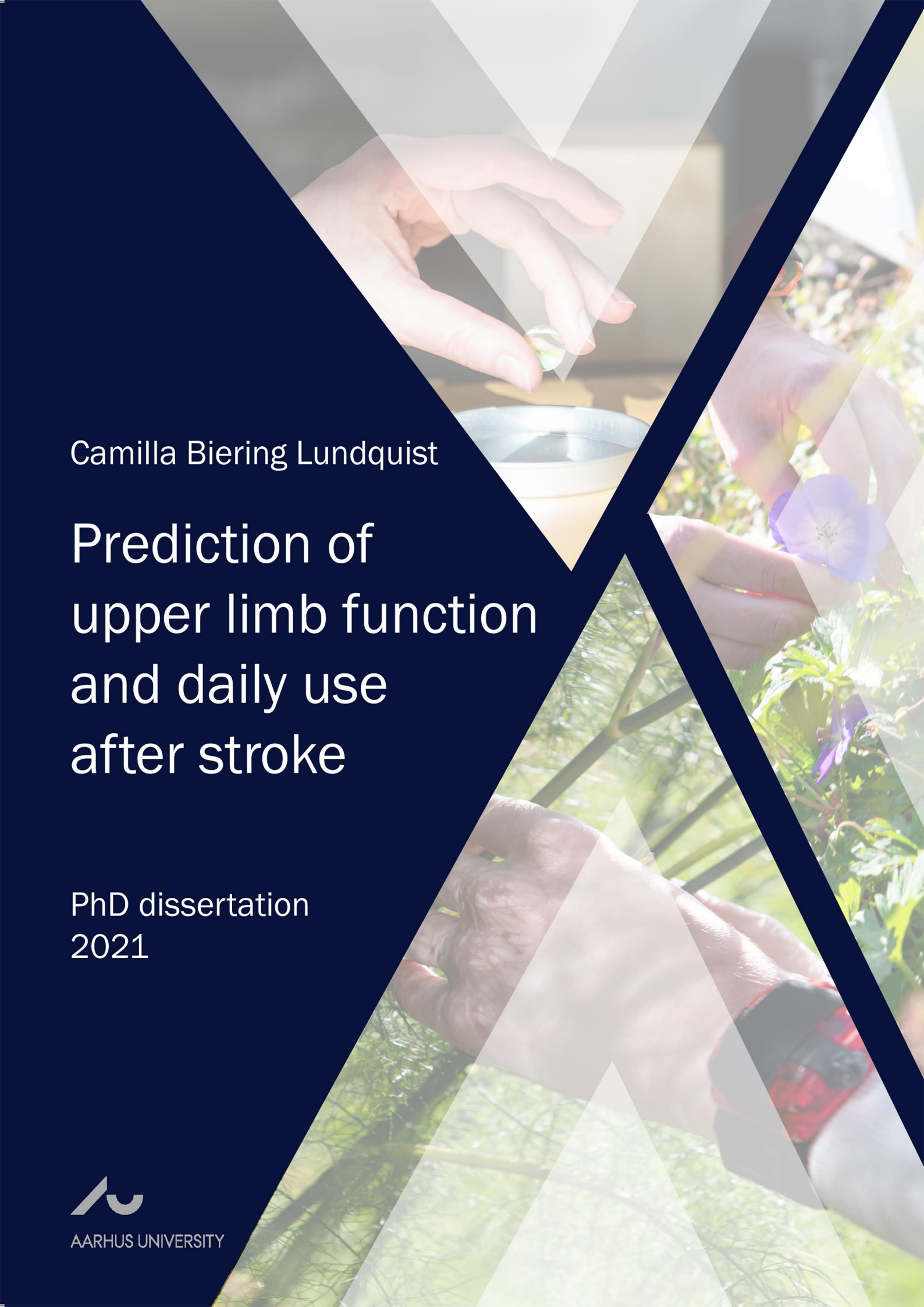


Camilla Biering Lundquist

Prediction of upper limb function and daily use after stroke

PhD dissertation
2021

AARHUS UNIVERSITY

Prediction of upper limb function and daily use after stroke

PhD thesis

Camilla Biering Lundquist

Graduate school of Health
Aarhus University

Hammel Neurorehabilitation Centre
and University Research Clinic

2021

Supervisors

Principal supervisor:
Iris Charlotte Brunner, Associate professor, PT, PhD. Department of Clinical Medicine, Hammel Neurorehabilitation Centre and University Research Clinic, Aarhus University, Aarhus, Denmark

