010.06.004.
157. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. In: Services. USDoHaH, editor.1974.
158. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-4. doi: 10.1001/jama.2013.281053.
159. Kaiser K. Protecting respondent confidentiality in qualitative research. *Qualitative Health Research.* 2009;19(11):1632-41. doi: 10.1177/1049732309350879.
160. Sawatzky R, Kwon J-Y, Barclay R, Chauhan C, Frank L, van den Hout WB, Nielsen LK, Nolte S, Sprangers MAG. Implications of response shift for micro-, meso-, and macro-level healthcare decision-making using results of patient-reported outcome measures. *Quality of Life Research.* 2021;30(12):3343-57. doi: 10.1007/s11136-021-02766-9.
161. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007–2013). *Contemporary Clinical Trials.* 2015;43:1-9. doi: 10.1016/j.cct.2015.04.004.

162. Beurskens AJ, Bültmann U, Kant I, Vercoulen JH, Bleijenberg G, Swaen GM. Fatigue among working people: validity of a questionnaire measure. *Occupational and environmental medicine.* 2000;57(5):353-7.
163. Chen K, Marsh EB. Chronic post-stroke fatigue: It may no longer be about the stroke itself. *Clinical Neurology & Neurosurgery.* 2018;174:192-7. doi: 10.1016/j.clineuro.2018.09.027.
164. Wang SS, Wang JJ, Wang PX, Chen R. Determinants of fatigue after first-ever ischemic stroke during acute phase. *PLoS ONE [Electronic Resource].* 2014;9(10):e110037. doi: 10.1371/journal.pone.0110037.
165. Wang J, Li F, Xiao L, Peng F, Sun W, Li M, Liu D, Jiang Y, Guo R, Li H, Zhu W, Xu G, Liu X. Depressed TSH level as a predictor of poststroke fatigue in patients with acute ischemic stroke. *Neurology.* 2018;91(21):e1971-e8. doi: 10.1212/WNL.0000000000006534.
166. Abdulla FA, Al-Khamis FA, Alsulaiman AA, Alshami AM. Psychometric properties of an Arabic version of the fatigue severity scale in patients with stroke. *Topics in Stroke Rehabilitation.* 2019;26(6):448-55. doi: 10.1080/10749357.2019.1628465.
167. Teng CH, Phonyiam R, Davis LL, Anderson RA. Adaptation to poststroke fatigue in stroke survivors and their care partners: a scoping review. *Disability & Rehabilitation.* 2022:1-15. doi: 10.1080/09638288.2022.2084775.
168. Jablonski A. The multidimensional characteristics of symptoms reported by patients on hemodialysis. *Nephroogyl Nursing Journal.* 2007;34(1):29-37; quiz 8.
169. Cella D, Lai Js, Chang CH, Peterman A, Slavin M. Fatigue in cancer patients compared with fatigue in the general United States population. *Cancer.* 2002;94(2):528-38. doi: 10.1002/cncr.10245.
170. Lenaert B, van Kampen N, van Heugten C, Ponds R. Real-time measurement of post-stroke fatigue in daily life and its relationship with the retrospective Fatigue Severity Scale. *Neuropsychological Rehabilitation.* 2020:1-15. doi: 10.1080/09602011.2020.1854791.
171. Lerdal A, Gay CL, Aouizerat BE, Portillo CJ, Lee KA. Patterns of morning and evening fatigue among adults with HIV/AIDS. *Journal of Clinical Nursing.* 2011;20(15-16):2204-16. doi: 10.1111/j.1365-2702.2011.03751.x.
172. Rothrock NE, Kaiser KA, Cella D. Developing a valid patient-reported outcome measure. *Clinical Pharmacology and Therapy.* 2011;90(5):737-42. doi: 10.1038/clpt.2011.195.
173. Willis GB. Cognitive interviewing : a tool for improving questionnaire design. Thousand Oaks, Calif: Sage; 2005.
174. Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology.* 2008;8(1):33. doi: 10.1186/1471-2288-8-33.
175. Chen YK, Qu JF, Xiao WM, Li WY, Weng HY, Li W, Liu YL, Luo GP, Fang XW, Ungvari GS, Xiang YT. Poststroke fatigue: risk factors and its effect on functional status and health-related quality of life. *International Journal of Stroke.* 2015;10(4):506-12. doi: 10.1111/ijs.12409.
176. Hawkins L, Lincoln NB, Sprigg N, Ward NS, Mistri A, Tyrrell P, Worthington E, Drummond A. The Nottingham Fatigue After Stroke (NotFAST) study: results from follow-up six months after stroke. *Topics in Stroke Rehabilitation.* 2017;24(8):592-6. doi: 10.1080/10749357.2017.1368912.

177. Field J, Holmes MM, Newell D. PROMs data: can it be used to make decisions for individual patients? A narrative review. *Patient Related Outcome Measures.* 2019;10:233-41. doi: 10.2147/prom.S156291.

I

Journal of Psychosomatic Research 124 (2019) 109759

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

journal homepage: www.elsevier.com/locate/jpsychores

Review article

Lack of content overlap and essential dimensions – A review of measures used for post-stroke fatigue

Ingrid Johansen Skogestad[a,b,*], Marit Kirkevold[b], Bent Indredavik[c,d], Caryl L. Gay[e,f], Anners Lerdal[b,e], On behalf of the NORFAST (Norwegian Study of Fatigue After Stroke) Group

[a] Lovisenberg Diaconal Hospital, Oslo, Norway
[b] Department of Nursing Science, Institute of Health and Society, Faculty of Medicine, University of Oslo, Norway
[c] Department of Neuromedicine and Movement Science, NTNU, Trondheim, Norway
[d] Department of Stroke, St. Olavs Hospital, Trondheim, Norway
[e] Department of Research and Development, Lovisenberg Diaconal Hospital, Oslo, Norway
[f] Department of Family Health Care Nursing, University of California, San Francisco, USA

ARTICLE INFO

Keywords:
Content validity
Fatigue
Mixed methods
PROM
Rehabilitation
Stroke

ABSTRACT

Introduction: Post-stroke fatigue (PSF) is a highly prevalent and disabling condition. A major obstacle in PSF research is the lack of consensus on how to assess and diagnose fatigue after stroke. A wide variety of patient reported outcome measures (PROMs) are currently being used, none of which are developed specifically for stroke patients. The objectives of this study are to evaluate content validity in individual fatigue PROMs, and to identify similarities and differences through cross-comparison of PROMs.

Methods: We used a novel mixed-methods approach to evaluate content validity in fatigue PROMs. First, we performed a qualitative content analysis of items in eleven fatigue PROMs used in stroke populations, and then we used descriptive statistics and a similarity coefficient to investigate similarities and differences across instruments.

Results: The analysis of 156 items in eleven PROMs revealed 83 different items each representing a distinct attribute of fatigue. The results show that currently used fatigue PROMs omit important PSF-specific items, do not take into account the multidimensional nature of PSF and lack content overlap.

Summary: The wide variety of items and lack of overlap between fatigue PROMs illuminates the need for researchers to report why a specific PROM was used. PROMs that capture the specific experiences of patients with PSF are also needed to advance research on PSF and its etiology and treatment.

1. Introduction

Post-stroke fatigue (PSF) affects 50% of stroke survivors. This disabling condition has negative impacts on patients' rehabilitation, functioning and return to work [1–4]. PSF is often described as a feeling of physical or mental exhaustion, which may develop in connection with routine activities, following an acute stroke [5,6]. Despite recent efforts, no effective PSF treatment exists [7]. Lack of consensus on how to assess and diagnose fatigue after stroke is a major obstacle in PSF research [8]. Currently, PSF researchers and clinicians use a wide variety of patient reported outcome measures (PROMs), none of which are developed specifically for stroke [9,10].

Despite evidence that PSF is a multidimensional phenomenon, this is generally not reflected in the PROMs used in research [8]. The selection of PROMs is often based on psychometric testing such as evaluation of internal structure, reliability and responsiveness [11]. Based on this limited evidence, researchers conclude that the instruments are valid and reliable. However, testing of validity and reliability involves evaluation of a multitude of measurement properties. The first, and the one with most clinical impact, is content validity [12–14]. Content validity is an assessment of whether the content of PROMs reflects the construct that is to be measured, and refers to the relevance, comprehensiveness and comprehensibility of PROMs in relation to the construct, target population and context of use [13]. To achieve accurate interpretations of research results related to PSF, it is essential that PROMs include stroke-relevant items and cover relevant PSF dimensions [11]. There are three major concerns regarding item content when assessing PSF with a PROM not developed specifically for stroke

* Corresponding author at: Lovisenberggata 17, N-0456 Oslo, Norway.
E-mail addresses: ingridjohansen.skogestad@lds.no (I.J. Skogestad), marit.kirkevold@medisin.uio.no (M. Kirkevold), bent.indredavik@ntnu.no (B. Indredavik), caryl.gay@ucsf.edu (C.L. Gay), anners.lerdal@medisin.uio.no (A. Lerdal).

https://doi.org/10.1016/j.jpsychores.2019.109759
Received 29 March 2019; Received in revised form 19 June 2019; Accepted 30 June 2019

survivors. First, the PROM is likely to include items that are irrelevant for a stroke population. Second, the PROM might fail to identify issues that are a specific feature of PSF [15]. Third, items often refer to the impact of fatigue on physical functioning and participation in everyday activities, not considering that these dimensions might be directly limited by other stroke sequela, thus conflating the fatigue scores [10]. Item content can affect all other measurement properties. Irrelevant items may reduce internal consistency, structural validity, and interpretability of PROMs. Missing dimensions may reduce validity and responsiveness [15]. In addition, the use of different PROMs across studies might affect the generalizability and replicability of PSF research [16].

Based on the lack of a stroke-specific instrument to measure PSF, the overall aims of this study are to explore the item relevance, missing items and comprehensiveness in dimensions of PROMs used in PSF research, and to identify similarities and differences through cross-comparison of PROMs.

2. Methods

We performed a systematic literature search to identify all studies using PROMs to assess PSF. We then applied a mixed-methods approach in order to analyze the selected PROMs individually and to do a cross-comparison of items, sub-dimensions and dimensions [17].

2.1. Data collection

2.1.1. Search strategy

The systematic literature search was based on PRISMA-guidelines and performed in Medline, EMBASE, Cinahl and PsychINFO up to May 24, 2018. We developed the search strategy and conducted the search in collaboration with specialized research librarians. Search terms contained words equivalent to stroke, fatigue and rehabilitation (Supplemental Table I). There were no limitations regarding publication date, but we only included studies in English, Norwegian, Danish or Swedish.

2.1.2. Eligibility criteria

We included studies of individuals with stroke (both first incidence or recurrent, ischemic or hemorrhagic) that had PSF as a primary outcome, and studies testing psychometric properties of fatigue PROMs in a stroke population. Since the search intent was to investigate the frequency of use of different fatigue PROMs, all study designs (except reviews) were included. We defined the following exclusion criteria: lack of primary data, duplicate studies, < 50% of the study sample affected by stroke, lack of recognized fatigue PROMs, conference abstracts, case reports, and studies not available in full text [12]. In addition, we considered relevant studies from reference lists of the included studies and major PSF reviews. The first author (IJS) scrutinized the titles, abstracts, and full-text articles and considered inclusion based on the above-mentioned criteria. This resulted in 78 included studies (Fig. 1) that used 24 different PROMs for measuring fatigue (Supplemental Table II).

2.1.3. Selection of PROMs

The 24 different PROMs were used in anywhere from one to 39 studies on PSF. Eight PROMs were used in five or more studies and the rest were used in three studies or less. Given the aim to analyze the most commonly used PROMs in a stroke population, the research group (IJS, MK and AL) decided to set the inclusion cut-off at PROMs used in five studies or more. This resulted in inclusion of Fatigue Severity Scale [FSS] [18], Visual Analog Scale-Fatigue [VAS-F], Short Form-36 Vitality [SF-36 Vitality] [19], Checklist Individual Strength Fatigue Subscale [CIS] [20], Fatigue Assessment Scale [FAS] [21], Multidimensional Fatigue Inventory [MFI-20] [22], Mental Fatigue Scale [MFS] [23], and Modified Fatigue Impact Scale [M-FIS] [24]. In addition, we chose to include Profile of Mood States-Fatigue [POMS] [25], Neurological Fatigue Index-Stroke [NFI-Stroke] [26], and Dutch Multifactorial Fatigue Scale [DMFS] [27], as these PROMs were partly developed or tested in a stroke population [27–29]. This resulted in eleven PROMs included in the final analysis.

2.2. Qualitative data analysis

The eleven PROMs and affiliated items were subject to a qualitative content analysis [30,31], consisting of two stages: data preparation and organizing of data [32].

2.2.1. Preparation of data

The eleven PROMs contained a total of 156 items (i.e. questions), ranging from one to 38 items in each of the separate PROMs (Supplemental Table III). We analyzed the items, but not the scoring method (Supplemental Table III), as this was beyond the scope of this study. NVivo (Version 11.4.1.1064) was used to keep track of the analysis [33].

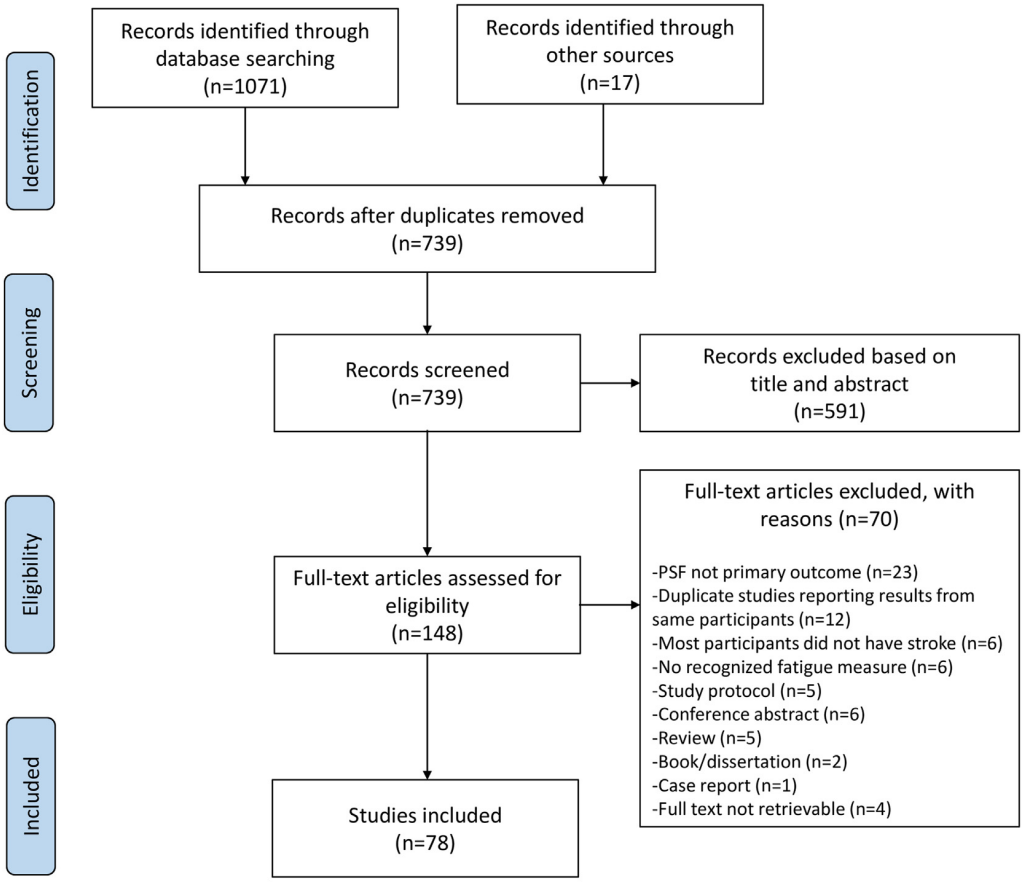
2.2.2. Organizing of data

This second stage consisted of analyzing and comparing items, organizing items into sub-dimensions, and identifying overarching dimensions. The first step in this inductive analytical process included reading and re-reading all PROMs separately to get a sense of each instrument in its entirety. Then all 156 items were compared to each other and classified in two different ways; as *single items* (meaning no other items were similar to that specific item) or *similar items* (this included items that were either identical in wording or had substantially the same meaning). E.g. *"Physically, I feel in a good shape"* and *"Physically, I feel I am in excellent condition"* were interpreted as similar items having the same substantial meaning, whereas the following borderline case *"My muscles have felt weak"* and *"Sometimes I lose my body strength"* were interpreted as different items because they lack a common concept. Studies have shown that even small differences in the items used, i.e. asking individuals to state if they feel "fatigued" vs. "lack of energy" may lead to considerable heterogeneity in prevalence rates [34]. Thus, in order to group items as similar, they had to feature very comparable content. Despite some analytical challenges, most single items had substantial differences such as "*I can follow conversations without getting tired*" and "*I have had difficulties making decisions*".

After comparison and classification of the total 156 items, the analysis resulted in 83 unique items, each representing a distinct attribute of fatigue. The 83 items were organized into sub-dimensions, and then we identified four central dimensions (Supplemental Fig. I). In order to display all items in Fig. 2, the items were given shortened names (e.g., "I feel rested" was called "rested"). Despite presentation of the analysis as a linear process, the analysis was iterative and we went back and forth in the analytical process. The analysis was conducted independently by IJS and MK and was continuously discussed in the research group until agreement was reached.

2.3. Statistics

Results from the qualitative analysis of items, sub-dimensions and dimensions were quantified and presented with descriptive statistics. Based on the analysis of items as either single or similar (0 or 1), we estimated statistical content overlap between PROMs using the Jaccard Index (i.e. Jaccard similarity coefficient). This is a method for comparing the similarity and diversity of sample sets and is defined as the size of the intersection divided by the size of the union of sample sets (i.e. the number of shared items divided by the total number of items in any two PROMs) [35].



Note : Flow chart reported based on the PRISMA-guidelines (2009).

Fig. 1. Flow chart of literature search and inclusion of studies [50].

3. Results

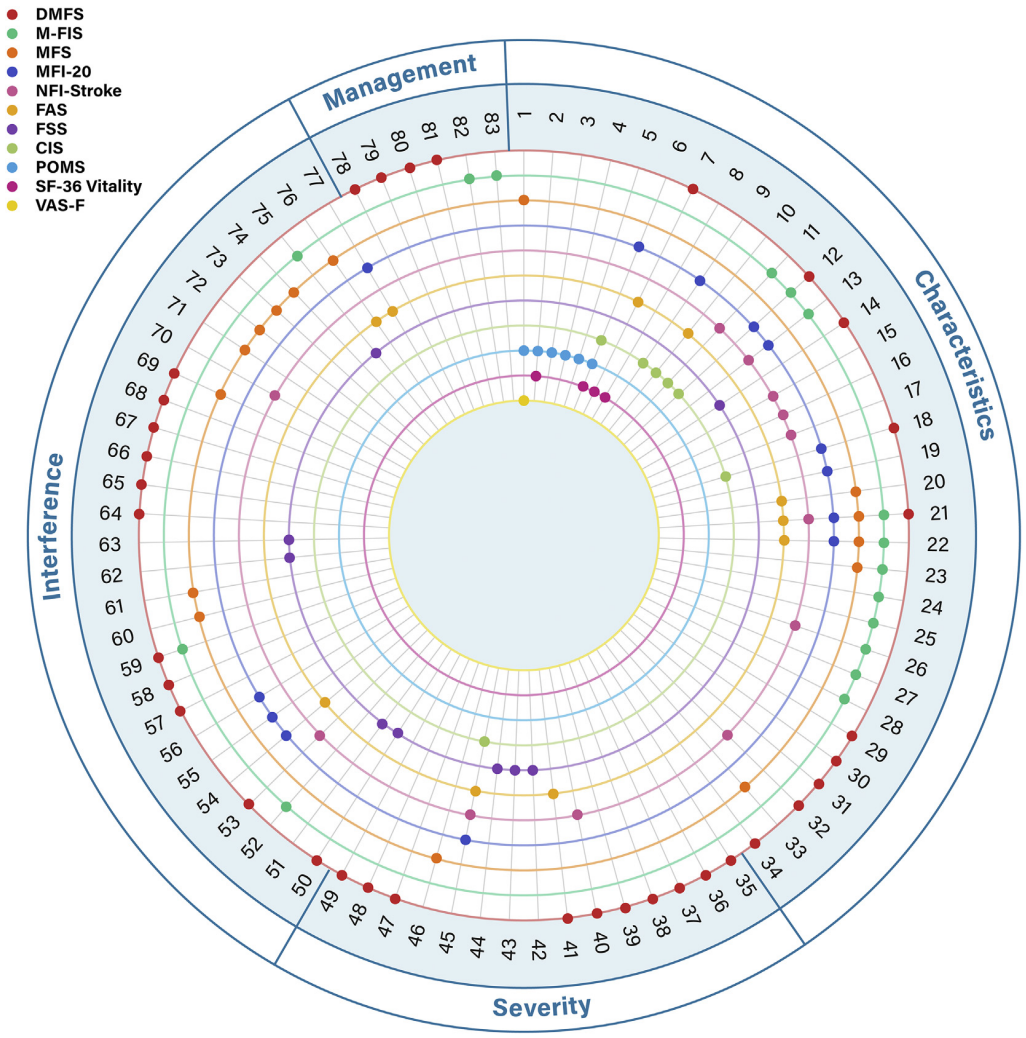
Analysis of the eleven PROMs included in this study revealed 83 unique items. We found four item dimensions and each dimension had sub-dimensions (Fig. 2). The dimension *characteristics* appeared early in the analysis, and involves typical attributes related to fatigue awareness and perception. Two sub-dimensions were identified: quality and diurnal variations. Quality was further organized into three groups: "general" items measuring the patient's overall feelings/condition (e.g. fatigued, tiredness and level of energy), "physical" items related to the state of the body (e.g. weak, fit, body strength), and "mental" items measuring cognitive functioning (e.g. concentration, thinking, coordination). Several of the "physical" and "mental" items are not directly related to fatigue (e.g. physically, I feel I am in excellent condition, and I have been forgetful). Diurnal variations include items measuring timing of fatigue (e.g. I get fatigued in the afternoon), and are considered a part of the *characteristics* dimension as it may describe individual fatigue patterns.

The *severity* dimension involves items related to the degree to which fatigue is bothering a person and to the intensity of fatigue. This dimension has three sub-dimensions. Impact measures how much suffering fatigue causes for the person (e.g. serious complaint, severe fatigue, bothered by fatigue). Onset speed includes items measuring how easily tired and fatigued a patient is and is related to severity because it describes the degree of vulnerability to reduced capacity. Onset speed is not considered to be a defining attribute of fatigue, and thus does not fit within the *characteristics* dimension.Recovery time includes items measuring the possibility and time needed to regain restoration. This relates to *severity* because a long recovery time might reflect how seriously one's condition is affected.

The *interference* dimension has three sub-dimensions.Behavioral interference items measure general activities that are affected when fatigued (e.g. hinders duties, do little, increased sleep). Physical interference items relate to how fatigue affects one's physical condition (e.g., fatigue prevents sustained physical functioning, body aches when fatigued). These items explicitly measure physical reactions during fatigue, and differ from the physical items under the quality sub-dimension of *characteristics* since those items reflect physical functioning as a part of the overall feeling or characteristics of fatigue (e.g. "fatigue interferes with physical function" vs. "physically, I feel I am in a good condition"). Mental interference items measure mental function when fatigued (e.g. cannot think when fatigued, make mistakes when fatigued). Items measuring sensitivity to stress, light and sound, as well as lack of motivation are considered part of mental interference.

The fourth dimension contains items related to *management* of fatigue and has one sub-dimension, coping, which measures how the patient deals with their fatigue (e.g. by avoiding overtiredness, planning to rest, and limiting physical activities).

Of the 83 items, 75% were categorized as either characteristics or interference, and 66% appear only in one PROM (Table 1). The most-used PROMs (FSS, VAS-F and SF-36 Vitality) cover only a few dimensions. Most items in FSS measure *interference*, and VAS-F and SF-36 Vitality only measure quality (Table 2). The majority of all items in this study are classified in the quality sub-dimension or the *interference* dimension (Table 1). MFS and DMFS are the only PROMs that include an item about diurnal variations of fatigue (Table 2). None of the



Characteristics

Quality

General Items

1. Fatigued
2. Worn out
3. Sluggish
4. Weary
5. Sleepy
6. Tired
7. Energy
8. Full of pep
9. Rested

Physical items

10. Exhausted
11. Feel weak
12. Physical condition
13. Physical capacity
14. Exercise causes fatigue
15. Body strength
16. Heavy limbs
17. Body unable to keep up
18. Feel fit
19. Feel active

Mental items

20. Mentally exhausted
21. Concentration
22. Thinking
23. Memory
24. Less alert
25. Organizing thoughts
26. Coordination
27. Attention
28. Making decisions
29. Conversations
30. Tired by emotions
31. Noise-induced fatigue
32. Thinking-induced fatigue

Diurnal variations

33. 24-hour variations
34. Fatigued in afternoon

Severity

Fatigue impact

35. Affects whole life
36. Serious complaint
37. Complaints get worse
38. Suffer terribly
39. Severe fatigue
40. Tired every day
41. Bothered by fatigue
42. Causes problems
43. Most disabling symptom

Onset speed

44. Easily fatigued
45. Easily tired

Recovery time

46. Long mental recovery
47. Recover easily
48. Fatigued next day
49. Wake up rested

Interference

Behavioral

50. Duties
51. Work, family, social life
52. Limited ability
53. Unable to go further
54. Think I do little
55. Get little done
56. Few plans
57. Say things I regret
58. Other people notice
59. Need to rest
60. Decreased sleep
61. Increased sleep

Physical

62. Interferes with physical function
63. Prevents sustained function
64. Induces headache
65. Body aches

Mental

66. Letting thoughts go
67. Unable to think
68. Makes mistakes
69. Emotional tendency
70. Speech problems
71. Stress sensitivity
72. Light sensitivity
73. Noise sensitivity
74. Irritability
75. Low motivation
76. Problems starting things
77. Feel no desire

Management

Coping

78. Let myself get tired out
79. Avoid overtiredness
80. Plan to rest
81. Finish tasks
82. Limit physical activity
83. Pace physical activity

Note. *Each colored circle represents a PROM and each dot represents one item. We have analyzed all the items and categorized them in dimensions (characteristics, severity, interference and management) and sub-dimensions.*

Fig. 2. Distribution of 83 different items in eleven fatigue PROMs used to measure fatigue in stroke survivors. Organized in relation to four fatigue dimensions.

Table 1
Number and percentage of items (n = 83) appearing in each dimension and sub-dimension and in one or more PROMs.

Dimension	Number of items	% of items
Sub-dimension		
Characteristics	34	41%
Quality	32	39%
General	9	11%
Physical	10	12%
Mental	13	16%
Diurnal variations	2	2%
Severity	15	18%
Fatigue impact	9	11%
Onset speed	2	2%
Recovery time	4	5%
Interference	28	34%
Behavioral	12	14%
Physical	4	5%
Mental	12	14%
Management	6	7%
Coping	6	7%
Number of PROMs containing each item		
1	55	66%
2	18	22%
3	5	6%
4	4	5%
6	1	1%

Note. Percentages calculated based on the 83 unique fatigue items identified through the qualitative analysis as displayed in Fig. 3.

PROMs in this study contain items on pre-stroke fatigue or a comprehensive assessment of diurnal variations, fatigue trajectory or aggravating factors (Table 2), which may have implications for management of PSF [2]. In total, this indicates a lack of multidimensionality among the fatigue PROMs in this study.

There is considerable lack of content overlap between PROMs, and the three most commonly used PROMs (FSS, VAS-F and SF-36 Vitality) do not overlap at all with each other (Table 3). FSS does not overlap with other PROMs except for the DMFS (38 items). MFI-20 has the highest degree of overlap with other PROMs, with 33% overlap with CIS and 27% overlap with FAS (Table 3).

4. Discussion

This study shows that the most commonly used PSF PROMs do not address potentially relevant aspects specific to PSF, such as diurnal variations and pre-stroke fatigue, do not account for the multidimensional nature of PSF and lack content overlap. Our analysis suggests that these PSF PROMs have important limitations, which might impair further progress in PSF research and patient care.

The analysis revealed 83 unique fatigue-related items in eleven PROMs used in stroke research. We classified the items in four main dimensions covering different fatigue-related aspects. Despite the consensus that fatigue is a multidimensional phenomenon [36–38], the three most widely used PROMs in PSF research (FSS, VAS-F, and SF-36 Vitality) mainly assess fatigue characteristics or interference. One previous study on PSF PROMs considered FSS to have insufficient face validity [28], and it is unclear why FSS remains the preferred PROM in stroke research.

One of the core characteristics of the experience of PSF is the unpredictability of when and why fatigue occurs [39]. Fatigue is not a constant symptom, but fluctuates throughout the day [2,40]. However, only two PROMs in this study had included this dimension, having one item each measuring diurnal variations. These two items lack specific details needed to investigate possible fluctuations, such as percent of fatigue during a typical day or week, as suggested in the Lynch case definition of post-stroke fatigue [41].

Most PROMs in this study, such as the FSS, were developed based on expert opinion and theory, and designed for conditions other than stroke. Many items in these PROMs refer to factors that are often directly limited by the stroke. For example, most items in the group "physical" (Fig. 2) under the quality sub-dimension are examples of normal physical stroke sequela (e.g. feel weak and low physical capacity), and these items are indistinguishable from the general consequences of the stroke and might bias the results [10,28].

The considerable lack of item overlap between the PROMs of post-stroke fatigue showed in this study, may affect the replicability and generalizability of PSF research. Current PSF research yields inconsistent findings regarding prevalence rates and associated factors, such as depression, anxiety, cognitive functioning, pain, sleep problems, lesion location, and co-morbidities [1,7,34,42]. Falconer et al. found significant variation in the prevalence of PSF depending on the PROMs and items used for detecting it, which could partly be explained by differences in terminology and descriptors used to assess fatigue [34].

Knowledge on the etiology of PSF is limited [38], but some studies report associations between PSF and biological as well as immunological factors [43]. Given the hypothesis that PSF is caused by cerebrovascular pathology, it would be of interest to discern pre-stroke fatigue conditions from PSF in order to elaborate our etiological understanding of PSF [2,44]. However, the PROMs in this study have no items measuring pre-stroke fatigue. Only a few previous studies have included pre-stroke fatigue assessment [45,46], and found that that 27–31% of stroke survivors reported having fatigue before their stroke [45,47–48]. In further studies it could be of relevance to distinguish

Table 2
Distribution of items in each PROM organized into fatigue dimensions and sub-dimensions, % (n).

	Characteristics		Severity			Interference			Management	Total number of items, N
PROM	Quality	Diurnal variations	Fatigue impact	Onset speed	Recovery time	Behavioral	Physical	Mental	Coping	
FSS	11% (1)	0% (0)	22% (2)	11% (1)	0% (0)	22% (2)	22% (2)	11% (1)	0% (0)	9
VAS-F	100% (1)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	1
SF-36V	100% (4)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	4
CIS	86% (6)	0% (0)	0% (0)	14% (1)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	7[a]
FAS	50% (5)	0% (0)	10% (1)	10% (1)	0% (0)	10% (1)	0% (0)	20% (2)	0% (0)	10
MFI-20	61% (8)	0% (0)	0% (0)	8% (1)	0% (0)	23% (3)	0% (0)	8% (1)	0% (0)	13[a]
MFS	33% (5)	7% (1)	0% (0)	0% (0)	7% (1)	13% (2)	0% (0)	40% (6)	0% (0)	15
M-FIS	68% (11)	0% (0)	0% (0)	0% (0)	0% (0)	13% (2)	0% (0)	6% (1)	13% (2)	16[a]
POMS	100% (6)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	6
NFI-Stroke	67% (8)	0% (0)	8% (1)	8% (1)	0% (0)	8% (1)	0% (0)	8% (1)	0% (0)	12
DMFS	26% (9)	3% (1)	20% (7)	0% (0)	9% (3)	14% (5)	6% (2)	11% (4)	11% (4)	35[a]

Note. Table showing the percentage and number (n) of items in each PROM measuring the different fatigue dimensions and sub-dimensions.
[a] This PROM contains items that were determined in the qualitative analysis to be similar, and thus, the total number of unique items in each instrument in this table is less than in the original version of the instrument.

Table 3
Overlap of items as measured by the Jaccard index for eleven PROMs.

PROM	FSS	VAS-F	SF-36 V	CIS[a]	FAS	MFI-20[a]	MFS	M-FIS[a]	POMS	NFI-Stroke	DMFS[a]
FSS	1.000	.000	.000	.000	.000	.000	.000	.042	.000	.000	.048
VAS-F	.000	1.000	.000	.000	.000	.000	.067	.000	.167	.000	.000
SF-36V	.000	.000	1.000	.100	.077	.063	.000	.000	.250	.000	.026
CIS[a]	.000	.000	.100	1.000	.133	.333	.000	.095	.083	.118	.050
FAS	.000	.000	.077	.133	1.000	.278	.190	.083	.000	.100	.071
MFI-20[a]	.000	.000	.063	.333	.278	1.000	.077	.160	.056	.136	.067
MFS	.000	.067	.000	.000	.190	.077	1.000	.107	.050	.038	.042
M-FIS[a]	.042	.000	.000	.095	.083	.160	.107	1.000	.000	.167	.063
POMS	.000	.167	.250	.083	.000	.056	.050	.000	1.000	.000	.000
NFI-Stroke	.000	.000	.000	.118	.100	.136	.038	.167	.000	1.000	.093
DMFS[a]	.048	.000	.026	.050	.071	.067	.042	.063	.000	.093	1.000

Note: The Jaccard index measures similarity between two sets of data with a range from 0 (no overlap) to 1 (total overlap), and is calculated as the number of shared items divided by the total number of items in any two PROMs. The PROMs are arranged by the most used to the least used instrument.

[a] This PROM contains items that were determined in the qualitative analysis to be similar, and thus, the total number of unique items in each instrument in this table is less than in the original version of the instrument (i.e. CIS 7 items, MFI-20 13 items, M-FIS 16 items and DMFS 35 items).

pre-stroke fatigue from PSF. However, potential recall bias and response shift might influence the validity of PROM answers if the recall interval is lengthy, thus collecting pre-stroke data is only advisable in the acute phase after stroke [49].

There are some limitations to this study. First, we analyzed a selected group of fatigue PROMs. Inclusion of other PROMs could have yielded a different result. However, we believe that the inclusion criteria cover the most relevant PROMs, as we only included the most commonly used fatigue PROMs, or those at least partly developed for a stroke population. Second, during the analysis of items there were some borderline cases, such as whether *single* or *similar* refer to different facets of fatigue and accordingly should be classified as separate items. We carefully discussed how conservative this part of the analysis should be. Generally, we chose to classify items as similar rather than different. Thus, if anything, the results overestimate, rather than underestimate, the homogeneity and overlap between the PROMs.

5. Summary

The wide variety of PROMs and dimensions used to assess PSF demonstrates lack of consensus regarding what needs to be assessed when diagnosing and measuring fatigue in stroke populations. Studies on PSF should report why they used a particular fatigue PROM, and what dimension of fatigue they intended to measure (characteristics, severity, interference and/or management). In order to move the research on the etiology of PSF forward, PROMs that capture the relevant experiences of patients with PSF are needed.

Acknowledgement

We would like to acknowledge specialized research librarians Mia Alexandra Ølnes and Fredrik Solvang Pettersen at Lovisenberg Diaconale University College for their help with the systematic literature search. We would also like to thank our graphic designer Eira Taksdal for designing Fig. 2 and our statistician Milada Småstuen at University of Oslo for cross-checking the Jaccard index.

Sources of funding

Ingrid Johansen Skogestad is funded by the non-profit Norwegian Health Association doctoral studentship. The funding source had no involvement in conducting or reporting of this study.

Disclosures

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpsychores.2019.109759.