Co-supervisors:
Jørgen Feldbæk Nielsen, Professor, MD, DMSc. Department of Clinical Medicine, Hammel Neurorehabilitation Centre and University Research Clinic, Aarhus University, Aarhus, Denmark

Tine Tjørnhøj-Thomsen, Professor, MSc. National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark

Evaluation committee

Inger Mechlenburg, Professor, PT, MSc, DM (chairperson and moderator of the defence). Department of Orthopaedic Surgery, Aarhus University Hospital. Aarhus, Denmark

Geert Verheyden, Professor, PT. Department of Rehabilitation Sciences KU Leuven, Belgium

Thomas Platz, Professor, MD. Department for Neurorehabilitation, BDH-Klinik Greifswald, Germany

Preface

This PhD project originates from an interest in upper limb impairment after stroke and a desire to use UL prediction models in clinical practice. When a patient asks: "Will I ever be able to use my arm and hand again?" or "when can I hold a fork while eating?" those questions could be answered with more certainty in the future. Knowledge of upper limb prognosis can be used for the benefit of the patient when setting goals or choosing UL interventions.

The clinical use of upper limb prediction models has been a topic of focus at Hammel Neurorehabilitation Centre since 2017, when a group of physiotherapists and occupational therapists employed within research or professional development examined and discussed the evidence and potential implementation of UL prediction models. Based on these discussions, the most relevant model for clinical use at individual level, appeared to be the Predict Recovery Potential (PREP2) algorithm. The main reason was, that compared to other prediction models, the predictive accuracy of PREP2 for patients with severe upper limb impairment was high.

However, several organizational obstacles prevented an implementation of PREP2 at a local level. The first part of PREP2, the Shoulder Abduction Finger Extension (SAFE) test, is designed to be performed within the first 72 hours after stroke, while patients are frequently admitted to RHN at a later point. Due to the limited time window to obtain the prediction, implementation of PREP2 was not feasible.

In light of the challenges with implementation of PREP2 in the clinical settting, this PhD project was initialized. Its aim was to investigate the accuracy of PREP2 when obtained at a later point in time than originally proposed. If accuracy would still be high, this would pave the way for an easier incorporation of the algorithm in clinical practice. Whereas the PREP2 predicts upper limb function, real life daily

use of arm and hand are often more relevant to patients and therapists. To be truly meaningful, improvements in upper limb function must translate into improved use of the arm and hand in daily life. Thus, also the prediction of daily use of the arm and hand was examined. Finally, as the success of a future implementation will to a large extend depend on the health care professionals, a qualitative study was conducted to explore therapists´ perceptions of facilitators and barriers for a future implementation. Though the results were not always as expected, conducting these three studies was an exciting process.

Camilla Biering Lundquist, February 20th 2021

This thesis is based on the following papers:

I. C.B. Lundquist, J.F. Nielsen, F.G. Arguissain, I. Brunner
Accuracy of the upper limb prediction algorithm PREP2 applied 2 weeks poststroke. A prospective longitudinal study.
Published in: Neurorehabilitation and Neural Repair 1-11 2020

II. C.B. Lundquist, J.F. Nielsen, I. Brunner
Prediction of upper limb use three months after stroke. A prospective longitudinal study.
Submitted to: Disability and Rehabilitation

III. C.B. Lundquist, H. Pallesen, T. Tjørnhøj-Thomsen, I. Brunner
Exploring physiotherapists' and occupational therapists' perceptions of the upper limb prediction algorithm PREP2 after stroke in a rehabilitation setting. A qualitative study.
Submitted to: BMJ Open.

Contents

List of figures

List of tables

List of abbreviations

ARAT	Action Research Arm Test
CART	Classification and Regression Tree
CCR	Correct Classification Rate
CFIR	Consolidated Framework for advancing Implementation Research
CI	Confidence Interval
FMA	Fugl-Meyer Motor Assessment Upper Extremity
FIM	Functional Independence Measure
IQR	Inter Quartile Range
MEP	Motor-evoked Potentials
NIHSS	National Institute of Health Stroke Scale
OT	Occupational Therapist
PI	Prediction Interval
PREP2	Predict Recovery Potential algorithm, version 2
PT	Physiotherapist
RHN	Hammel Neurorehabilitation Centre and University Research Clinic
SAFE	Shoulder Abduction Finger Extension
SSS	Scandinavian Stroke Scale
SD	Standard Deviation
TMS	Transcranial Magnetic Stimulation
Twopd	Two-point discrimination
UL	Upper limb

Definitions

Algorithm: A set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem.[1]

Biomarker: A stroke recovery biomarker can be defined as "an indicator of disease state that can be used as a measure of underlying molecular/cellular processes that may be difficult to measure directly in humans."[2] A Motor-evoked Potential (MEP) is an example of a biomarker, used in the present PhD project.

Neglect: Unilateral visuospatial neglect can be defined as "the inability to detect, respond to, and orient toward novel and significant stimuli occurring in the hemispace contralateral to a brain lesion."[3]

Prediction: A statement about what you think will happen in the future[4]

Use ratio: The use ratio is measured with wrist-worn accelerometers and defined as the total hours of paretic UL use divided by total hours of non-paretic use. A use ratio of 0.5 indicates that the paretic UL is active 50% of the time the non-paretic UL is active.

English summary

Background: Prediction of UL function and daily use is relevant for targeted rehabilitation of patients with stroke. In this PhD project a prospective, observational longitudinal study was conducted to examine prediction of UL function (Study I) and prediction of UL use (Study II). A qualitative study was conducted to explore physiotherapists' and occupational therapists' perceptions of upper limb prediction models (Study III).

Study I: The aim was to examine the prognostic accuracy of an existing UL algorithm, the Predict Recovery Potential algorithm (PREP2), when the time window to obtain the prediction was expanded to two weeks after stroke.
Methods: Patients were assessed in accordance with the PREP2 approach. However, two main components, the shoulder abduction finger extension (SAFE) score and motor-evoked potentials (MEPs) were obtained two weeks after stroke. UL function at 3 months was predicted in one of four categories and compared to the actual outcome at three months, as assessed by the Action Research Arm Test. The prediction accuracy of the PREP2 was quantified using the correct classification rate (CCR).
Results: A total of 91 patients were included. Overall CCR of the PREP2 was 60% (95% CI 50-71%). Within the four categories, CCR ranged from the lowest value at 33% (95% CI 4-85%) for the category *Limited* to the highest value at 78% (95% CI 43 -95%) for the category *Poor*. In the present study, the overall CCR was significantly lower than the 75% accuracy reported by the PREP2 developers.

Study II: The primary aim was to examine if UL impairment after stroke could predict UL use in daily life. The secondary aim was to identify additional predictors of

UL use and characteristics of patients who did not achieve normal UL use.
Methods: UL impairment was assessed with Fugl-Meyer Motor Assessment Upper Extremity (FMA) two weeks after stroke. UL use was assessed three months after stroke with wrist-worn accelerometers, and expressed as a use ratio. The use ratio is the total hours of paretic UL use divided by total hours of non-paretic use.
The predictive value of FMA for UL use ratio, was assessed in a linear regression model. In addition, the association was adjusted for secondary variables. Use ratio was dichotomized into normal and non-normal, and non-normal use was assessed by logistic regression.
Results: Eighty-seven patients were included. FMA score predicted 38% of the variance in UL use ratio and an adjusted regression model predicted 55%. The statistically significant predictors were FMA, MEP status and neglect. The 95% prediction intervals of the regression lines were wide. Non-normal use could be predicted with a high accuracy based on MEP- and/or neglect. For the remaining patients, with MEP and without neglect, non-normal use could be predicted at a sensitivity of 0.80 and a specificity of 0.83.