References

[1] T.B. Cumming, A.B. Yeo, J. Marquez, L. Churilov, J.-M. Annoni, U. Badaru, et al., Investigating post-stroke fatigue: an individual participant data meta-analysis, J. Psychosom. Res. 113 (2018) 107–112 Oct.
[2] A. Lerdal, L.N. Bakken, S.E. Kouwenhoven, G. Pedersen, M. Kirkevold, A. Finset, et al., Poststroke fatigue–a review, J. Pain Symptom Manag. 38 (2009) 928–949.
[3] E.L. Glader, B. Stegmayr, K. Asplund, Poststroke fatigue: a 2-year follow-up study of stroke patients in Sweden, Stroke. 33 (2002) 1327–1333.
[4] G. Andersen, D. Christensen, M. Kirkevold, S.P. Johnsen, Post-stroke fatigue and return to work: a 2-year follow-up, Acta Neurol. Scand. 125 (2012) 248–253.
[5] M. Kirkevold, D. Christensen, G. Andersen, S.P. Johansen, I. Harder, Fatigue after stroke: manifestations and strategies, Disabil. Rehabil. 34 (2012) 665–670.
[6] F. Staub, J. Bogousslavsky, Fatigue after stroke: a major but neglected issue, Cerebrovasc. Dis. 12 (2001) 75–81.
[7] S. Wu, M.A. Kutlubaev, H.Y. Chun, E. Cowey, A. Pollock, M.R. Macleod, M. Dennis, E. Keane, M. Sharpe, G.E. Mead, Interventions for post-stroke fatigue, Cochrane Database Syst. Rev. (7) (2015) CD007030, , https://doi.org/10.1002/14651858.CD007030.pub3.
[8] M. Nadarajah, H.T. Goh, Post-stroke fatigue: a review on prevalence, correlates, measurement, and management, Top. Stroke Rehabil. 22 (2015) 208–220.
[9] S. Choi-Kwon, P.H. Mitchell, J.S. Kim, Nursing interventions for poststroke fatigue, Stroke. 46 (2015) e224–e227.
[10] S.F. Tyson, P. Brown, How to measure fatigue in neurological conditions? A systematic review of psychometric properties and clinical utility of measures used so far, Clin. Rehabil. 28 (2014) 804–816.
[11] A.J. Dittner, S.C. Wessely, R.G. Brown, The assessment of fatigue: a practical guide for clinicians and researchers, J. Psychosom. Res. 56 (2004) 157–170.
[12] C.A.C. Prinsen, L.B. Mokkink, L.M. Bouter, J. Alonso, D.L. Patrick, H.C.W. de Vet, et al., COSMIN guideline for systematic reviews of patient-reported outcome measures, Qual. Life Res. 27 (2018) 1147–1157.
[13] C.B. Terwee, C.A.C. Prinsen, A. Chiarotto, M.J. Westerman, D.L. Patrick, J. Alonso, et al., COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study, Qual. Life Res. 27 (2018) 1159–1170.
[14] A. Chiarotto, R.W. Ostelo, M. Boers, C.B. Terwee, A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain, J. Clin. Epidemiol. 95 (2018) 73–93.
[15] S.P. McKenna, Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science, BMC Med. 9 (2011) 86.
[16] N. Kruithof, M.H. Van Cleef, S.M. Rasquin, T.J. Bovend'Eerdt, Screening poststroke fatigue; feasibility and validation of an instrument for the screening of poststroke fatigue throughout the rehabilitation process, J. Stroke Cerebrovasc. Dis. 25 (2016) 188–196.
[17] E.I. Fried, The 52 symptoms of major depression: lack of content overlap among seven common depression scales, J. Affect. Disord. 208 (2017) 191–197.
[18] L.B. Krupp, N.G. LaRocca, J. Muir-Nash, A.D. Steinberg, The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus, Arch. Neurol. 46 (1989) 1121–1123.
[19] J.E. Ware Jr., C.D. Sherbourne, The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection, Med. Care. 30 (1992) 473–483.
[20] J.H. Vercoulen, C.M. Swanink, J.F. Fennis, J.M. Galama, J.W. van der Meer, G. Bleijenberg, Dimensional assessment of chronic fatigue syndrome, J. Psychosom. Res. 38 (1994) 383–392.

[21] H.J. Michielsen, J. de Vries, G.L. van Heck, F.J.R. van de Vijver, K. Sijtsma, Examination of the dimensionality of fatigue: the construction of the Fatigue Assessment Scale (FAS), Eur. J. Psychol. Assess. 20 (2004) 39–48.
[22] E.M. Smets, B. Garssen, B. Bonke, J.C. De Haes, The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue, J. Psychosom. Res. 39 (1995) 315–325.
[23] B. Johansson, A. Starmark, P. Berglund, M. Rodholm, L. Ronnback, A self-assessment questionnaire for mental fatigue and related symptoms after neurological disorders and injuries, Brain Inj. 24 (2010) 2–12.
[24] J.D. Fisk, P.G. Ritvo, L. Ross, D.A. Haase, T.J. Marrie, W.F. Schlech, Measuring the functional impact of fatigue: initial validation of the fatigue impact scale, Clin. Infect. Dis. 18 (Suppl. 1) (1994) S79–S83.
[25] D.M. McNair, M. Lorr, An analysis of mood in neurotics, J. Abnorm. Psychol. 69 (1964) 620–627.
[26] R.J. Mills, C.A. Young, J.F. Pallant, A. Tennant, Development of a patient reported outcome scale for fatigue in multiple sclerosis: the neurological fatigue index (NFI-MS), Health Qual. Life Outcomes 8 (2010) 22.
[27] A.C. Visser-Keizer, A. Hogenkamp, H.J. Westerhof-Evers, I.J.L. Egberink, J.M. Spikman, Dutch multifactor fatigue scale: a new scale to measure the different aspects of fatigue after acquired brain injury, Arch. Phys. Med. Rehabil. 96 (2015) 1056–1063.
[28] G. Mead, J. Lynch, C. Greig, A. Young, S. Lewis, M. Sharpe, Evaluation of fatigue scales in stroke patients, Stroke. 38 (2007) 2090–2095.
[29] R.J. Mills, J.F. Pallant, M. Koufali, A. Sharma, S. Day, A. Tennant, et al., Validation of the neurological fatigue index for stroke (NFI-stroke), Health Qual. Life Outcomes 10 (2012) 51.
[30] M. Sandelowski, Triangles and crystals: on the geometry of qualitative research, Res. Nurs. Health 18 (1995) 569–574.
[31] J.D. Comins, M.R. Krogsgaard, J. Brodersen, Ensuring face validity in patient-related outcome scores–a matter of content, Knee. 20 (2013) 72–78.
[32] S. Elo, H. Kyngas, The qualitative content analysis process, J. Adv. Nurs. 62 (2008) 107–115.
[33] C. Houghton, K. Murphy, B. Meehan, J. Thomas, D. Brooker, D. Casey, From screening to synthesis: using nvivo to enhance transparency in qualitative evidence synthesis, J. Clin. Nurs. 26 (2017) 873–881.
[34] M. Falconer, S. Walsh, J.A. Harbison, Estimated prevalence of fatigue following stroke and transient ischemic attack is dependent on terminology used and patient gender, J. Stroke Cerebrovasc. Dis. 19 (2010) 431–434.
[35] M. Levandowsky, D. Winter, Distance between Sets, Nature. 234 (1971) 34.
[36] L. Whitehead, The measurement of fatigue in chronic illness: a systematic review of unidimensional and multidimensional fatigue measures, J. Pain Symptom Manag. 37 (2009) 107–128.
[37] J.L. Hinkle, K.J. Becker, J.S. Kim, S. Choi-Kwon, K.L. Saban, N. McNair, et al., Poststroke fatigue: emerging evidence and approaches to management: a scientific statement for healthcare professionals from the American Heart Association, Stroke. 48 (2017) (e159-e70).
[38] M. Acciarresi, J. Bogousslavsky, M. Paciaroni, Post-stroke fatigue: epidemiology, clinical characteristics and treatment, Eur. Neurol. 72 (2014) 255–261.
[39] G. Eilertsen, H. Ormstad, M. Kirkevold, Experiences of poststroke fatigue: qualitative meta-synthesis, J. Adv. Nurs. 69 (2013) 514–525.
[40] J.H. White, K.R. Gray, P. Magin, J. Attia, J. Sturm, G. Carter, et al., Exploring the experience of post-stroke fatigue in community dwelling stroke survivors: a prospective qualitative study, Disabil. Rehabil. 34 (2012) 1376–1384.
[41] J. Lynch, G. Mead, C. Greig, A. Young, S. Lewis, M. Sharpe, Fatigue after stroke: the development and evaluation of a case definition, J. Psychosom. Res. 63 (2007) 539–544.
[42] S. Wu, A. Barugh, M. Macleod, G. Mead, Psychological associations of poststroke fatigue: a systematic review and meta-analysis, Stroke. 45 (2014) 1778–1783.
[43] R. Aarnes, J. Stubberud, A. Lerdal, A literature review of factors associated with fatigue after stroke and a proposal for a framework for clinical utility, Neuropsychol. Rehabil. (2019) 1–28.
[44] G.E. Mead, C. Graham, P. Dorman, S.K. Bruins, S.C. Lewis, M.S. Dennis, et al., Fatigue after stroke: baseline predictors and influence on survival. Analysis of data from UK patients recruited in the international stroke trial, PLoS One 6 (2011) e16988.
[45] A. Lerdal, L.N. Bakken, E.F. Rasmussen, C. Beiermann, S. Ryen, S. Pynten, et al., Physical impairment, depressive symptoms and pre-stroke fatigue are related to fatigue in the acute phase after stroke, Disabil. Rehabil. 33 (2011) 334–342.
[46] F. McKechnie, S. Lewis, G. Mead, A pilot observational study of the association between fatigue after stroke and C-reactive protein, J. R. Coll. Physicians Edinb. 40 (2010) 9–12.
[47] T. Egerton, A. Hokstad, T. Askim, J. Bernhardt, B. Indredavik, Prevalence of fatigue in patients 3 months after stroke and association with early motor activity: a prospective study comparing stroke patients with a matched general population cohort, BMC Neurol. 15 (2015) 181.
[48] S.S. Wang, J.J. Wang, P.X. Wang, R. Chen, Determinants of fatigue after first-ever ischemic stroke during acute phase, PLoS One 9 (2014) e110037.
[49] E. Kwong, N. Black, Retrospectively patient-reported pre-event health status showed strong association and agreement with contemporaneous reports, J. Clin. Epidemiol. 81 (2017) 22–32.
[50] D.Liberati Moher, A. Tetzlaff, J.DG Altmann, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statment, J. Clin. Epidemiol. 62 (2009) 1006–1012.

II

Skogestad et al. Journal of Patient-Reported Outcomes (2021) 5:35
https://doi.org/10.1186/s41687-021-00307-z

Journal of Patient-Reported Outcomes

RESEARCH Open Access

Post-stroke fatigue: an exploratory study with patients and health professionals to develop a patient-reported outcome measure

Ingrid Johansen Skogestad[1,2*†], Marit Kirkevold[2,3†], Petra Larsson[4,5†], Christine Råheim Borge[5,6†], Bent Indredavik[7,8†], Caryl L. Gay[6,9†] and Anners Lerdal[5,6]

Abstract

Background: Post-stroke fatigue (PSF) is commonly reported and described as disabling by patients recovering from stroke. However, a major challenge is how to accurately diagnose and assess PSF. Therefore, the aim of this study was to explore PSF as it is experienced by stroke survivors and described by health professionals to guide future development of a PSF-specific PROM.

Methods: Individual semi-structured interviews were conducted with stroke survivors experiencing PSF ($n = 9$) and three focus groups were conducted with health professionals ($n = 16$). Data were analyzed through inductive content analysis.

Results: The analysis revealed four themes illustrating the experience and descriptions of PSF: 1) PSF characteristics, 2) interfering and aggravating factors, 3) management, and 4) PSF awareness, which refers to stroke survivors first becoming aware of PSF after their initial hospital admission.

Conclusion: This study highlights the complexity and multidimensionality of PSF. The results from this study will guide future development of a PSF-PROM and support its content validity.

Keywords: Fatigue, Stroke, Rehabilitation, Qualitative research, Patient-reported outcome measure

Introduction

Post-stroke fatigue (PSF) is one of the most common symptoms 3 months after stroke [1] and can have negative implications for patients' rehabilitation, physical function, activities in daily life, and quality of life [2–4]. Despite the high prevalence (25–85%) [5] and disabling nature, evidence-based interventions to prevent and treat PSF are currently lacking [6]. A major challenge is to achieve accurate diagnostics of PSF, a prerequisite for the development of novel preventive and therapeutic measures [7].

PSF can be defined as lack of energy, or increased need to rest, every day or nearly every day, leading to difficulties partaking in everyday activities [8]. The diagnosis of PSF is traditionally based on patient-reported outcome measures (PROMs). However, a recent review showed that the PROMs most commonly used to measure fatigue in stroke survivors have several limitations [9]. This is in line with a previous review, which did not find any fatigue PROM that met critical criteria for an ideal instrument [10]. Moreover, existing fatigue PROMs mostly include only one or two fatigue dimensions, such

* Correspondence: ingridjohansen.skogestad@lds.no
†Ingrid Johansen Skogestad, Marit Kirkevold, Petra Larsson, Christine Råheim Borge, Bent Indredavik and Caryl L. Gay contributed equally to this work.
[1]Medical Department, Lovisenberg Diaconal Hospital, Oslo, Norway
[2]Department of Nursing Science, Faculty of Medicine, Institute of Health and Society, University of Oslo, Oslo, Norway
Full list of author information is available at the end of the article

as the intensity or impact of fatigue, and do not assess other potentially relevant dimensions of the fatigue experience [9]. In addition, existing fatigue PROMs are not developed specifically for stroke survivors [11]. The former shortcomings of these fatigue PROMs may partly be explained by the lack of involvement of patients and health professionals in the instrument development process [9], which is strongly recommended in guidelines [12, 13] and often a part of the health technology assessment for medicinal product approval and reimbursement [14, 15].

Although relatively few qualitative studies have been published on PSF, they consistently describe fatigue as having several dimensions [16]. This includes core characteristics of PSF [16], different factors contributing to fatigue [17, 18], and various aspects of daily life affected by fatigue [19]. However, previous qualitative studies have not aimed to guide item development in a new PSF-specific PROM. Qualitative studies, through interviews or focus groups, are the preferred method to establish content validity in new PROMs [20]. Qualitative studies have the advantage of being able to directly engage with the experts, who can provide a comprehensive understanding of the construct to be measured. Experts include both patients and health professionals. Patients have the symptom experience, and health professionals have clinical experience treating these patients and can describe typical characteristics, consequences and management of the symptom [13, 20–24]. A qualitative study will provide a deeper understanding of PSF, how it impacts life, and management strategies, which is critical to addressing the limitations of existing fatigue PROMs and informing the content and structure of a PROM specifically developed to measure PSF.

The overall aim of this study was to explore PSF as it is experienced by stroke survivors and described by health professionals to guide future development of a new PSF-specific PROM.

Methods

Design

In this qualitative study, individual semi-structured interviews were conducted with stroke survivors with PSF and multi-disciplinary focus groups with health professionals who provide clinical care to stroke patients. This study was conducted in Norway as part of a larger research project, which includes three sub-studies with the overall aim of developing and testing a new PROM for PSF. This study has followed COSMIN criteria for establishing content validity in PROMs [13], as well as the COREQ checklist for qualitative studies (Online resource 1) [25].

Participants

Stroke survivors

Nine stroke survivors were included in this study. The inclusion criteria were: (1) stroke within the last 2 years, (2) 18 years or older, and (3) meeting the diagnostic criteria for PSF as defined by a clinical interview [8]. For the first criterion, all types of stroke were included, as defined by the International Classification of Diseases 10th edition (ICD-10) [26] and included codes for ischemic stroke (I63), non-traumatic intracerebral haemorrhage (I61), and stroke, not specified as haemorrhage or infarction (I64). The time period of stroke within the last 2 years was chosen because the level of post-stroke fatigue has previously been shown to remain constant for up to 2 years [27]. A purposive sampling strategy was used and aimed to recruit participants with diversity in age, gender, physical impairment, communication disorders and living accommodations. As these demographic and clinical variables may influence patients' PSF experiences, such a purposive sampling strategy was intended to provide different perspectives and descriptions of PSF [23, 24]. Five stroke participants were recruited through a Facebook page for a stroke user organization, and four participants were recruited from the stroke outpatient clinic at a hospital in Oslo, Norway. Additional information is provided in the COREQ checklist including sampling strategy, study design, analysis and research team (Online resource 1).

Heath professionals

A total of 16 health professionals participated in 3 focus groups. All health professionals were involved in the clinical care of stroke patients, and recruited from different levels of health care services and from different disciplines. Participants with varied ages, genders, professions and years of clinical experience were recruited to obtain diverse perspectives on PSF.

Data collection

The first author conducted all the interviews and focus groups, and a co-author co-moderated the focus groups, observing and taking notes. Based on the study's aim and previous literature reviews [2, 16], an interview guide was developed to ask: 1) how PSF is described from the perspective of stroke survivors, 2) what it is like to live with PSF, from the perspectives of both stroke survivors (based on personal experience) and health professionals (based on clinical observation), and 3) how PSF is managed by stroke survivors and by health professionals (Online resource 2).

Most interviews and focus groups lasted 60–80 min, while one interview lasted 45 min. All interviews and focus groups were audio-recorded and transcribed verbatim by the first author. Reflection notes were made

immediately after completing each interview/focus group, which aimed to describe contextual information as well as immediate reflections on the data. In addition to interviews, we obtained the following data from the stroke survivors: demographics, stroke characteristics and clinical outcome, modified Rankin Scale (MRS) [28], fatigue (Fatigue Severity Scale [FSS]) [29], depression and anxiety (Hospital Anxiety and Depression Scale [HADS]) [30, 31], cognitive function (Montreal Cognitive Assessment [MOCA]) [32], and health-related quality of life (EuroQol five dimension scale [EQ-5D]) [33, 34]. Focus group participants answered questionnaires about their age, gender, profession, and years of experience with stroke.

Data analysis

Methods for inductive content analysis were applied with the aim to identify, analyse and report themes and categories of stroke survivors' experiences and health professionals' descriptions of PSF [35, 36]. The analysis included reading, open coding, organizing and abstracting codes into sub-categories, categories and themes, and reporting the results. Data collection and analysis occurred simultaneously as an iterative process. The decision to stop data collection was based on a comprehensive evaluation including considerations of the study aim, interview quality and when analysis revealed no new categories in the additional data [35, 37].

Upon completion of each interview, transcripts were imported to NVIVO (v.11), a qualitative data analysis software used to enhance efficiency and transparency in the analytical process [38]. Analysis of the two populations was done separately and was combined for reporting the results. The transcripts were read and re-read several times to get a sense of the whole material. The coding process started after conducting the first three interviews and continued consecutively throughout the process of data collection and analysis. First, the material was coded by its manifest content, i.e. the level of interpretation and abstraction was low at this point. The analytical units in each transcript were given one or more individual codes. All codes were grouped according to their content and formed sub-categories and categories. These categories were first formed separately within each transcript, before further analyses of these categories across the data material in each of the two populations. This method allowed for transparency in the process of finding major and minor categories, patterns between categories, and similarities and differences across the stroke survivors and health professionals. All of the sub-categories were abstracted to categories, which further represented four themes. Examples are presented in Table 1.

Questionnaire data were summarized using descriptive statistics (i.e., frequencies, medians, ranges) to describe the characteristics of participating stroke survivors and health professionals.

Ethical considerations

The study was approved by the Regional Medical and Health Ethics Committee of Southeastern Norway (REK), with reference [reference removed due to blinding]. All participants received written information about the study, gave written consent and were informed about their ability to withdraw from the study at any time before publication of results.

Results

Analysis of the data material revealed four themes that illustrate the experiences and descriptions of PSF: characteristics, interference and aggravating factors, management, and PSF awareness (Fig. 1). Results from the individual interviews with stroke survivors and from the focus groups with health professionals were mostly consistent, and focus group findings are only reported when they contributed additional information. Characteristics of the 9 participating stroke survivors are summarized in Table 2 and characteristics of the 16 participating health professionals are summarized in Table 3.

Characteristics of PSF

In the data material from both stroke survivors and health professionals, PSF was described as a complex and multifactorial phenomenon. For the theme

Table 1 Analysis, examples of themes, categories, sub-categories and codes

Theme	Categories	Sub-categories	Code
Characteristics	Quality	Mental fatigue	Mentally exhausted
			Not physically tired
			Head feels heavy
Interference and aggravating factors	Cognitive performance		Communication difficulties when fatigued
			Concentration difficulties when fatigued
			Fatigued by decision-making
			Prolonged attention induces fatigue

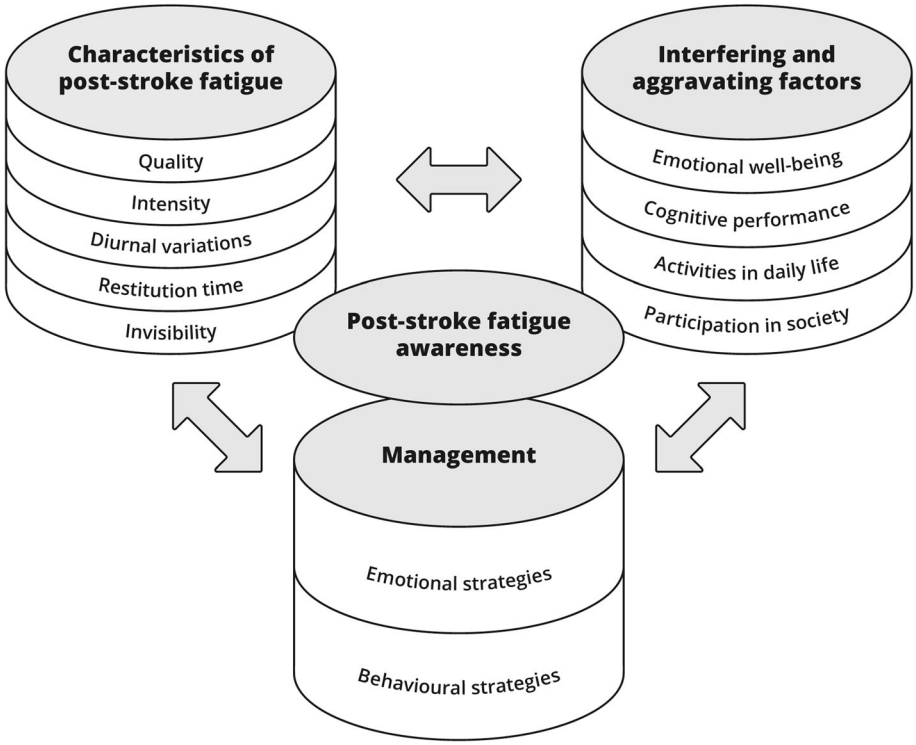


Fig. 1 Conceptual model of post-stroke fatigue, including themes and categories. PSF has different characteristics that affected how PSF interfered with patients' lives. These interfering and aggravating factors had to be managed, and the use of management strategies could again influence the characteristics of PSF and whether or how PSF continued to interfere with their lives. Fatigue was initially interpreted as a normal reaction to having a stroke, and stroke survivors first became aware of PSF sometime after their initial hospital admission

characteristics of PSF five categories were identified: *quality, intensity, diurnal variations, restitution time, and invisibility* (Fig. 1).

Perceptions of PSF varied between the stroke survivors, and different descriptions of fatigue *quality* included three sub-categories: mental fatigue, physical fatigue, and general fatigue. Some described feeling mentally fatigued, whereas for others, fatigue presented itself physically as a bodily sensation. However, most used general terms, such as:

> *"I need to rest my head, I get exhausted in my head, and then I also become tired in my body."* (Participant 4)

> *"I have not done anything other than just sit still […] It is not really tiredness either, it is just a completely different experience, not tired and not sleepy, but a combination of those two, but in a COMPLETELY different way than before the stroke."* (Participant 7)

The *intensity* of fatigue spanned from total exhaustion that prevented the stroke survivors from completing ordinary duties, to manageable fatigue. The stroke survivors also experienced *diurnal variations* of fatigue. Most had days or times during a day without fatigue, and some of them described waking up refreshed, whereas others were fatigued in the morning. The stroke survivors described different patterns of diurnal variations, but a common feature was that the levels of fatigue varied throughout the day. The stroke survivors could also have days or weeks of feeling better or worse. These diurnal and periodic variations of fatigue were a prominent characterization of their experiences.

> *"It [fatigue] is bad in the morning, but then it gets better. It is usually best at mid-day, and then early afternoon it is a dead break […] and in the evening … every evening it is just as if I have used all the energy […] then I am very tired..."* (Participant 7)

When the stroke survivors experienced fatigue, they needed long *restitution time*. The actual recovery time needed varied, but in general, longer restitution time was needed after the stroke. They further described how PSF was *invisible* and that other people had difficulties recognizing their fatigue:

Table 2 Demographics, clinical characteristics and health-related quality of life of the stroke survivors with PSF (individual interview participants)

Stroke Survivor Characteristics	n	Median (range)
Time since stroke, in months		21 (3–24)
3–8 months	4	
20–24 months	5	
Age in years[a]	9	59 (23–80)
Gender		
Male (female)	4 (5)	
Living arrangements		
Living with a partner	6	
Living with children or other family member	2	
Living alone	1	
Residence		
Urban area	5	
Rural area	4	
Education		
Upper secondary education	3	
Higher education < 4 years	3	
Higher education ≥ 4 years	3	
Work status		
Pre-stroke		
Full time work/studies (100%)	5	
Retired	4	
Post-stroke		
Disability leave (100%)	2	
Partial sick leave (50% – 70%)	3	
Retired	4	
ICD-10 Classification[b]		
Non-traumatic intracerebral hemorrhage (I61)	3	
Cerebral infarction (I63)	6	
Degree of disability at stroke onset (mRS)[c]		
Moderate severe disability (mRS 4)	1	
Moderate disability (mRS 3)	1	
Slight disability (mRS 2)	5	
No significant disability (mRS 1)	2	
Communication disorder at stroke onset[d]		
Aphasia (self-reported)	4	
Normal speech	5	
Living situation (at the time of the interview)		
Dependent living (assistance provided)	1	
Independent living (no assistance provided)	8	
Rehabilitation services (at the time of the interview)[e]		
Physiotherapy (weekly)	3	
None	6	
Fatigue (FSS7)[f]		6.4 (4.7–7)

Table 2 Demographics, clinical characteristics and health-related quality of life of the stroke survivors with PSF (individual interview participants) *(Continued)*

Stroke Survivor Characteristics	n	Median (range)
Severe fatigue (FSS ≥ 5)	7	
Moderate fatigue (FSS = 4–4.9)	2	
Depression and anxiety (HADS total score)[g]		20 (9–23)
Likely case of depression and/or anxiety (HADS ≥19)	5	
Possible case of depression and/or anxiety (HADS 15–18)	2	
Normal symptoms of depression and anxiety (HADS < 15)	2	
MoCA[h]		26 (22–29)
Mild cognitive impairment (MoCA 18–25)	4	
No cognitive impairment (MoCA ≥26)	5	
Self-reported health EQ-VAS[i]	9	40 (30–80)

[a] The individual ages of the participants were: *23-54 - 54 - 55 - 59 - 74 - 76 - 79 - 80*
[b] The ICD-10 classification is based on the participants retrospective self-report ($n = 5$) or collected from their medical record ($n = 4$)
[c] *MRS* Modified Rankin Scale measures the degree of stroke impairment or dependence in daily activities. Scores can range from no symptoms (0) to death (6). MRS in this study was rated by IJS, based on retrospective self-report ($n = 5$) or medical records ($n = 4$) [28]
[d] Communication disorder at stroke onset was based on self-report. None of the participants had significant aphasia at the time of the interview
[e] None of the participants received any other rehabilitation services at the time of the interview such as occupational therapy, speech-language therapy etc.
[f] *FSS7* Fatigue Severity Scale 7-item version is scored on a 7-point scale ranging from strongly disagree (1) to strongly agree (7). An individual mean score of ≥5 indicates severe fatigue, score 4–4.9 indicates moderate fatigue, and score < 4 indicates no/mild fatigue [29]
[g] *HADS* Hospital Anxiety and Depression Scale is a screening instrument developed to identify depression and anxiety in medical patients. HADS total score ≥ 19 indicates a case of depression and anxiety, a score between 15 and 18 indicates a possible case, and scores below 15 indicates no symptoms of depression or anxiety [30, 31]
[h] *MoCA* Montreal Cognitive Assessment is a screening instrument developed to detect mild cognitive impairment. Scores over 25 indicate normal cognition and scores between 18 and 25 indicate mild cognitive impairment [32]. In stroke patients, normal scores range 20–27 during the chronic post-stroke phase [39]
[i] EQ-VAS assesses overall health-related quality of life, ranging from *worst possible* (0) to *best possible* (100) health [33]. Mean EQ-VAS score in a Norwegian general population sample is 77.9 (SD = 19.5) [34]

> *"It is my level of energy … the invisible complaints that are a challenge […] I look quite well, but I have been sick and I am still marked by that … "* (Participant 3)

The stroke survivors also had difficulties finding appropriate words to describe PSF to others and to explain how their fatigue was distinct from regular tiredness. This invisibility of PSF constituted the final category of PSF characteristics.

Table 3 Background characteristics of health professionals (focus group participants)

Health Professional Characteristics	Median (range)
Age	48.5 (26–58)
Years of experience with stroke patients	15.5 (3–32)
	n
Gender	
Male (female)	7 (9)
Place of work	
Stroke unit at a local hospital	6
Stroke rehabilitation hospital	5
Community home care	5
Profession	
Physiotherapist	5
Occupational therapist	5
Nurse	3
Speech therapist, physician, clinical psychologist	3

Interfering and aggravating factors

The stroke survivors described how fatigue *interfered* with their lives and how different factors *aggravated* fatigue. Four categories of interfering and aggravating factors were identified: *emotions, cognitive performance, activities in daily life, and participation in society.* Having fatigue interfered with all these aspects of their lives, and in addition, different factors in all four categories could aggravate their fatigue.

PSF also affected the stroke survivors' *emotions*. Having an acute stroke was a frightening experience for the stroke survivors, and the continuous presence of fatigue after the stroke perpetuated their perception of feeling unwell. This contradicted the stroke survivors' understanding of their stroke as a one-time incident from which they had fully or mostly recovered. The stroke survivors also lacked motivation, worried that people would perceive them as lazy and experienced sadness related to how fatigue interfered with their lives:

> *"I have to say that it is quite depressing. Several times, like after that Sunday, I thought: Do you know what? Now you have been wasting a whole day on*

nothing. Nothing, you have not done anything." (Participant 2)

Frustration with being fatigued was also described by the health professionals:

"It takes time [to understand that they are fatigued] and it is a test of patience that is difficult for the patients to accept. They get really frustrated because they are so tired." (Focus group 3)

While having fatigue interferes with their emotions, the stroke survivors also described that experiencing irritation or sadness could also trigger or aggravate fatigue.

Having fatigue also interfered with stroke survivors' *cognitive performance.* During periods of fatigue, the stroke survivors experienced difficulties in communicating, interacting socially, concentrating and maintaining attention. The stroke survivors further described that attempts to concentrate on a task, make decisions, and sustain attention towards a subject could often aggravate the fatigue. Health professionals reported that cognitively demanding rehabilitation activities, such as hand training, often resulted in patients getting excessively fatigued:

"Previously a lot of patients performed physical activities regularly, but now they get tired in a completely different way. Performing upper limb rehabilitation … it is not a lot of repetitions before they get mentally exhausted, because they really need to concentrate and keep focusing." (Focus group 2)

PSF also interfered with the ability to perform regular *activities in daily life*, such as activities outside the home, household chores or sustained activities without rest, and even getting dressed or taking a shower in extreme instances. These activities in daily life were also reported to aggravate fatigue.

PSF interference with activities was further complicated by both the intensity and the diurnal variations of fatigue, leading to difficulties with anticipating their day-to-day capacity. This unpredictability was described as an essential and problematic consequence of having fatigue, and made life with PSF more challenging to manage:

"And also, when you have something that you need to do … you don't know when it [fatigue] will come or if you will get tired from things." (Participant 2)

The stroke survivors described that fatigue in general was an obstacle for them to *participate in society*. Stroke survivors of working age all reported that they worked less, or not at all, due to their fatigue. Being social, engaging in hobbies, and keeping in contact with friends and family were experienced as difficult due to fatigue.

"It don't have the same energy to be with them [my family] […] Before, we used to babysit our grandchildren […] but now, I don't at all have the capacity to do the nice things anymore, and that's too bad … " (Participant 7)

Management

Both stroke survivors and health professionals described a continuous process of trying to find a balance and adapting to life with PSF. The stroke survivors gradually learned to recognize PSF characteristics, how PSF interfered with their life and its aggravating factors. The health professionals in this study described how they observed, advised and supported this process. Two different management strategies were identified: *emotional* and *behavioural strategies*.

Both stroke survivors and health professionals described different *emotional strategies.* Sometime after their stroke, the stroke survivors acknowledged that fatigue was part of their life and accepted that they had to adjust accordingly. Some also tried to ignore fatigue, and carried out activities despite knowing that participating would induce severe fatigue. The health professionals underlined the importance of patients needing to experience on their own how fatigue affected them, and that this was important in accepting the new situation:

"They are used to having a lot of energy and suddenly they don't. And I see that several patients need to go through everything a couple of times, where they get really fatigued, until they start prioritizing, and understand that this is how things need to be." (Focus group 2)

The stroke survivors described different *behavioural strategies* aimed at either preventing or relieving fatigue. Preventive strategies included limiting the activities performed, prioritizing, planning, structuring the days, resting in advance and seeking information.

"It is a delicate balance between activities and rest. If I overdo it … sometimes I feel very good, and then I get on with training, but then mostly the next day and even the day after is ruined." (Participant 1)

The health professionals helped the patients to acknowledge their fatigue, advised them not to spend all their energy at once and helped them to find a balance between activities and rest:

> "*I try to put on the breaks sometimes, because some of them are very eager […] I advise them to take a break, this is a marathon, not a sprint.*" (Focus group 1)

To relieve fatigue, both stroke survivors and health professionals reported different strategies such as withdrawing from a situation, resting, and sleeping. For some stroke survivors, resting involved sitting still and solving a crossword puzzle (physical rest), some needed rest in a dark and quiet room devoid of stimulus (mental rest), whereas others described a combination of both these resting strategies.

PSF awareness

Both stroke survivors and health professionals interpreted fatigue in the early phase as a normal reaction to being acutely ill. Later in the rehabilitation phase, depending on when they recovered from other stroke sequelae, the stroke survivors expected to return to their pre-stroke level of energy. However, when their fatigue did not resolve, their *awareness of PSF* became gradually evident. The health professionals reported that most stroke survivors were exhausted and tired in the early phase, but they did not define it as PSF, as fatigue during this stage was perceived as temporary:

> *"We don't call it fatigue in the acute phase, that's more after a while when we can see how the damage manifests itself, because it is natural to be very tired and exhausted in the beginning."* (Focus group 3)

Both stroke survivors and health professionals described PSF as evident when fatigue started to be in conflict with the patient's and society's expectations of performance, often occurring after their initial hospitalization due to their stroke.

> *"When I came home from rehab … I was supposed to start doing things, inviting people and being social, things that I love to do. Then I did not have energy to do the things I did before. […] we have this tradition inviting a lot of people. I was looking forward to it. But I was SO tired, and I did not understand it. I just sat there crying, I can't do this, we have to call everybody and cancel."* (Participant 5)

Discussion

In this study exploring the PSF experiences of stroke survivors and PSF descriptions by health professionals, four themes were identified: PSF characteristics, interference and aggravating factors, management, and PSF awareness. PSF was described as an experience of mental, physical or general feeling of exhaustion and tiredness, with a discrepancy between the level of activity and the level of fatigue. PSF interfered with, and was aggravated by, emotions, cognitive performance, activities in daily life and participation in society. To manage PSF, both emotional and behavioral strategies were used. It took time before patients were aware of PSF, and it often became evident when fatigue resulted in inability to carry out expected daily activities.

Themes and categories of PSF

The first theme, *characteristics of PSF*, contained five categories important for the characterization of PSF. In line with previous studies [18, 40], *quality* was described as mental, physical, and general fatigue. Likewise, the *diurnal variations* in fatigue intensity found in this study are in agreement with previous studies on PSF [18]. Further, intensity of and diurnal variations in PSF have also been found to be distinct from fatigue in other chronic conditions, such as multiple sclerosis [41]. Despite these prevalent findings, existing fatigue PROMs mostly lack items on quality subtypes and diurnal variations [9]. Interestingly, stroke survivors in this study experienced fatigue despite being in good physical condition, supporting a previous review that found no association between PSF and physical fitness [42]. In contrast, several existing fatigue PROMs include impaired physical condition as an indicator of fatigue. However, these fatigue PROMs are not developed or designed specifically to assess fatigue in a stroke population [9]. This emphasizes the problems of using a generic PROM to assess PSF and suggests that assessment of PSF requires a disease-specific PROM, which is currently not available.

Another theme of PSF in this study was *interfering and aggravating factors*. This included the categories: *emotions, cognitive performance, activities in daily life* and *participation in society*. Previous studies have shown that PSF leads to frustration and emotional disturbances [17], interferes with cognition [43], and also impacts activities at a social, family and community level [17, 19].

The third theme of PSF was *management*, including *emotional* and *behavioral strategies*. Accepting fatigue and adjusting expectations, applying energy-conservation strategies and resting both in advance of and after activity, as well as being physically active and receiving support from others have been previously described as strategies for managing PSF [16, 18, 19, 43]. Although limited evidence exists on the effectiveness of different management strategies, improved assessment and identification of such strategies will enable future studies to compare their effectiveness.

Diagnostic criteria and PSF awareness

A major limitation of existing fatigue PROMs is the lack of clear diagnostic criteria. The stroke survivors in this

study experienced fatigue in the early phase after stroke, similar to observations made by health professionals. Both stroke survivors and health professionals described how PSF interfered with the rehabilitation process. Nevertheless, both patients and health professionals interpreted PSF during this stage as a normal response, and did not necessarily recognize it as a significant problem. This is in agreement with a previous study reporting that fatigue first became evident during hospital admission, but the impact on role loss was not realized until after discharge [43]. In contrast, the case definition for PSF developed by Lynch et al. [8] contains individual criteria for detecting PSF during hospitalization, as the authors acknowledge PSF in the early phase to be important. In addition, a longitudinal observational study found that having PSF in the early phase after a stroke was an independent risk factor for poor physical health 18 months after stroke [3]. These prior studies highlight the importance of assessing and diagnosing PSF in the early phase after stroke.

As fatigue is a common symptom in the general population, not all fatigue experienced after a stroke should necessarily meet the definition of PSF. In individuals with a previous (pre-stroke) history of fatigue, PSF should only be considered when the feeling of fatigue is substantially different in its characteristics, and/or severely increased in intensity. For others, PSF should be considered when the feelings of fatigue are new and persistent after the stroke. However, to accurately distinguish newly developed post-stroke fatigue from pre-existing fatigue through the use of PROMs could be challenging. Including a retrospective item asking about pre-stroke fatigue history could be an important first step to investigating the potential similarities and differences between pre- and post-stroke fatigue.

PSF as a multidimensional phenomenon

This study highlights the complexity and multidimensionality of PSF, which included closely interacting emotional, cognitive, physical and social aspects. When measuring complex constructs such as fatigue, a multidimensional measurement instrument is preferable in order to have a detailed assessment of all relevant dimensions [20]. For example, both symptom intensity and symptom interference measures are considered vital, as stroke survivors might report fatigue as very distressing and significantly interfering with daily life, despite reporting relatively low fatigue intensity, and vice versa. This is in agreement with symptom experience in cancer patients, showing a non-linear relationship between symptom severity and symptom interference [44]. In order to have a more comprehensive assessment of PSF that includes all relevant dimensions, there is a need for a new PSF-specific PROM.