Study III: The aim was to explore how physiotherapists (PTs) and occupational therapists (OTs) perceive UL prediction models.
Methods: Four focus group interviews with 3-6 PTs and OTs were conducted. Data was analysed using a thematic content analysis. Meaning units were identified and subthemes formed. Information gained from all interviews was synthesized.
Results: Four main themes emerged: *Current Practice; Perceived Benefits; Barriers;* and *Preconditions for Implementation*. The participants knew of UL prediction algorithms, but few had a profound knowledge. PREP2 was considered a potentially helpful tool when planning treatment and setting goals. Main barriers were concern about prediction accuracy and potential dilemmas of confronting the patients with a negative prognosis. Preconditions for implementation included tailoring the implementation to a specific unit, sufficient time for acquiring new skills, and a supporting organization.

Conclusion: In Study I, the PREP2 obtained two weeks post stroke was unsuited for clinical implementation. However, PREP2 showed potential to predict either *excellent UL function* in already well-recovered patients or *poor UL function* in patients with persistent severe UL impairment who were MEP-.
In Study II, UL function at baseline predicted increased UL use in daily life. Individual predictions were difficult due to large outcome variations. However, non-normal UL use could be predicted reliably based on the absence of MEPs and/or presence of neglect.
In study III, experienced neurological therapists were sceptical towards prediction algorithms due to the lack of precision of the algorithms and concerns about ethical dilemmas. However, the PREP2 algorithm was regarded as potentially useful.

Danish summary

Baggrund: Prædiktion af armfunktion og prædiktion af daglig brug af arm og hånd kan anvendes til at målrette rehabiliteringen af patienter med følger efter apopleksi. I denne ph.d. afhandling undersøges prædiktion af armfunktion (Studie I) og prædiktion af daglig armbrug (Studie II) i et prospektivt, longitudinelt studie. Fysioterapeuters og ergoterapeuters holdninger til armprædiktionsmodeller undersøges i et kvalitativt studie (Studie III).

Studie I: Formålet var at undersøge præcisionen af en eksisterende algoritme for prædiktion af arm og håndfunktion, når denne anvendes på et senere tidspunkt i patientforløbet end oprindeligt tiltænkt.
Metode: Inkluderede patienter blev undersøgt to uger efter deres apopleksi. Patienterne fik i overensstemmelse med Predict Recovery Potential (PREP2) algorithmen prædikteret deres kommende armfunktion tre måneder efter apopleksi i en af fire kategorier, der hver svarede til et interval af scores på Action Research Arm Test. Præcisionen af algoritmen blev udregnet ved correct classification rate (CCR), hvor de prædikterede kategorier for armfunktion blev sammenholdt med den reelt opnåede armfunktion.
Resultater: I alt 91 patienter blev inkluderet. Overodnet set var CCR af PREP2 60% (95% CI 50-71%). Inden for de fire kategorier spændte CCR fra en laveste værdi på 33% (95% CI 4-85%) for kategorien *Begrænset Armfunktion* til en højeste værdi på 78% (95% CI 43 -95%) for kategorien *Ringe Armfunktion*. Præcisionen af algoritmen i studie I var statistisk significant lavere end de 75%, der blev fundet i den oprindelige population, hvor algoritmen blev anvendt få dage efter apopleksiens opståen.