Study strengths and limitations

The study met all relevant COSMIN criteria, which are considered the gold standard for establishing content validity in PROM development [13]. Most of the COREQ criteria are also met, except returning transcripts and participant checking, as well as repeat interviews. We aimed to include a heterogeneous sample of participants in order to explore a broad range of experiences with PSF, but only nine stroke survivors participated in the study. In addition, stroke survivors were asked retrospectively about their fatigue experiences in the early phase and were interviewed up to 24 months post-stroke, introducing possible recall bias. Although the median age of the stroke survivors in this sample was low (59 years) compared to the median age for stroke in Norway (76 years) [45], our sample is too small for quantitative comparisons and the individual ages reflect an age distribution that is representative of the stroke populations in many countries. Another strength of this study was that the perspectives of health professionals working with stroke patients were also included, and results from these focus groups were largely consistent with results from the individual interviews with stroke survivors. Further, the overall aim of this study was to explore PSF to guide future development of a PSF-specific PROM. The results from this study will serve as the basis for item generation in the new PSF PROM. The drafted PROM will then be pilot-tested with cognitive interviews, giving the new participants the opportunity to add, modify and remove items.

Conclusion

This study highlights the complexity and multidimensionality of PSF, which included closely interacting emotional, cognitive, physical and social aspects. Fatigue was interpreted as a normal reaction in the early phase after stroke, and awareness of PSF first emerged when PSF came into conflict with the patient's and society's expectations of performance. Since stroke survivors might not immediately recognize their fatigue, health professionals can help patients to comprehend and adapt to living with fatigue. The results of this study will form the basis for item generation and the development of a comprehensive PSF-specific PROM. Further studies will follow COSMIN-methodology for PROM development, which will include: drafting the PSF PROM, pilot-testing it with cognitive interviews, and field-testing the PROM in a larger sample to explore dimensions and potentially reduce items; further evaluation of the final PROM's measurement properties will then be conducted in a cross-sectional sample [20].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s41687-021-00307-z.

Additional file 1 : **Online resource 1**. Consolidated criteria for reporting qualitative studies (COREQ): 32-item checklist.

Additional file 2 : **Online resource 2**. Research questions and interview guides for individual interviews and focus groups.

Acknowledgements

The authors would like to thank the participants in this study for contributing their time and their experiences to this research. We would like to further acknowledge our research group, the Norwegian Study of Fatigue after Stroke (NORFAST) including user consultant Hilde Magelssen.

Code availability

Not applicable.

Authors' contributions

The following authors contributed to the study: Study design: IJS, MK, CRB, BI, CLG and AL, data collection: IJS, PL and CRB, Data analysis: IJS, MK, PL, CRB, and AL. All authors participated in the interpretation of the findings and developing the manuscript, and have read and approved the final submitted manuscript.

Funding

Ingrid Johansen Skogestad is funded by the Norwegian non-profit *National Association for Public Health* doctoral studentship. The funding source had no involvement in conducting or reporting of this study.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

The study was approved by the Regional Medical and Health Ethics Committee of Southeastern Norway (REK), with reference # 2017/1741. All participants received written information about the study, gave written consent and were informed about their ability to withdraw from the study at any time before publication of results.

Competing interests

The authors declare that there are no conflicts of interest.

Author details

[1]Medical Department, Lovisenberg Diaconal Hospital, Oslo, Norway. [2]Department of Nursing Science, Faculty of Medicine, Institute of Health and Society, University of Oslo, Oslo, Norway. [3]Department of Nursing and Health Promotion, Oslo Metropolitan University, Oslo, Norway. [4]Surgical Department, Lovisenberg Diaconal Hospital, Oslo, Norway. [5]Department for Interdisciplinary Health Sciences, Faculty of Medicine, Institute of Health and Society, University of Oslo, Oslo, Norway. [6]Department of Research and Development, Lovisenberg Diaconal Hospital, Oslo, Norway. [7]Department of Neuromedicine and Movement Science, NTNU, Trondheim, Norway. [8]Department of Stroke, St. Olavs Hospital, Trondheim, Norway. [9]Department of Family Health Care Nursing, University of California San Francisco, San Francisco, USA.

Received: 15 December 2020 Accepted: 1 April 2021
Published online: 21 April 2021

References

1. Skaner, Y., Nilsson, G. H., Sundquist, K., Hassler, E., & Krakau, I. (2007). Self-rated health, symptoms of depression and general symptoms at 3 and 12 months after a first-ever stroke: A municipality-based study in Sweden. *BMC Family Practice*, *8*(1), 61. https://doi.org/10.1186/1471-2296-8-61.
2. Lerdal, A., Bakken, L. N., Kouwenhoven, S. E., Pedersen, G., Kirkevold, M., Finset, A., & Kim, H. S. (2009). Poststroke fatigue--a review. *Journal of Pain and Symptom Management*, *38*(6), 928–949. https://doi.org/10.1016/j.jpainsymman.2009.04.028.
3. Lerdal, A., & Gay, C. L. (2013). Fatigue in the acute phase after first stroke predicts poorer physical health 18 months later. *Neurology*, *81*(18), 1581–1587. https://doi.org/10.1212/WNL.0b013e3182a9f471.
4. Kitzmuller, G., Mangset, M., Evju, A. S., Angel, S., Aadal, L., Martinsen, R., et al. (2019). Finding the way forward: The lived experience of people with stroke after participation in a complex psychosocial intervention. *Qualitative Health Research*, *29*(12), 1711–1724. https://doi.org/10.1177/1049732319833366.
5. Cumming, T. B., Yeo, A. B., Marquez, J., Churilov, L., Annoni, J. M., Badaru, U., … English, C. (2018). Investigating post-stroke fatigue: An individual participant data meta-analysis. *Journal of Psychosomatic Research*, *113*, 107–112. https://doi.org/10.1016/j.jpsychores.2018.08.006.
6. Acciarresi, M., Bogousslavsky, J., & Paciaroni, M. (2014). Post-stroke fatigue: Epidemiology, clinical characteristics and treatment. [Review]. *European Neurology*, *72*(5–6), 255–261. https://doi.org/10.1159/000363763.
7. Nadarajah, M., & Goh, H. T. (2015). Post-stroke fatigue: A review on prevalence, correlates, measurement, and management. *Topics in Stroke Rehabilitation*, *22*(3), 208–220. https://doi.org/10.1179/1074935714Z.0000000015.
8. Lynch, J., Mead, G., Greig, C., Young, A., Lewis, S., & Sharpe, M. (2007). Fatigue after stroke: The development and evaluation of a case definition. *Journal of Psychosomatic Research*, *63*(5), 539–544. https://doi.org/10.1016/j.jpsychores.2007.08.004.
9. Skogestad, I. J., Kirkevold, M., Indredavik, B., Gay, C. L., Lerdal, A., & Group, N (2019). Lack of content overlap and essential dimensions - a review of measures used for post-stroke fatigue. *Journal of Psychosomatic Research*, *124*, 109759. https://doi.org/10.1016/j.jpsychores.2019.109759.
10. Whitehead, L. (2009). The measurement of fatigue in chronic illness: A systematic review of unidimensional and multidimensional fatigue measures. *Journal of Pain and Symptom Management*, *37*(1), 107–128. https://doi.org/10.1016/j.jpainsymman.2007.08.019.
11. Mead, G., Lynch, J., Greig, C., Young, A., Lewis, S., & Sharpe, M. (2007). Evaluation of fatigue scales in stroke patients. *Stroke*, *38*(7), 2090–2095. https://doi.org/10.1161/STROKEAHA.106.478941.
12. Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1-eliciting concepts for a new PRO instrument. *Value in Health*, *14*(8), 967–977. https://doi.org/10.1016/j.jval.2011.06.014.
13. Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., … Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, *27*(5), 1159–1170. https://doi.org/10.1007/s11136-018-1829-0.
14. U. S. Department of Health Human Services (2006). *F. D. A. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance*. (2006/10/13 ed., (vol. 4, p. 79).
15. EMA (2005). *Reflection paper on the regulatory guidance for the use of healthrelated quality of life (HRQL) measures in the evaluation of medicinal products*. European Medicines Agency (https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-regulatory-guidance-use-healthrelated-quality-life-hrql-measures-evaluation_en.pdf).
16. Eilertsen, G., Ormstad, H., & Kirkevold, M. (2013). Experiences of poststroke fatigue: Qualitative meta-synthesis. *Journal of Advanced Nursing*, *69*(3), 514–525. https://doi.org/10.1111/jan.12002.
17. Young, C. A., Mills, R. J., Gibbons, C., & Thornton, E. W. (2013). Poststroke fatigue: The patient perspective. *Topics in Stroke Rehabilitation*, *20*(6), 478–484. https://doi.org/10.1310/tsr2006-478.
18. Barbour, V. L., & Mead, G. E. (2012). Fatigue after stroke: The Patient's perspective. *Stroke Research and Treatment*, *2012*, 863031–863036. https://doi.org/10.1155/2012/863031.
19. Flinn, N. A., & Stube, J. E. (2010). Post-stroke fatigue: Qualitative study of three focus groups. *Occupational Therapy International*, *17*(2), 81–91. https://doi.org/10.1002/oti.286.
20. de Vet, H., Terwee, C., Mokkink, L., & Knol, D. (2011). *Measurement in medicine: A practical guide (practical guides to biostatistics and epidemiology)*. Cambridge University Press. https://doi.org/10.1017/CBO9780511996214.
21. Ricci, L., Lanfranchi, J.-B., Lemetayer, F., Rotonda, C., Guillemin, F., Coste, J., & Spitz, E. (2019). Qualitative methods used to generate questionnaire items:

A systematic review. *Qualitative Health Research*, *29*(1), 149–156. https://doi.org/10.1177/1049732318783186.
22. Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3.
23. Brod, M., Tesler, L. E., & Christensen, T. L. (2009). Qualitative research and content validity: Developing best practices based on science and experience. *Quality of Life Research*, *18*(9), 1263–1278. https://doi.org/10.1007/s11136-009-9540-9.
24. Magasi, S., Ryan, G., Revicki, D., Lenderking, W., Hays, R. D., Brod, M., … Cella, D. (2012). Content validity of patient-reported outcome measures: Perspectives from a PROMIS meeting. *Quality of Life Research*, *21*(5), 739–746. https://doi.org/10.1007/s11136-011-9990-8.
25. Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, *19*(6), 349–357. https://doi.org/10.1093/intqhc/mzm042.
26. WHO (2004). *ICD-10 international statistical classification of diseases and related health problems: Tenth revision*, (2nd ed.,). World Health Organization.
27. Christensen, D., Johnsen, S. P., Watt, T., Harder, I., Kirkevold, M., & Andersen, G. (2008). Dimensions of post-stroke fatigue: A two-year follow-up study. *Cerebrovascular Diseases*, *26*(2), 134–141. https://doi.org/10.1159/000139660.
28. van Swieten, J. C., Koudstaal, P. J., Visser, M. C., Schouten, H. J., & van Gijn, J. (1988). Interobserver agreement for the assessment of handicap in stroke patients. *Stroke*, *19*(5), 604–607. https://doi.org/10.1161/01.str.19.5.604.
29. Lerdal, A., Gay, C., & Lee, K. (2013). Curvilinear relationship between age and post-stroke fatigue among patients in the acute phase following first-ever stroke. *International Journal of Physical Medicine and Rehabilitation*, *1*(141). https://doi.org/10.4172/2329-9096.1000141.
30. Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x.
31. Herrmann, C. (1997). International experiences with the hospital anxiety and depression scale--a review of validation data and clinical results. *Journal of Psychosomatic Research*, *42*(1), 17–41. https://doi.org/10.1016/s0022-3999(96)00216-4.
32. Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x.
33. EuroQol Research Foundation (2019). EQ-5D-5L user guide.
34. Stavem, K., Augestad, L. A., Kristiansen, I. S., & Rand, K. (2018). General population norms for the EQ-5D-3 L in Norway: Comparison of postal and web surveys. *Health and Quality of Life Outcomes*, *16*(1), 204. https://doi.org/10.1186/s12955-018-1029-1.
35. Elo, S., & Kyngas, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x.
36. Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687.
37. Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, *26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444.
38. Houghton, C., Murphy, K., Meehan, B., Thomas, J., Brooker, D., & Casey, D. (2017). From screening to synthesis: Using nvivo to enhance transparency in qualitative evidence synthesis. *Journal of Clinical Nursing*, *26*(5–6), 873–881. https://doi.org/10.1111/jocn.13443.
39. Chiti, G., & Pantoni, L. (2014). Use of Montreal cognitive assessment in patients with stroke. *Stroke*, *45*(10), 3135–3140. https://doi.org/10.1161/STROKEAHA.114.004590.
40. Kirkevold, M., Christensen, D., Andersen, G., Johansen, S. P., & Harder, I. (2012). Fatigue after stroke: Manifestations and strategies. *Disability and Rehabilitation*, *34*(8), 665–670. https://doi.org/10.3109/09638288.2011.615373.
41. Eilertsen, G., Ormstad, H., Kirkevold, M., Mengshoel, A. M., Soderberg, S., & Olsson, M. (2015). Similarities and differences in the experience of fatigue among people living with fibromyalgia, multiple sclerosis, ankylosing spondylitis and stroke. *Journal of Clinical Nursing*, *24*(13–14), 2023–2034. https://doi.org/10.1111/jocn.12774.
42. Duncan, F., Kutlubaev, M. A., Dennis, M. S., Greig, C., & Mead, G. E. (2012). Fatigue after stroke: A systematic review of associations with impaired physical fitness. *International Journal of Stroke*, *7*(2), 157–162. https://doi.org/10.1111/j.1747-4949.2011.00741.x.
43. White, J. H., Gray, K. R., Magin, P., Attia, J., Sturm, J., Carter, G., & Pollack, M. (2012). Exploring the experience of post-stroke fatigue in community dwelling stroke survivors: A prospective qualitative study. *Disability and Rehabilitation*, *34*(16), 1376–1384. https://doi.org/10.3109/09638288.2011.645111.
44. Doorenbos, A., Given, B., Given, C., Verbitsky, N., Cimprich, B., & McCorkle, R. (2005). Reducing symptom limitations: A cognitive behavioral intervention randomized trial. *Psychooncology*, *14*(7), 574–584. https://doi.org/10.1002/pon.874.
45. Fjærtoft, H., Indredavik, B., Mørch, B., Skogseth-Stephani, R., Halle, K. K., & Varmdal, T. (2018). In N. Hjerneslagsregister (Ed.), *Norwegian stroke registry annual report*.

Publisher's Note

III

Development and testing of the Norwegian Fatigue Characteristics and Interference Measure (FCIM) for stroke survivors: cognitive interviews and Rasch analysis

Ingrid Johansen Skogestad[a,b]*, Anders Kottorp[c], Petra Larsson[d,e], Therese Marie Moen[f], Caryl L. Gay[f,g], Christine Råheim Borge[e,f], Anners Lerdal[e,f]

[a] Medical Department, Lovisenberg Diaconal Hospital, Oslo, Norway

[b] Department of Public Health Science, Institute of Health and Society, Faculty of Medicine, University of Oslo, Norway

[c] Faculty of Health and Society, Malmö University, Malmö, Sweden

[d] Surgical Department, Lovisenberg Diaconal Hospital, Oslo, Norway

[e] Department of Interdisciplinary Health Sciences, Institute of Health and Society, Faculty of Medicine, University of Oslo, Norway

[f] Research Department, Lovisenberg Diaconal Hospital, Oslo, Norway

[g] Department of Family Health Care Nursing, University of California, San Francisco, USA

*Corresponding author. injo@lds.no ORCID: 0000-0001-8252-258X

Author contributions

Study conception and design were performed by Ingrid Johansen Skogestad, Anders Kottorp and Anners Lerdal. Data collection was performed by Ingrid Johansen Skogestad and analyses were performed by Ingrid Johansen Skogestad, Anders Kottorp, Petra Larsson, Therese Marie Moen, Christine Råheim Borge and Anners Lerdal. The first draft, and the revisions, of the manuscript was written by Ingrid Johansen Skogestad and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the participating patients and user consultants for their valuable contributions to this study.

Funding

Ingrid Johansen Skogestad is funded by the Norwegian non-profit National Association for Public Health doctoral studentship. The funding source had no involvement in conducting or reporting of this study.

Accessibility of FCIM

The authors retain copyright for the Norwegian Fatigue Characteristics and Interference Measure (FCIM). For any inquiry regarding the availability, use, translation and scoring of FCIM, contact the first author Ingrid Johansen Skogestad; https://orcid.org/0000-0001-8252-258X

Word count: 3997

Introduction

Fatigue is a common and debilitating symptom for stroke survivors, potentially affecting every part of their daily activities and quality of life [1]. Stroke clinicians are currently advised to systematically assess fatigue in all patients [2, 3]. However, there is insufficient evidence to support the use of any routine treatment or prevention strategies [4, 5], and the pathophysiology of post-stroke fatigue (PSF) is largely unknown [1]. A critical barrier in PSF research is the lack of a patient-reported outcome measure (PROM) for fatigue with sound psychometric properties [6, 7].

There is growing recognition that content validity is the most important measurement property of a PROM [8]. Content validity can be defined as "the degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured" [9]. It is advised that content validity should be demonstrated before evaluating other psychometric properties [10, 11]. However, establishing content validity in PROMs that assess unobservable constructs such as fatigue is challenging and requires several steps involving qualitative methods [9]. In a prior study [12], we explored the experience of fatigue in a qualitative study with stroke survivors and health professionals. These findings resulted in the development of a conceptual framework that outlined PSF as a multidimensional phenomenon. Two important dimensions were fatigue characteristics (e.g., intensity, timing) and fatigue interference (i.e. emotional, cognitive, activity and social impacts of fatigue). A clear definition of the construct to be measured is a prerequisite for item development [8]. Despite this, recent fatigue measures lack a clear definition, measure various aspects, and lack high-quality evidence of content validity [6].

In stroke research, the Fatigue Severity Scale (FSS) is the most-used PRO for fatigue. However, the FSS lacks evidence of content validity and evidence of its psychometric properties is limited, particularly among people with stroke [6, 13, 14]. For example, the FSS does not assess important features such as mental versus physical fatigue or diurnal variations [6, 15]. Although other, more complex PROMs for fatigue exist, a recent review found no multidimensional fatigue questionnaires had been adequately validated in people with stroke [16]. With the growing recognition of PSF as a unique clinical entity and the increasing use of PROMs as endpoints in clinical studies, there is a clear need to develop a PSF-specific PROM with robust measurement properties that capture multidimensional features of PSF.

After establishing content validity, guidelines also recommend extensive field testing to obtain insight into the structural validity of the data [9, 11]. Structural validity can be defined as "the degree

to which the scores of a measurement instrument are an adequate reflection of the dimensionality of the construct to be measured" [17]. Rasch analysis is a powerful method to examine item characteristics in detail [9, 18]. However, the fatigue instruments currently used in stroke populations have limited evidence of their content validity. Studies have often moved straight to assessing other types of validity, reliability and responsiveness [13, 16, 19]. This approach could introduce several potential biases in these instruments' scores [8]. To move forward in PSF research, there is need for a fatigue PROM designed for and validated in the stroke population and that has been developed following advanced PROM guidelines [16]. The aim of this study was, therefore, to develop such a measure and evaluate its content validity, structural validity and internal consistency.

Method

Design

This study had a mixed-methods design involving three iterative steps (Fig. 1), as described by de Vet et al. [9]. An expert panel developed the instrument's initial items, cognitive interviews were conducted to evaluate content validity, and an online questionnaire was used to collect data for a Rasch analysis evaluating structural validity and internal consistency. Reporting of the item development and cognitive interviews follows the consolidated criteria for reporting qualitative research [20], and reporting of the Rasch analysis follows the Rasch reporting guideline for rehabilitation research [21].

Item development

We first established an expert panel, using convenience sampling to include a balanced group with diverse backgrounds. The expert panel consisted of three stroke researchers and clinicians from Lovisenberg Diaconal Hospital (LDS), two PROM researchers (one from LDS and one external expert on Rasch analysis) and two stroke patients with fatigue serving as user consultants. One user consultant was already involved in the research project and recruited the second through her network. Both were experienced user consultants in different health projects and active in a user organization for stroke survivors. The first author (IJS), who is a nurse and current PhD-student, led six meetings (four live at LDS and two on Zoom) each lasting 1-4 hours (with breaks). Between meetings, all group members read and commented on written versions of the instrument. The expert panel developed the items and response categories for the fatigue instrument based on our conceptual framework of fatigue derived from a previously published qualitative study. In that study, we defined PSF as "an experience of mental, physical or general feeling of exhaustion and tiredness, with a discrepancy between the level of activity and the level of fatigue" [12]. The expert panel

developed the initial instrument informed by PROM development guidelines [8, 22-25]. The aim was to develop an instrument that could measure characteristics and interference (in two subscales), have the potential to measure improvement (for use in intervention studies), assess the level and duration of fatigue experienced prior to stroke (i.e. pre-stroke fatigue) and be relatively short (feasible). The expert panel developed version 1.0 of the instrument, named the Fatigue Characteristics and Interference Measure (FCIM), through regular meetings from January through June 2020. We also consulted a speech therapist for advice on adapting the instrument for people with aphasia and/or reading difficulties. Finally, to ensure item clarity and test interview technique, we performed two pilot interviews with the user consultants.

Cognitive interviews – developing evidence of content validity

Next, we conducted cognitive interviews with stroke patients. The aim was to establish the content validity of FCIM version 1.0 by evaluating comprehensibility, retrieval, judgement and communication, comprehensiveness and relevance [24, 26]. A stroke user organization in Norway invited members to participate via text messages, e-mail and its Facebook page. Inclusion criteria were prior stroke during the last two years, over 18 years, and living within driving distance from Oslo. 15 stroke survivors were purposively sampled to ensure variability in age, gender, fatigue severity and time since stroke diagnosis. Face-to-face individual interviews were conducted during August and September 2020. Eight interviews were conducted in participants' homes, and seven at LDH in Oslo. Participants completed FCIM version 1.0 as part of the interview, and a questionnaire including a 7-item version of the FSS (FSS7) [14] and information about their sociodemographic characteristics and relevant medical history.

Cognitive interviews were conducted with the Three Step Test Interview (TSTI) technique [27], and followed a semi-structured interview guide (Online Resource 1). TSTI was developed as an aid to identify problems in newly-developed instruments. It consists of the following three steps:

1) *Observing* the response behavior and concurrent *thinking aloud*. The interviewer takes notes and observes behavior such as hesitation and correction of response category. The participants are also instructed to think aloud and verbalize their thoughts when filling out the instrument. The aim is to make the participants' immediate thoughts about the instrument observable for the interviewer.
2) *Follow-up probing* considering the behavior or expressed thoughts collected in step 1, where the aim is to clarify and complete only the primary data previously collected.
3) *Debriefing* aimed at eliciting experiences and opinions, such as potential problems, possible improvements and the instrument's completeness.

Audio-recordings and field notes were taken during the interviews. Recordings were transcribed verbatim and subject to a deductive content analysis facilitated by NVivo 12 [28]. Each item was analyzed separately, and we used a categorization matrix based on Tourangeau's four-stage cognitive model [26], which includes comprehension, retrieval, judgement and response. Finally, we assessed the completeness of the instrument as a whole. After 10 interviews, some changes were made to the instrument, and then again after completing all 15 interviews, resulting in FCIM version 2.0.

Rasch analysis – developing evidence of structural validity and internal consistency

We conducted a cross-sectional study with a convenience sample of stroke patients (N=169). Participants were recruited through the website of a stroke user organization. Inclusion criteria were adults with a self-reported stroke diagnosis who could read Norwegian. Data were collected between January and March 2021. Participants responded to an online questionnaire including sociodemographic information, relevant medical history, FSS7 and FCIM version 2.0. A response to every question was required.

FCIM version 2.0 was analyzed using a two-faceted Rasch model, which calculated the probability of a specified response for both persons and items along the same linear scale (representing the latent trait). This enables transformation of ordinal raw scores into an interval-level variable (called logits). Winsteps (version 5.2.0.0), R (version 4.1.2) and SPSS (version 28.0.0.0) were used to conduct statistical analyses and generate graphs. We applied a Partial Credit Model (PCM) since the FCIM includes items with different response categories. Then we assessed rating scale functioning according to Linacre's guidelines to determine whether the scale was suitable for Rasch analysis [29] (Table 1). We then proceeded to the main aim of the Rasch analysis to evaluate FCIM's structural validity and internal consistency (Table 1).

[Table 1 here]

Rasch analysis transforms ordinal raw scores into an interval-level scale and the fit statistics estimate how well the items' and persons' raw data fit the model assumptions [30]. Fit statistics are presented in infit and outfit unstandardized mean square (MNSQ) and standardized fit statistics (z-values). The MNSQ residuals show the degree of randomness and a MNSQ value of 1.0 indicates perfect fit. Values less than 1.0 indicate overfit to the model (i.e. the observations are too predictable), while values greater than 1.0 indicate underfit (i.e. there is more randomness in the data than expected in the Rasch model). Consistent with earlier empirical studies [14], we evaluated standardized infit statistics as they are more sensitive to unexpected item response patterns targeted to the person. This is in contrast to outfit, which is more sensitive to unexpected observations on very easy or hard

items, with high outfit MNSQ often resulting from a few random responses by low performers [30, 31].

Structural validity. Each subscale's unidimensionality was assessed using Principal Components Analysis (PCA) of the residuals [32]. Although the conceptual framework of fatigue is multidimensional, statistical unidimensionality of each subscale/dimension needs to be ensured. Local dependency between items was evaluated with Yen's Q3 statistics, which compute raw score residual correlations [33]. Local independence means that the variance left after removing the contribution to the latent trait is only random, normally-distributed, noise [32]. If two items are locally dependent, it indicates either that they add to some other dimension, or that they duplicate some feature of each other (called redundancy-dependency) [32].

Internal consistency. Targeting of persons was reported with the mean measure score (theta) for persons and mean standard error. Well-targeted measures have a mean location for persons around 0 (same as for items, which is always set to 0). A positive mean value for persons indicates higher levels of fatigue compared to the average of the scale and a negative value indicates lower levels of fatigue [34]. We also reported Wright item maps that, on the same scale, display both the individual participants' ability measures and the individual items' difficulty calibrations (including the step calibrations [Andrich thresholds]) [32]. Precision was evaluated using the person and item separation index with associated reliability [32]. Consistency in item correlations was assessed with Kuder-Richardson Formula 20 (KR-20), which is comparable to Cronbach's alpha.

Score-to-measure conversion. Rasch analysis converts the raw scores to interval data (logits). To facilitate clinical interpretation of FCIM, we provide score-to-measure tables for both subscales.

The groups of persons with misfitting MNSQ and z-values were compared to the rest of the sample concerning age, gender and fatigue (FSS severe fatigue vs. no/mild/moderate fatigue (combined)), using Student's t-test or Fisher's exact test, as appropriate.

Differential Item Functioning (DIF) analysis was conducted to investigate whether subgroups had significantly different responses to items despite equal levels of the underlying trait [32].

Ethical considerations

The study was conducted following the Declaration of Helsinki and was approved by the Regional Medical and Health Ethics Committee of Southeastern Norway (REK)(reference 2017/1741). All participants provided written informed consent.

Results

Instrument development by expert panel

First, the expert panel generated 160 items related to our conceptual framework [12]. Then, the panel aimed to select the best-worded and most relevant, comprehensive and discriminating items [17]. The FCIM is based on a reflective model and was developed after thorough discussions and according to relevant guidelines [8, 24]. Based on the conceptual framework, the instrument was divided into two subscales: characteristics and interference. This initial version consisted of our definition of PSF and 32 items with a 7-day recall period. The characteristics subscale had 10 items with a 5-point Likert scale ranging from “not at all” to “very much”. Three item pairs (fatigued items 1 and 2, mental fatigue items 3 and 4 and physical fatigue items 5 and 6) were designed to be the same question, but with slightly different wording (Table 2). We included all three item pairs to investigate in the next steps which items were preferred by stroke patients. The interference subscale had 20 items with a 5-point Likert scale ranging from “never” to “all the time”. In addition, we included two pre-stroke fatigue items to help distinguish post-stroke fatigue from pre-existing levels of fatigue. Based on the speech therapist’s input, we used bold font for essential words in each item. This process resulted in FCIM version 1.0.

Content validity testing by cognitive interviews

FCIM version 1.0 was then assessed in 15 cognitive interviews with stroke patients. Characteristics of the participants are presented in Table 3. The interviews lasted between 24 and 92 (mean 53) minutes. Details about the cognitive interviews results are displayed in an item-tracking matrix (Online Resource 2). Preliminary analysis after the first 10 interviews showed difficulties with comprehension of the PSF definition, and minor difficulties with comprehension and judgement in seven items in the interference subscale. We edited the instrument and presented the updated version in the next five interviews, resulting in improved understanding of these items and the PSF definition. After 15 interviews, we changed the ordering of items in the interference subscale and removed two items due to comprehension problems and lack of relevance. We also flagged five items (2, 4, 6, 27 and 28) because of similarity, comprehension and judgement issues. Despite potential issues, we decided to temporarily keep the flagged items and investigate their performance in the Rasch analysis. This resulted in FCIM version 2.0 with 10 characteristics items, 18 interference items and two pre-stroke fatigue items (Table 2). Except for the five flagged items, we found no significant problems relating to comprehension, judgement, relevance or completeness of the items.

[Table 2 here]

[Table 3 here]

Evaluating structural validity and internal consistency with Rasch analysis

FCIM version 2.0 was further evaluated in a sample of 169 patients with stroke who responded to an online questionnaire (Table 3). There were no missing data. First, we evaluated the functioning of both subscales against Linacre's guidelines (Table 1) [29], and both subscales fulfilled the criteria of having unimodal distribution with peaks in the center, more than 10 observations in each category, outfit MNSQ <2.0, and step calibrations (Andrich thresholds) that advanced monotonically between 1 and 5 logits.

Characteristics subscale. The first iteration with all 10 items revealed misfit in item 10, so we removed this item (Table 4). The second iteration displayed overfit in item 2, which was also removed. In the third iteration, all eight items demonstrated acceptable fit to the Rasch model. Then we assessed dimensionality of the remaining eight items by a PCA of the residuals. The variance explained by the latent trait was just above 55%, however the eigenvalue in the 1st contrast was slightly elevated, justifying further investigation. Residual correlations between items 3 and 4, as well as items 5 and 6, were above the critical value. This was expected since these item pairs were almost identical and previously flagged from the cognitive interviews. We removed items 4 and 6 and re-ran the analysis. Removing these locally-dependent items resulted in improved overall fit to the Rasch model (Table 4, Online Resource 3). The remaining six items also demonstrated evidence of local independence, with no positive correlations. The mean person response was about 1 logit higher than the mean item measure. The Wright map (Fig. 2) shows that the subscale might work best for low and moderate levels of fatigue. In addition, the 6-item subscale demonstrated acceptable KR-20 and person separation (Table 4). Slightly exceeding our criterion, 10 persons (5.9%) demonstrated misfit to the Rasch model. However, no significant differences in the group of misfits compared to the rest of the sample were found in relation to age, gender or fatigue. No DIF was detected in relation to gender, but significant DIF was found for item 8 in relation to age, with the age group 61-83 being more likely to agree with this item than the age group 45-60 (p=0.0064). The final 6-item characteristics subscale demonstrated evidence of good structural validity and internal consistency. Characteristics subscale raw scores range from six to 30 and correspond to Rasch person measures of -5.96 to 6.37 logits. A score-to-measure table is provided in Online Resource 4.

Interference subscale. Item goodness-of-fit statistics from the first iteration of the 18-item interference subscale indicated misfit in items 15 and 16 (Table 4). After removal of these two items, the remaining 16 items demonstrated acceptable fit with infit MNSQ values that met our specified

criteria. PCA of the residuals showed that 62% of the variation was explained by the latent trait, however, as the eigenvalue in the 1st contrast was greater than 2, and some of the residual item correlations were above 0.3 (indicating locally-dependent items), further investigation was justified. Items 14 and 18 both demonstrated a higher-than-expected residual correlation with two other items (Table 4), thus we decided to remove items 14 and 18. We additionally removed items 27 and 28, since they were flagged as redundant items from a content validity perspective in the cognitive interviews. In the final subscale, the mean person measure was 0.55 logits higher than the mean item measure, and the Wright map (Fig. 3) shows that the subscale works well across all levels of fatigue. The final interference subscale also had a high person separation index and reliability. Fourteen persons (8.28%) demonstrated misfit to the Rasch model on the interference subscale, which exceeded our criterion (Table 4). However, no statistically significant differences were found in the group of misfitting persons compared to the rest of the sample in relation to age, gender or fatigue. No significant DIF was found in relation to gender or age. Interference subscale raw scores range from 12 to 60 and correspond to Rasch person measures of -7.54 to 8.09 logits. A score-to-measure table is provided in Online Resource 5. In sum, the final 12-item interference subscale demonstrated good fit to the Rasch model and good structural validity and internal consistency (Table 4, Online Resource 6).

Pearson correlation indicated a strong positive relationship ($r=0.77$, $p<.001$) between the 6-item characteristic subscale and 12-item interference subscale.

[Table 4 here]

Discussion

In this study, we developed and evaluated the Norwegian Fatigue Characteristics and Interference Measure (FCIM), a new 20-item PROM for PSF.

Several previous studies have concluded that fatigue is a multidimensional phenomenon [12, 35]. In this study, based on our conceptual framework, the fatigue dimensions of characteristics and interference were separated into two subscales. Although they were highly correlated, these two dimensions were confirmed by the cognitive interviews and the principal components analysis, suggesting that characteristics and interference constitute dimensions that need to be assessed separately to avoid missing important features of patients' PSF experience. In clinical settings, differentiating these subscales may also support targeting of different types of interventions. For example, a specific intervention might have no effect on fatigue's intensity (measured by items in the characteristics dimension), but could alter fatigue's interference with the person's activities. Hence,

PROMs that fail to capture the multidimensionality of PSF may lack sufficient sensitivity and specificity for use in clinical intervention trials.

A high-quality PROM needs to be feasible in addition to having evidence of validity and reliability. To avoid respondents losing concentration or becoming fatigued, it is recommended that the instrument not be too extensive or time-consuming [9]. Thus, we aimed for a relatively small number of items in the final FCIM. Three items were removed due to high infit MNSQ values indicating that answers to these items were more unexpected than predicted by the Rasch model. This might indicate that these items capture an additional construct [32]. For example, a possible explanation of misfit in item 10 is that evening fatigue can be commonly experienced even by people with low levels of overall fatigue, as it might reflect normal circadian rhythms of lower energy levels later in the day [36]. Keeping this item could possibly bias the results.

In our previous study, we found that stroke survivors qualitatively described fatigue in a variety of ways [12], a finding also reflected in existing fatigue PROMs using a wide range of different expressions to capture fatigue [6]. Even slight differences in item wording are known to affect responses [37]. To identify the best fatigue wording, we included three item pairs with slightly different wording in the characteristics subscale. Based on the cognitive interviews, we flagged these items and retained them for the Rasch analysis. Not surprisingly, Rasch analysis indicated local dependence or overfit to the model in these items. It seems likely that the residual correlations between these items indicated duplication (redundancy-dependency) rather than multidimensionality [33]. In retrospect, we could have removed these items after the cognitive interviews, however, it is unlikely that keeping them changed our overall results.

Another limitation of our study was the loss of some separation ability due to the subscales' item reduction, although the instrument's structural validity and internal consistency increased, as did its feasibility. Person misfit was slightly higher than expected and should be monitored more closely in larger studies, as a larger sample (>200) would offer more powerful data on person and context [22]. In addition, we used a convenience sampling method, and despite our sample's diverse sociodemographic and clinical characteristics, a random-sampling method could yield more robust results.

An advantage of our study is that FCIM is developed based on qualitative data. Our previous review of PROMs used in PSF research showed that existing instruments included items confounded by other post-stroke sequela [6], such as "Do you feel weak?" Including such items could bias the results in a stroke population. Fatigue characteristics and interference are common in other diagnoses, and

FCIM has the potential to be used across different patient groups after assessment of its measurement properties.

Conclusion

In this study, we have developed the FCIM, a patient-reported outcome measure for post-stroke fatigue that includes two subscales measuring fatigue's characteristics and interference. The FCIM also includes two pre-stroke fatigue items. This study has shown that the FCIM comprehensively captures the essential experiences of fatigue, and thus has high content validity. Using Rasch analysis, we removed misfitting and locally-dependent items, which resulted in two unidimensional subscales demonstrating structural validity and internal consistency. Further assessment of the FCIM is necessary before it can be used in clinical studies. The most important next step is to investigate the FCIM's ability to detect change over time (i.e. responsiveness), and its relationship to other instruments (construct validity). Responsiveness is especially important for determining whether the FCIM can be used as an outcome measure for intervention studies. Future studies should also evaluate the FCIM's measurement properties in other patient populations, as fatigue severity and fatigue interference are common outcomes of many diseases/disorders. While FCIM is currently only available in Norwegian, we aim to translate and cross-culturally validate the instrument in an English-speaking sample.