Studie II: Hovedformålet var at undersøge, om armfunktion to uger efter apopleksi kunne prædiktere daglig brug af arm og hånd tre måneder efter apopleksi. Derudover at identificere yderligere prædiktorer for daglig brug af arm og hånd samt at karakterisere patienter, som ikke opnåede normal brug af arm og hånd.
Metode: Armfunktion blev undersøgt med Fugl-Meyer undersøgelse af armfunktion (FMA) to uger efter apopleksi. Daglig brug af arm og hånd blev målt med accelerometre på begge håndled tre måneder efter apopleksi og angivet som en use ratio. Use ratio angiver antal timer med aktivitet i den afficerede arm i forhold til antal timer med aktivitet i den ikke-afficerede arm. Den prædiktive værdi af FMA for use ratio blev undersøgt med linear regression. Efterfølgende blev associationen justeret for sekundære variabler. Use ratio blev dichitomiseret i normal og ikke-normal og ikke-normal brug blev undersøgt med logistisk regression.
Resultater: I alt 87 patienter blev inkluderet. FMA prædikterede 38% af variation i use ratio og en justeret model prædikterede 55%. De statistisk signifikante prædiktorer var FMA, MEP status og neglekt. 95% prædiktionsintervallet for regressionslinjerne var brede. Ikke-normal brug af arm og hånd kunne prædikteres med høj præcision ud fra fravær af MEP og/eller neglekt. For de restende patienter, som havde MEP og ikke havde neglekt, kunne ikke-normal brug af arm og hånd prædikteres med en sensitivitet på 0.80 og en specificitet på 0.83.

Studie III: Formålet var at undersøge fysio- og ergoterapeuters holdninger til armprædiktionsmodeller.
Metode: Der blev afholdt fire fokusgruppeinterviews med hver 3-6 terapeuter. Data blev analyseret med tematisk indholdsanalyse. Meningsbærende enheder blev identificeret og grupperet i undertemaer og information på tværs af alle interviews blev sammenfattet i fire hovedtemaer.
Resultater: De fire hovedtemaer var: *Nuværende praksis; Fordele; Barrierer;* og *Betydning for Implementering*. Deltagerne havde begrænset kendskab til armprædiktionsmodeller men PREP2 blev anset som et potentielt brugbart redskab i forbindelse med tilrettelæggelse af behandling og målsætning. De primære barrierer for implementering var dels at modellernes blev anset for at være for upræcise

samt dels at skulle konfrontere patienter med en negativ prognose. En kommende implementeringsstrategi vil skulle tilpasses det enkelte afsnit, der skal afsættes tid til at tilegne sig nye færdigheder, og organisationen skal understøtte implementeringen.

Konklusion: På baggrund af studie I konkluderes, at PREP2 ikke bør implementeres i klinisk praksis, hvis den anvendes to uger efter apopleksi. Dog kan PREP2 bruges til med stor sikkerhed at forudsige enten *Fremragende Armfunktion* for patienter med god armfunktion to uger efter apopleksi, eller *Ringe Armfunktion* for patienter med begrænset eller ingen armfunktion i kombination med ingen MEP.
I studie II kunne funktion i arm og hånd to uger efter apopleksi prædiktere brug af arm og hånd. Prædiktion på individniveau var dog upræcis. Med stor sikkerhed kunne det fastslås, at patienter der ikke havde MEP ikke opnåede normalt brug af arm og hånd. Ligeledes kunne det fastståes, at patienter med neglekt ikke opnåede normal brug af arm og hånd.
I studie III var de erfarne neuroterapeuter skeptiske over for armprædiktionsmodeller. Dette skyldes primært at modellernes blev ansat for at være for upræcise samt bekymringer vedrørende negative prognoser. Dog blev PREP2 algoritmen anset som et potentielt nyttigt redskab.

Introduction

This PhD project aims to answer if upper limb (UL) prediction models can be used for prediction of UL function and UL use for the benefit of therapists and patients in a rehabilitation setting. The project is divided into three studies, all centering on UL prediction models, but viewing the topic from different angles and approaching it accordingly, using a combination of quantitative (Study I & II) and qualitative methods (Study III).

In Study I, the accuracy of an existing algorithm for prediction of UL function is examined, when the time window to obtain the prediction is expanded to two weeks after stroke.

In Study II, prediction of UL use is examined. The underlying rationale of Study II is that patients who engage in physical rehabilitation mainly seek improvement in movement performance within their daily lives. Thus, from a patient perspective, the prediction of UL use may be even more relevant than the prediction of UL function.