Table 1 Overview of measurement properties and criteria assessed for and through Rasch analysis

Measurement properties	Aim	Criteria
Rating scale functioning	Assess item responses and their fit to the Rasch model assumptions	Assessment of distribution At least 10 observations of each category Categories advance monotonically Outfit MNSQ <2.0 Step calibrations (Andrich Thresholds) advance between 1 and 5 logits
Item goodness-of-fit	Assess individual items' fit to the Rasch model	Infit MNSQ between 0.7-1.3 [38]
Structural validity	Principal components analysis (PCA) of the residuals to assess unidimensionality and local independence	Criteria for unidimensionality: raw variance explained by measure >50% of the total variance. Unexplained variance in residual contrasts <2 eigenvalues [32] Yen's Q3 correlations between residuals <0.3 [33, 39]
Internal consistency	Targeting of persons with mean and S.E of θ (theta)	Report direction and distance from mean item measure at 0 [39]
Internal consistency	Assess person separation index and reliability Assess item separation index and reliability	Person separation >2 Person reliability ≥0.8 Item separation >3 Item reliability >0.9 [32]
Internal consistency	Assess consistency in correlations between items	KR-20 ≥0.70 for each unidimensional scale or subscale [39]
Person goodness-of-fit	Detect improbable item-score patterns	Infit MNSQ ≤1.4 and/or associated z-value <2 [14]
Differential item function (DIF)	Assessed across gender and three age groups (24-44, 45-60, 61-83)	DIF was analyzed using Mantel Chi-Square test for polytomous data with a Bonferroni-adjusted p-value = 0.01 [14, 40]

Table 2 Version 2.0 of the Norwegian Fatigue Characteristics and Interference Measure (FCIM)

Version 2.0 of the Norwegian Fatigue Characteristics and Interference Measure (FCIM)*

In this questionnaire, we want to assess post-stroke fatigue. It is very normal to feel tired during periods when you have a lot to do, but being fatigued means that you are more tired than you would expect considering what you have done.

Please choose the answer that best describes how you have been feeling in the past 7 days.

To what degree did you feel: Not at all – A little bit – Somewhat – Quite a bit – Almost always

1. Fatigued
2. *Exhausted*
3. Mentally fatigued
4. *Tired in your head*
5. Physically fatigued
6. *Tired in your body*
7. Fatigued in the morning
8. Fatigued around noon
9. Fatigued in the afternoon
10. *Fatigued in the evening*

Challenges due to fatigue Never – Rarely – Sometimes – Often – All the time

11. How often were you so fatigued that you had problems **concentrating**?
12. How often were you so fatigued that you had problems **making decisions**?
13. How often were you so fatigued that you had problems **following a conversation**?
14. *How often were you so fatigued that you had problems **gathering your thoughts**?*
15. *How often were you so fatigued that you had problems taking a **bath or shower**?*
16. *How often were you so fatigued that you had problems getting **dressed/undressed**?*
17. How often did you have problems **starting** your tasks/activities because of fatigue?
18. *How often did you have problems **completing** your tasks/activities because of fatigue?*
19. How often did you have to **give up on** your tasks/activities because of fatigue?
20. How often have tasks/activities **taken more time** because of fatigue?
21. How often did you avoid **activities outside your home** because of fatigue?
22. How often has it been difficult to **plan activities** ahead of time because of fatigue?
23. How often have you felt fatigued even if you have **not done anything?**
24. How often did you avoid **physical activity** because of fatigue?
25. How often did you limit your **social activities** because of fatigue?
26. How often were you too fatigued to be **together with your family**?
27. *How often did you avoid **engaging in hobbies or leisure activities** because of fatigue?*
28. *How often did you avoid **pleasant activities** because of fatigue?*

Pre-stroke fatigue

29. Before you had a stroke, to what degree did you feel fatigued then?
    Not at all – A little bit – Somewhat – Quite a bit – Almost always
30. Before you had a stroke, for how long had you felt fatigued? (If item 29 is greater than 'Not at all')
    1 month or less – 2-6 months – 7-12 months – more than a year

*The FCIM was developed in Norwegian and was translated above by the first author in the interest of this publication only. At the time of publication, the measure has not yet been translated into English according to the current standards for translation of measures for research purposes. This version should not be considered the final version and should not be used.

Note: The 10 items in italics (2, 4, 6, 10, 14, 15, 16, 18, 27 and 28) were evaluated in FCIM version 2.0, but were not included the final 20-item version 3.0.

Table 3 Patient characteristics for the cognitive interview and Rasch analysis samples

Patient characteristics	Cognitive interviews (n=15)	Rasch analysis (n=169)
Age, mean (range)	55.5 (40-75)*	52.4 (24-83)
24-44	2	35 (20.7)
45-60	7	98 (58.0)
61-83	4	36 (21.3)
Male, n (%)	7 (46.7)	66 (39.1)
Education, n (%)		
Primary school	2 (13.3)	11 (6.5)
Secondary school	5 (33.3)	62 (36.7)
Higher education <4 years	4 (26.6)	61 (36.1)
Higher education ≥ 4 years	4 (26.6)	35 (20.7)
Marital status, n (%)		
Married	11 (73.3)	95 (56.2)
Unmarried	3 (20.0)	38 (22.5)
Widowed	0	4 (2.4)
Divorced/separated	1 (6.6)	32 (18.9)
Years since stroke, n (%)		
1-24 months	13 (86.6)	69 (41)
>2 years	2 (13.3)	100 (59)
Type of stroke, n (%)		
Cerebral infarction	13 (86.6)	126 (75)
Hemorrhage	2 (13.3)	34 (20)
Unknown/other	0	9 (5)
Work status, n (%)		
Working (full or part time)	5 (33.3)	60 (35.5)
Stroke-related speech disorder, n (%)	Not collected	55 (32.5)
Weekly rehabilitation with speech therapist, n (%)	1 (6.7)	12 (7.1)
Fatigue Severity Scale (FSS7) total score**, n (%)		
No/mild fatigue (1-3.9)	3 (20.0)	21 (12.4)
Moderate fatigue (4-4.9)	5 (33.3)	23 (13.6)
Severe fatigue (5-7)	7 (46.7)	125 (74.0)

All data are self-reported.

*Age is missing for two cognitive interview participants.

**The FSS total score is calculated as the mean of all item scores, can range from 1 to 7 and can be divided into the categories above [41].

Table 4 Overview of Rasch analysis results at each step in the iterative item removal process

Psychometric property	Characteristics subscale		Interference subscale		
Item removal step	Step #1	Step #2	Step #1	Step #2	Step #3
Item goodness-of-fit statistics (Infit MNSQ) 1st iteration 2nd iteration	 Item 10 (1.38) Item 2 (0.56)		 Item 16 (1.41) Item 15 (1.48)		
Removed item (reason)	Item 10 (misfit) Item 2 (misfit and flagged)	Item 4 (flagged and locally dependent) Item 6 (flagged and locally dependent)	Item 16 (misfit) Item 15 (misfit)	Item 14 (locally dependent) Item 18 (locally dependent)	Item 27 (flagged) Item 28 (flagged)
Items left for further analysis	8	6	16	14	12
Principal components analysis (PCA) of the residuals: Latent trait % 1st contrast (eigenvalue)	 55.4% 2.3	 57.6% 1.5	 62.0% 2.5	 62.9% 2.3	 62.8% 2.0
Local independence (Yen's Q3 residual correlations between items)	Items 5 and 6 (0.55) Items 3 and 4 (0.37)	No positive correlations	Items 12 and 14 (0.37) Items 13 and 14 (0.37) Items 18 and 19 (0.33) Items 17 and 18 (0.31) Items 27 and 28 (0.25) Items 19 and 20 (0.24) Items 25 and 28 (0.24)	Items 11 and 12 (0.25) Items 19 and 20 (0.24) Items 17 and 23 (0.23)	Items 19 and 20 (0.20) Items 11 and 12 (0.20) Items 17 and 23 (0.20) Items 21 and 25 (0.20)
Targeting of persons* Mean person measure (mean model standard error)	1.08 (0.58)	0.95 (0.67)	0.52 (0.42)	0.57 (0.45)	0.55 (0.48)
KR-20 (internal consistency)	0.89	0.86	0.95	0.95	0.94
Person separation index (person reliability)*	2.45 (0.86)	2.21 (0.83)	4.11 (0.94)	3.87 (0.94)	3.52 (0.93)
Item separation index (item reliability)	4.11 (0.94)	3.63 (0.93)	5.51 (0.97)	5.87 (0.97)	6.28 (0.98)
Person misfit, n (%)	11 (6.5)	10 (5.9)	17 (10)	17 (10)	17 (10)

*With extreme and non-extreme persons

Fig. 1 The three iterative steps in the development of the Norwegian FCIM.

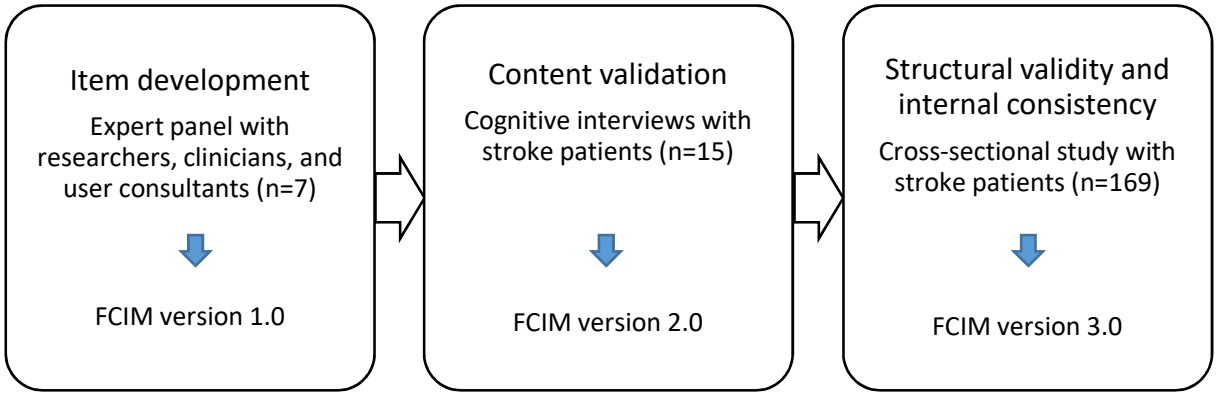


Fig. 2 Wright map displaying the 6-item characteristics subscale. The left side displays individual participants' ability measures (based on their mean logits), presented both as a total sample, and separately for females and males. The right side displays the individual items' difficulty calibrations, including the difficulty of each step calibration (Andrich thresholds). Both the participants' ability measures and the individual item difficulty calibrations are spaced along the common vertical axis with the logits presented on the right side [32].

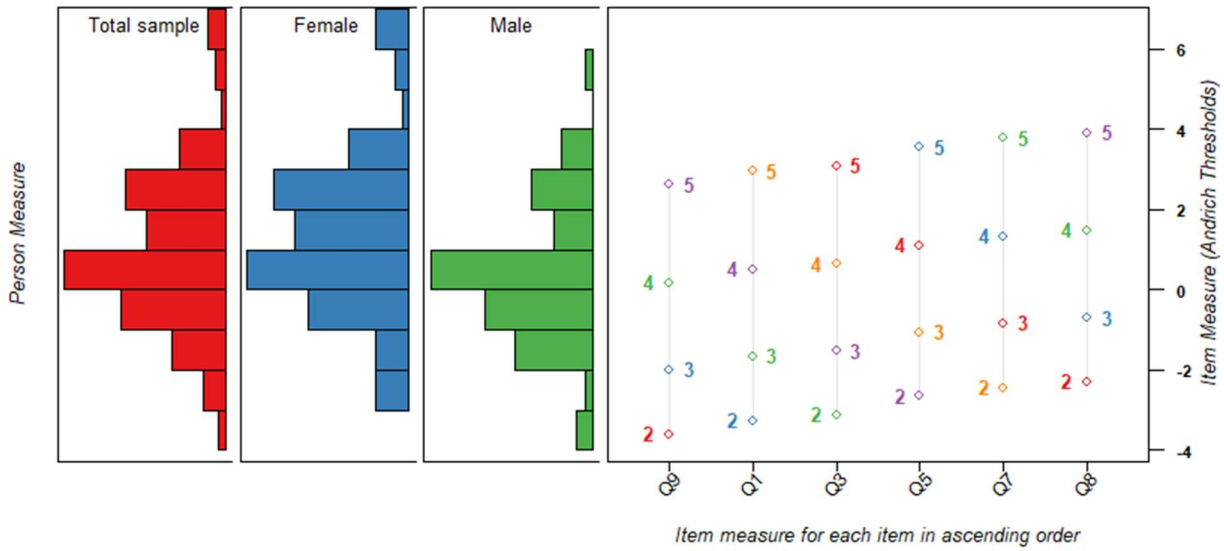
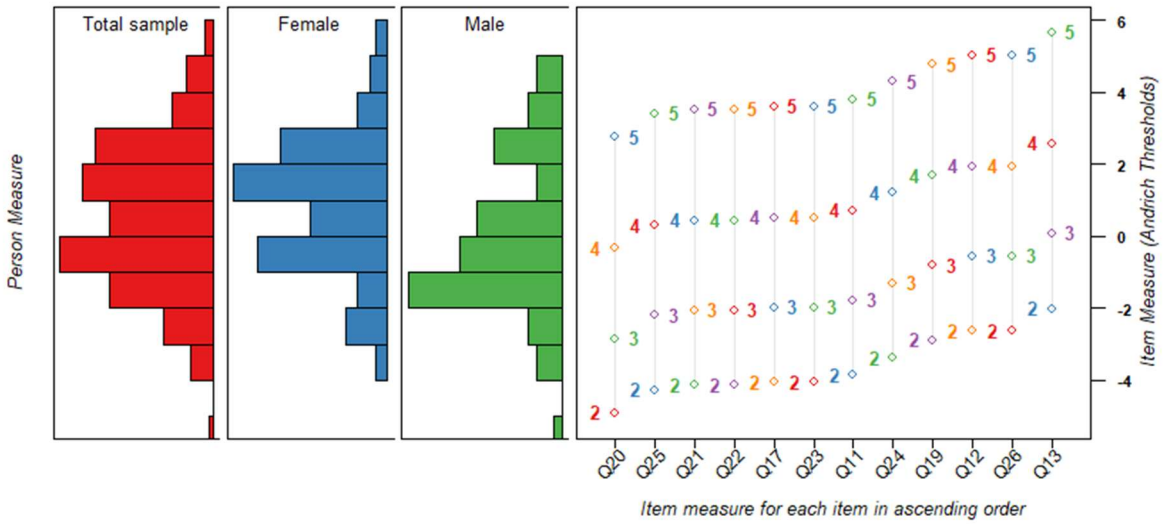


Fig. 3 Wright map displaying the 12-item interference subscale. The left side display individual participants ability measures (based on their mean logits), presented both as a total sample, and separately for females and males. The right side displays the individual items' difficulty calibrations, including the difficulty of each step calibration (Andrich thresholds). Both the participants' ability measures and the individual item difficulty calibrations are spaced along the common vertical axis with the logits presented on the right side [32].



References

1. Paciaroni M, Acciarresi M. Poststroke Fatigue. *Stroke.* 2019;50(7):1927-33. doi: 10.1161/STROKEAHA.119.023552.
2. Lanctot KL, Lindsay MP, Smith EE, Sahlas DJ, Foley N, Gubitz G, Austin M, Ball K, Bhogal S, Blake T, Herrmann N, Hogan D, Khan A, Longman S, King A, Leonard C, Shoniker T, Taylor T, Teed M, et al. Canadian Stroke Best Practice Recommendations: Mood, Cognition and Fatigue following Stroke, 6th edition update 2019. *International Journal of Stroke.* 2020;15(6):668-88. doi: 10.1177/1747493019847334.
3. Norwegian Directorate of Health. The Norwegian guideline for stroke treatment and rehabilitation [in Norwegian] 2017. https://www.helsedirektoratet.no/retningslinjer/hjerneslag.
4. Wu S, Kutlubaev MA, Chun HY, Cowey E, Pollock A, Macleod MR, Dennis M, Keane E, Sharpe M, Mead GE. Interventions for post-stroke fatigue. *Cochrane Database of Systematic Reviews.* 2015(7):CD007030. doi: 10.1002/14651858.CD007030.pub3.
5. Kennedy C, Kidd L. Interventions for post-stroke fatigue: A Cochrane review summary. *International Journal of Nursing Studies.* 2018;85:136-7. doi: 10.1016/j.ijnurstu.2017.11.006.
6. Skogestad IJ, Kirkevold M, Indredavik B, Gay CL, Lerdal A, Group N. Lack of content overlap and essential dimensions - A review of measures used for post-stroke fatigue. *Journal of Psychosomatic Research.* 2019;124:109759. doi: 10.1016/j.jpsychores.2019.109759.
7. Cumming TB, Yeo AB, Marquez J, Churilov L, Annoni J-M, Badaru U, Ghotbi N, Harbison J, Kwakkel G, Lerdal A, Mills R, Naess H, Nyland H, Schmid A, Tang WK, Tseng B, van de Port I, Mead G, English C. Investigating post-stroke fatigue: An individual participant data meta-analysis. *Journal of Psychosomatic Research.* 2018;113:107-112. doi: 10.1016/j.jpsychores.2018.08.006.
8. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Bouter LM, de Vet HCW, Mokkink LB. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research.* 2018;27(5):1159-70. doi: 10.1007/s11136-018-1829-0.
9. de Vet H, Terwee, C., Mokkink, L., & Knol, D. Measurement in Medicine: A Practical Guide (Practical Guides to Biostatistics and Epidemiology). Cambridge: Cambridge University Press; 2011.
10. U. S. Department of Health Human Services, F. D. A. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health and Quality of Life Outcomes. 2006/10/13 ed2006. p. 79.
11. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research.* 2018;27(5):1147-57. doi: 10.1007/s11136-018-1798-3.
12. Skogestad IJ, Kirkevold M, Larsson P, Borge CR, Indredavik B, Gay CL, Lerdal A. Post-stroke fatigue: an exploratory study with patients and health professionals to develop a patient-reported outcome measure. *Journal of Patient Reported Outcomes.* 2021;5(1):35. doi: 10.1186/s41687-021-00307-z.
13. Nadarajah M, Goh HT. Post-stroke fatigue: a review on prevalence, correlates, measurement, and management. *Topics in Stroke Rehabilitation.* 2015;22(3):208-20. doi: 10.1179/1074935714Z.0000000015.
14. Lerdal A, Kottorp A. Psychometric properties of the Fatigue Severity Scale-Rasch analyses of individual responses in a Norwegian stroke cohort. *International Journal of Nursing Studies.* 2011;48(10):1258-65. doi: 10.1016/j.ijnurstu.2011.02.019.
15. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology.* 1989;46(10):1121-3.