In Study III, the focus is shifted from the accuracy of prediction models to a future implementation of these models. Thus, the perceptions of the physiotherapists (PTs) and occupational therapists (OTs) potentially performing the UL predictions are explored. Study III may contribute with answers to why, despite a growing body of research, UL prediction models are not yet widely implemented in the clinical setting. Study III is a step to bridge the gap between evidence and practice that prevents the dispersion of new knowledge to the clinal setting.

By keeping this broad perspective on UL prediction models the aim is to contribute with new knowledge and broaden the understanding of the topic, thus bringing the implementation of UL prediction models in the clinical setting one step closer.

Background

Stroke and stroke epidemiology

Stroke is a leading cause of death and long-term disability in the western world.[5,6] The American Heart Association has estimated that the prevalence of stroke in adults is 2.7% in the United States, and each year approximately 795.000 people experience a stroke.[5] Approximately 610.000 of these are first attacks, and 185.000 are recurrent attacks. Of all strokes, 85 - 90% are ischemic and 10% are haemorrhages and the prevalence of stroke increases with age.[5,7] In Denmark 15.000 people annually experience a new stroke, equivalent to an incidence rate of 346 per 100.000.[8] In 2017, nearly 250.000 people in Denmark lived with a stroke[9] and the greatest cost of stroke in the country was associated with home care or practical aid after stroke.[8] High direct and indirect costs of brain disorders, including stroke, have been found, and the occurrence of stroke is expected to increase in the future.[10]

UL impairment is a frequent consequence of stroke and has been reported present in 48% of stroke survivors in the acute phase[11] and 30 - 66% of stroke survivors in the chronic phase.[12,13] Stroke survivors with impaired UL often experience subsequent functional limitations affecting activities of daily living.[14,15] Restrictions in participation and a consequent decline in health-related quality of life have been documented.[16]

The ICF in relation to the upper limb

The consequences of stroke in relation to the UL can be described within the International Classification of Functioning, Disability and Health (ICF) framework.[17] The ICF identifies three levels of human functioning: impairment, activity and participation level. Impairments are problems in body function or structure such as a significant deviation or loss, e.g. reduced range of UL movement, sensory dysfunctions or UL pain. Activity is the execution of a task, e.g. activities of daily living, or an action, and participation is involvement in a life situation. Activity limitations are difficulties a person may have in executing activities.[17]

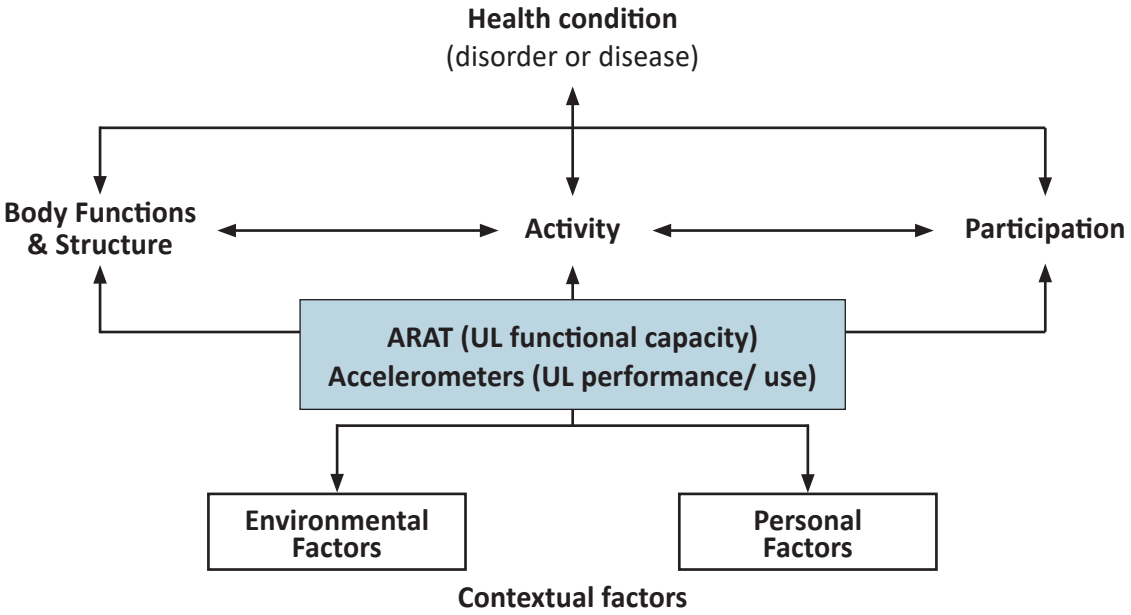
The ICF distinguishes between the capacity for use and actual performance.[17] Capacity, or function, indicates the highest probable level of functioning of a person at a point in time. Capacity is typically assessed in a standardized environment with a clinical test. Performance is what a person actually does in his or her usual, unstructured environment.[17] Performance may be assessed either via self-report with questionnaires or directly via wrist-worn accelerometers when a person engage in daily life activities.[18,19]

The UL capacity and performance are to some extend related and capacity is a prerequisite for UL performance. However, other factors than capacity influence performance. If capacity is higher than performance, then some aspect of the environment[17] or factors within the person, i.e. motivation or cognitive deficits, could be barriers to optimal performance.

Performance of daily life activities depends considerably on the recovery of motor functional capacity in the UL.[14,15] A major goal of UL rehabilitation is to facilitate that the paretic arm is engaged in activities of daily life and improvements in UL impairment and function should be transferred to improved UL performance in real life. The aim is an UL use pattern that resembles the pre-stroke levels as closely as possible.[20,21]

The present PhD project centres on both UL capacity and UL performance and outcome measures were chosen accordingly. In study 1, capacity or function at activity level was measured using the Action Research Arm Test (ARAT). In Study II, activity performance was measured using wrist-worn accelerometers (Figure 1). Both outcomes are described in more detail in the methods section.

Figure 1. Outcomes Used in the Present PhD project in Relation to ICF



ARAT: Action research Arm Test
Source: Modified from ICF figure[3]

Prediction of upper limp function

During the past two decades, several models for the prediction of UL function have been proposed.[22-34] According to this research, the initial UL function after stroke is the main predictor for UL recovery. In five prospective longitudinal

studies UL motor impairment was assessed with the Fugl-Meyer Motor Assessment Upper Extremity (FMA) within 2 weeks of stroke, and at 3 or 6 months after stroke.[24,28-31] These studies showed that most patients recover 70-80% of their maximum possible UL motor function within 3 to 6 months after stroke.[24,28-31] However, great variation between individuals exists[35] and a substantial number of patient with severe UL impairments improved markedly less than predicted.[24,29]

Whereas existing UL models are most accurate for predicting recovery in patients with mild to moderate UL impairment,[22-27] prediction of future UL function in patients with severe UL impairment may be improved by the use of a biomarker.[23,36-39] According to a recent consensus paper[2], a stroke recovery biomarker can be defined as "an indicator of disease state that can be used as a measure of underlying molecular/cellular processes that may be difficult to measure directly in humans." Thus, a biomarker can be used to predict a future outcome or recovery (defined as the change in the clinical score) or a treatment response.[40]

A biomarker widely used in UL prediction studies is the motor-evoked potentials (MEPs), motor contractions elicited by pulses of transcranial magnetic stimulation (TMS).[23,35-41] TMS is a safe, non-invasive tool, that can be used to stimulate the primary motor cortex and test the functional integrity of the ipsilesional corticospinal pathway, and thereby establish if MEPs are present.[42] According to a recent review, MEPs at rest was the only biomarker predicting UL function in stroke patients with severe UL impairment.[36] Patients in whom MEPs can be elicited in muscles of the affected UL limb have been found to experience a higher amount of UL improvement compared to patients without MEPs.[2,23,29,38,39,43]