16. Elbers RG, Rietberg MB, van Wegen EE, Verhoef J, Kramer SF, Terwee CB, Kwakkel G. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Quality of Life Research.* 2012;21(6):925-44. doi: 10.1007/s11136-011-0009-2.
17. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63(7):737-45. doi: 10.1016/j.jclinepi.2010.02.006.
18. Andrich D, Marais I. A course in Rasch measurement theory. *D Andrich y I Marais (Coords), Measuring in the Educational, Social and Health Sciences.* 2019:41-53.
19. Mead G, Lynch J, Greig C, Young A, Lewis S, Sharpe M. Evaluation of fatigue scales in stroke patients. *Stroke.* 2007;38(7):2090-5. doi: 10.1161/STROKEAHA.106.478941.
20. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care.* 2007;19(6):349-57. doi: 10.1093/intqhc/mzm042.
21. Mallinson T, Kozlowski AJ, Johnston MV, Weaver J, Terhorst L, Grampurohit N, Juengst S, Ehrlich-Jones L, Heinemann AW, Melvin J, Sood P, Van de Winckel A. Rasch Reporting Guideline for Rehabilitation Research (RULER): the RULER Statement. *Arch Phys Med Rehabil.* 2022;103(7):1477-86. doi: 10.1016/j.apmr.2022.03.013.
22. Mokkink LB, Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., & Terwee, C. B. . COSMIN Study Design checklist for Patient-reported outcome measurement instruments. Amsterdam, The Netherlands. 2019.
23. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1--eliciting concepts for a new PRO instrument. *Value in Health.* 2011;14(8):967-77. doi: 10.1016/j.jval.2011.06.014.
24. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. *Value in Health.* 2011;14(8):978-88. doi: 10.1016/j.jval.2011.06.013.
25. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research.* 2021;30(8):2197-218. doi: 10.1007/s11136-021-02822-4.
26. Tourangeau R. Cognitive Science and Survey Methods. In: T. B. Jabine MLS, J. M. Tanur, & R. Tourangeau, editor. Cognitive Aspects of Survey design: Building a Bridge between Disciplines. Washington, DC: National Academy Press.1984. p. 73-100.
27. Hak T, Veer KvD, Jansen H. The Three-Step Test-Interview (TSTI): An observation-based method for pretesting self-completion questionnaires. *Survey Research Methods.* 2008;2(3):143-50. doi: 10.18148/srm/2008.v2i3.1669.
28. Elo S, Kyngas H. The qualitative content analysis process. *Journal of Advanced Nursing.* 2008;62(1):107-15. doi: 10.1111/j.1365-2648.2007.04569.x.
29. Linacre JM. Optimizing rating scale category effectiveness. *Journal of Applied Measurement.* 2002;3(1):85-106.
30. Bond TG, Fox. CM. Applying the Rasch model: Fundamental measurement in the human sciences. . New Jersey: Psychology Press; 2013.
31. Wright BD, Masters GN. Rating scale analysis. MESA press; 1982.
32. Linacre JM. A User's Guide to Winsteps Ministeps Rasch-Model Computer Programs. Program Manual 5222022.

33. Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement.* 2017;41(3):178-94. doi: 10.1177/0146621616677520.
34. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis and Rheumatology.* 2007;57(8):1358-62. doi: 10.1002/art.23108.
35. Lerdal A, Bakken LN, Kouwenhoven SE, Pedersen G, Kirkevold M, Finset A, Kim HS. Poststroke fatigue--a review. *Journal of Pain and Symptom Management.* 2009;38(6):928-49. doi: 10.1016/j.jpainsymman.2009.04.028.
36. Valdez P. Circadian Rhythms in Attention. *Yale J Biol Med.* 2019;92(1):81-92.
37. Willis GB. Cognitive interviewing: A tool for improving questionnaire design. sage publications; 2004.
38. Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology.* 2008;8(1):1-11.
39. Mokkink LB, Prinsen CA, Patrick D, Alonso J, Bouter LM, De Vet HC, Terwee CB. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) User Manual. 2018.
40. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological).* 1995;57(1):289-300.
41. Lerdal A, Bakken LN, Rasmussen EF, Beiermann C, Ryen S, Pynten S, Drefvelin AS, Dahl AM, Rognstad G, Finset A, Lee KA, Kim HS. Physical impairment, depressive symptoms and pre-stroke fatigue are related to fatigue in the acute phase after stroke. *Disability and Rehabilitation.* 2011;33(4):334-42. doi: 10.3109/09638288.2010.490867.

Online resource 1

Cognitive Interview Guide
Introduction: We have developed a new instrument to assess fatigue in stroke patients. The aim of this interview is to test our new instrument and find out whether you find the items difficult to answer, if you understand what we are asking, if we lack response options etc. We are not so interested in what your actual answer is, but rather how you experience the process of answering. This interview will have three steps. First, I want you to answer the instrument, and concurrently think aloud while you are answering. In this step, I will only observe and cannot answer any questions that you might have. In the next step, I will follow up on the observations I made in the first step, and lastly, we will talk more freely about the instrument and your experiences.
Step 1 – Observation Observe the respondent's behavior and gather "think aloud" data. Look specifically for: -Questions that are being skipped -Correction of the chosen response category -Hesitation -Stress/insecurity
Step 2 – Follow up probing Ask questions about the observations made in step 1. For example: -You stopped by this question, what did you think then? -Did I hear you saying…?
Step 3 – De-briefing about experiences and opinions Ask questions about the respondent's experiences of and opinions about specific items and the instrument as a whole. For example: -How did you understand the introduction to the instrument? -How did you comprehend this term in this question? -Can you tell me why you answered alternative x on this question? -Do you have any suggestions to how this could have been formulated differently? -You had problems answering question x, can you tell me about how your situation is related to this question? -How do you understand the answer alternatives in this question? -Were any of the questions excessive/irrelevant? -Were there any questions or themes that you thought were missing? -What did you think of the instrument as a whole?

Online Resource 2

Item-tracking matrix of FSCIM Version 1.0					
Introduction		**Analytical summary**	**Example quotes**	**Changes made (after 10 or 15 interviews)**	**Results**
Post-stroke fatigue definition, long version		Problems with reading and understanding the long definition		Simplified the definition after 10 interviews	No problems in the last 5 interviews
Severity subscale					
Item	Short item name	Analytical summary	Example quotes	Changes made (after 10 or 15 interviews)	Results
1	Fatigued				
2	Exhausted	Several interpreted this item as the feeling you get after a hard physical exercise out or intensive work	*"Exhausted, that is what you are after a long day at work." Participant 10*		**Flagged item (removed after Rasch)**
3	Mentally fatigued				
4	Tired in your head	This was comprehended as being almost the same as item 3, but as a slightly wider concept, e.g. also including head pain and worrying for the future	*"I feel that tired in your head, that is similar to mentally fatigue […] but tired in your head can also be headache, or that your feelings are gathering up…" Participant 8*		**Flagged item (removed after Rasch)**
5	Physically fatigued				
6	Tired in your body	This was comprehended as being almost the same as item 5, but as also including body pain and normal tiredness after exercise.	"Being tired in your body… that could also be after physical exercise and that's a good feeling." Participant 5 "I think that is almost the same, physically fatigued and tired in your body." Participant 15		**Flagged item (removed after Rasch)**

7	Morning fatigue				
8	Fatigued around noon				
9	Afternoon fatigue				
10	Evening fatigue				**(removed after Rasch)**
Interference subscale					
11	Concentrating	Initial wording: "How often did you have problems concentrating due to fatigue?" This wording were comprehended as two questions in one: one about concentration and the other about fatigue.	*"I have problems with my concentration, but that has not anything to do with my fatigue… other than that it gets worse when I am fatigued." Participant 3*	Changed after 10 interviews to: "How often were you so fatigued that you had problems concentrating?	No problems in the last 5 interviews.
12	Making decisions	Initial wording: "How often did you have problems making decisions due to fatigue?" This wording was comprehended as two questions in one: one about making decisions and the other about fatigue.		Changed after 10 interviews to: "How often were you so fatigued that you had problems making decisions.	No problems in the last 5 interviews.
13	Following a conversation	Initial wording: "How often did you have problems following a conversation due to fatigue?" This wording was comprehended as two questions in one: one about following a conversation and the other about fatigue.	*"If I am out at social gatherings, then I fall out [of the conversation], because it gets too much different voices and sounds…" Participant 1*	Changed after 10 interviews to: "How often were you so fatigued that you had problems following a conversation.	No problems in the last 5 interviews.
14	Gathering thoughts	A few commented that this item was very similar to item 11, concentrating. Initial wording: "How often did you have problems gathering your thoughts due to fatigue?" This wording was comprehended as two questions in one: one about gathering thoughts and the other about fatigue.		Changed after 10 interviews to: "How often were you so fatigued that you had problems gathering your thoughts.	**No problems in the last 5 interviews. (Item 14 removed after Rasch)**

15	Bath or shower				**(removed after Rasch)**
16	Dress/undress				**(removed after Rasch)**
17	Starting activities	We detected problems with item 17-20. These items were interpreted differently than our intention. We changed individual wording of these items as described below. We also changed the ordering of items, and put these four items together in a more logical sequence. After completing these changes, we did not experience any problems in the last 5 interviews. Some respondents had problems with comprehension and judgement because of the overall wording of the item. Our aim was to investigate if they had problems **starting** things in general, but the respondents rather emphasized problems with daily tasks. Initial wording: "How often were you unable to get started with daily tasks due to fatigue?"	*"Daily tasks… what do you mean by daily tasks?" Respondent 4*	Changed the wording after 10 interviews to: "How often did you have problems starting your tasks/activities because of fatigue?"	No problems in the last 5 interviews.
18	Completing activities	We detected problems with comprehension of this item. It was interpreted as the focus was on household chores, and many did not have any specific chores. Our aim was to investigate if they had problems **completing** things in general, and our initial wording was not reflecting this aim. Initial wording: "How often did you have problems completing your tasks at home due to fatigue"	*"Several things I leave to my husband now, such as vacuuming." Respondent 6*	Changed the wording after 10 interviews to: "How often did you have problems completing your tasks/activities because of fatigue?"	No problems in the last 5 interviews. **removed after Rasch)**
19	Giving up on activities	We detected problems with comprehension, because of the overall wording of the item. Initial wording: "How often did you have to stop an activity due to fatigue?"		Changed the wording after 10 interviews to: "How often did you have to give up on	No problems in the last 5 interviews.

				your tasks/activities because of fatigue?"	
20	Taken more time	Judgement problems. Some choose to take more time as a coping mechanism. Initial wording: "How often have tasks taken longer time because of fatigue?"		Changed the wording after 10 interviews to: "How often have tasks/activities taken more time because of fatigue?"	No problems in the last 5 interviews.
21	Activities outside home				
22	Planning				
23	Not done anything				
24	Physical activity				
25	Social activity				
26	Family				
27	Hobbies	Comprehension and retrieval problems. Some respondents had no hobbies. Overlaps with other items. We decided to flag this item and investigate how it performed in the Rasch analysis.	*"Hobbies? I think this item is very similar to social activities. Is it common for adults to have hobbies other than physical exercise? Participant 10* *"This is difficult to answer since I don't have any hobbies. But I ended up answering never" Participant 11*		**Flagged item (removed after Rasch)**
28	Pleasant activities	Comprehension and judgement problems. Overlaps with other items. We decided to keep this item for further Rasch analysis.	*"…social activities, there I responded often, and then pleasant activities… These items are overlapping I think…" Participant 5*		**Flagged item (removed after Rasch)**

29	Rehabilitation	Initial wording was: "If you have rehabilitation therapy, how often did you avoid planned sessions due to fatigue?" Participants reported problems with comprehension, retrieval and judgement. Several did not understand what we meant with "rehabilitation therapy" and we changed the wording after 10 interviews. These changes made the item similar to item 24 and was then comprehended as general physical activity. Some participants were not engaged with rehabilitation activities. Considering these overall problems, we decided to remove this item.	*"I don't know what you mean by rehabilitation activities? Do you mean physical activity?" Participant 6.*	"If you are doing retraining after your stroke, how often were you so fatigued that you avoided exercises?	**Removed after 15 interviews due to problems and lack of relevance**
30	Paid work	Wording "If you are working, how often were you less effective at work due to fatigue?" This item had the same response alternatives as the other questions and some respondents answered this question even if they were on full disability leave. If included, this item should have a "not relevant" category since it was not relevant to the respondents who did not work. We decided that we only wanted to include items that were relevant for all participants, and that item 21 (activities outside home) could partly enable participants to report interference with work in that item.	*"If you are at work… I am not at work… What shall I answer on this item since I am not working?" Participant 13*		**Removed after 15 interviews due to lack of relevance**
31	Pre-stroke fatigue				
32	Pre-stroke fatigue	Problems with two overlapping response categories. Initial wording was "1 month" and "1-6 months". This was changed to "1 month or less" and "2-6 months"		Changed response categories after 10 interviews.	No problems in the last 5 interviews.
Overall judgement of completeness		Changed the order of almost all items in the interference subscale after 10 interviews. The pre-stroke items were moved from the start to the end of the instrument.			

Notes: Blank cells indicate that no problems were identified and/or no changes were made. Flagged items had problems identified in the cognitive interviews but were retained for further evaluation in Rasch analysis. All items flagged in the cognitive interviews were subsequently removed based on the Rasch analysis.

Online resource 3

The 6-item characteristics subscale showing item calibrations and fit statistics reported in hierarchical order, difficulty from most to least

Items	Measure (logits)	Std. error	Infit MnSq	Infit zstd	Outfit MnSq	Outfit zstd
8	0.60	0.12	0.88	-1.13	0.88	-1.08
7	0.46	0.12	1.15	1.32	1.16	1.44
5	0.24	0.12	1.20	1.73	1.22	1.89
3	-0.23	0.12	1.11	0.99	1.08	0.76
1	-0.36	0.12	0.71	-2.93	0.70	-3.07
9	-0.70	0.12	0.90	-0.95	0.93	-0.62

*Std., standard; MnSq, mean square; zstd, Z standard.

Online resource 4

Raw score to Rasch logits conversion table of the 6-item characteristics subscale of FCIM

FCIM characteristics raw score	FCIM characteristics Rasch logits
6	-5.96
7	-4.67
8	-3.85
9	-3.31
10	-2.88
11	-2.50
12	-2.16
13	-1.83
14	-1.51
15	-1.18
16	-0.86
17	-0.52
18	-0.17
19	0.19
20	0.55
21	0.93
22	1.32
23	1.72
24	2.14
25	2.58
26	3.04
27	3.55
28	4.16
29	5.04
30	6.37

Online resource 5

Raw score to Rasch logits conversion table of the 12-item interference subscale of FCIM

FCIM interference raw score	FCIM interference Rasch logits
12	-7.54
13	-6.27
14	-5.48
15	-4.97
16	-4.58
17	-4.26
18	-3.97
19	-3.71
20	-3.47
21	-3.24
22	-3.02
23	-2.81
24	-2.61
25	-2.40
26	-2.21
27	-2.01
28	-1.81
29	-1.62
30	-1.42
31	-1.22
32	-1.03
33	-0.82
34	-0.62
35	-0.42
36	-0.21
37	0.00
38	0.21
39	0.43
40	0.65
41	0.87
42	1.10
43	1.34
44	1.57
45	1.82
46	2.07
47	2.32
48	2.58
49	2.84
50	3.11
51	3.38
52	3.66

53	3.96
54	4.27
55	4.60
56	4.97
57	5.40
58	5.95
59	6.79
60	8.09

Online resource 6

The 12-item interference subscale showing item calibrations and fit statistics reported in hierarchical order, difficulty from most to least

Items	Measure (logits)	Std. error	Infit MnSq	Infit zstd	Outfit MnSq	Outfit zstd
13	1.58	0.12	1.06	0.57	1.09	0.79
26	0.95	0.12	0.91	-0.82	0.90	-0.90
12	0.95	0.12	1.11	1.02	1.14	1.30
19	0.69	0.12	1.03	0.35	1.03	0.36
24	0.20	0.13	1.11	1.05	1.12	1.09
11	-0.28	0.13	0.84	-1.50	0.82	-1.70
23	-0.48	0.13	0.98	-0.18	0.99	-0.03
17	-0.48	0.13	0.93	-0.64	0.90	-0.91
22	-0.56	0.13	1.18	1.57	1.17	1.52
21	-0.56	0.13	1.02	0.24	1.00	0.04
25	-0.69	0.13	0.98	-0.15	0.94	-0.52
20	-1.33	0.13	0.85	-1.40	0.82	-1.61

*Std., standard; MnSq, mean square; zstd, Z standard.

Appendix 1

Spørsmål om hjerneslaget og sosiodemografiske opplysninger

1. Hvor mange måneder er det siden hjerneslaget?

_____

2. Hva slags type hjerneslag hadde du?

☐ Hjerneblødning
☐ Hjerneinfarkt
☐ Vet ikke
☐ Annet

3a. Ble det tatt bildeundersøkelser av hodet ditt (CT eller MR)?

☐ Ja
☐ Nei
☐ Vet ikke

3b. Hvis ja, hva var resultatet av CT/MR?

☐ Hjerneblødning/hjerneinfarkt
☐ Det var normalt
☐ Vet ikke

4a. Hadde du ansiktsskjevhet eller krafttap i armer eller ben?

☐ Ja
☐ Nei
☐ Vet ikke

4b. Dersom ja, hvilken side av kroppen ble rammet?

☐ Høyre
☐ Venstre

4c. Har du hatt hjerneslag eller TIA tidligere?

☐ Ja
☐ Nei
☐ Vet ikke

5. Hadde du problemer med å snakke eller finne ord? (Som ikke var tilstede før hjerneslaget)

☐ Ja
☐ Nei
☐ Vet ikke/husker ikke

6. Hvor gammel er du?
Skriv antall år: ____

7. Kjønn
☐ Kvinne
☐ Mann

8a. Var du i lønnet arbeid før du fikk hjerneslag?
☐ Ja
☐ Nei

8b. Hvis ja, arbeidet du på fulltid eller deltid?
☐ Fulltid
☐ Deltid

8c. Er du i lønnet arbeid nå?
☐ Ja
☐ Nei

8d. Hvis ja, har du vært helt eller delvis sykemeldt de siste 14 dagene?
☐ Ja
☐ Nei

8e. Hvis ja, er du sykemeldt på grunn av problemer etter hjerneslaget?
☐ Ja
☐ Nei

9. Får du behandling for depresjon?
☐ Ja
☐ Nei

10. Hvor mange ganger har du mottatt rehabiliterings- og/eller omsorgstjenester i løpet av de siste 2 ukene?
☐ Fysioterapi, antall:____
☐ Hjemmesykepleie, antall:____
☐ Ergoterapi, antall:____
☐ Logopedtjenester, antall:____
☐ Psykolog/psykiater, antall:____
☐ Hjemmehjelp, antall:____
☐ Annet

11. Har du familie eller andre nærstående som du kan støtte deg på?
(sett ett eller flere kryss)

☐ Ingen
☐ Ektefelle/samboer
☐ Egne barn
☐ Foreldre
☐ Andre nærstående

12. Hvilken utdanning er den høyeste du har fullført?

☐ Grunnskole 7-10 år, framhaldsskole folkehøgskole
☐ Real- eller middelskole, yrkesskole, ett- eller toårig videregående skole
☐ Artium, økonomisk gymnas eller allmennfaglig retning i videregående skole
☐ Høgskole eller universitet, mindre enn 4 år
☐ Høgskole eller universitet, 4 år eller mer

13. Hva er din sivilstand?

☐ Gift/registrert partner
☐ Ugift
☐ Enke/enkemann
☐ Skilt
☐ Separert

14. Hvem bor du sammen med?
(sett ett eller flere kryss)

☐ Ektefelle samboer
☐ Barn/svigerbarn
☐ Bor alene
☐ Søster/bror
☐ Annen familie/slekt
☐ Bor på institusjon
☐ Andre