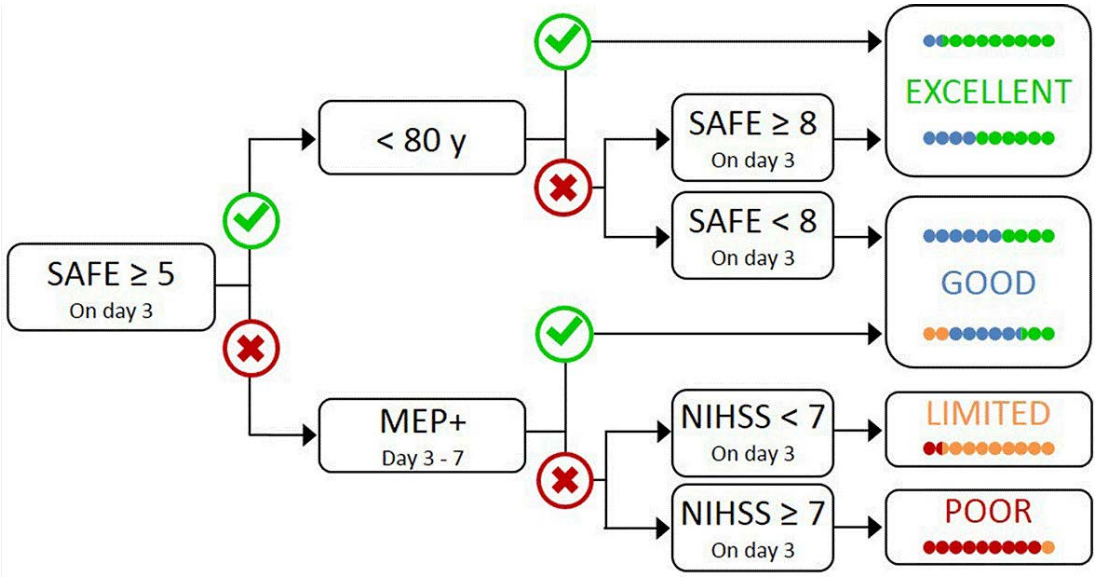
The Predict Recovery Potential (PREP2) algorithm is an UL prediction model that has incorporated information obtained from a biomarker.[22,23] The PREP2 stands out, as its accuracy for patients with severe paresis exceeds that of previous prediction models.[22,23,44] PREP2 predicts UL function at three months after stroke in one of four categories, based on the Action Research Arm Test (ARAT).[45,46] The

category *Excellent* includes the ARAT scores of 51 to 57, *Good* includes the ARAT scores of 34 to 50, *Limited* includes the ARAT scores of 13 to 33, *Poor* includes by the ARAT scores of 0 to 12.

The PREP2 algorithm is a process in three stages (see Figure 2, page 16).[23] In stage one, Shoulder Abduction and Finger Extension strength are scored separately between 0 to 5 (max). The two sub-scores are added to comprise a SAFE score of 0 to 10 (max). The second stage of PREP2 varies depending on the SAFE score. If the SAFE score ≥ 5 information on age (below or above 80 years) is used and the patient is predicted to have either *Excellent* or *Good UL function*. For patients with a SAFE score below 5, TMS is needed to test the function of motor pathways between the stroke-affected side of the brain and the affected arm. If MEPs can be elicited (MEP+) the patient is predicted to have a *Good UL function*. If MEPs cannot be elicited (MEP-) the patient's National Institute of Health Stroke Scale score (NIHSS), is used.[47] NIHSS is a measure of stroke severity and depending on the score, the patient will be predicted to achieve either *Limited* or *Poor UL function.*

The PREP2 was developed from an analysis of data derived from two longitudinal studies of patients, recruited within three days after stroke.[23] At three months after stroke the algorithm correctly predicted UL function for 156 of 207 patients (75%). Of the remaining 51 patients, PREP2 was too pessimistic for 1/3 and too optimistic for 2/3 of the patients. For patients with a SAFE ≥ 5 accuracy of prediction was 78%.[23] For patients with a SAFE score below 5, accuracy was only 55% if information on MEP status was not included. However, if information on MEP status was included, prediction accuracy for this subgroup of patients with severe UL paresis increased to 70%.[23]

Figure 2. The Predict Recovery Potential (PREP2) Algorithm.



SAFE: Shoulder Abduction and Finger Extension. < 80 y: Below 80 years old. MEP+: motor- evoked potentials present. NIHSS: National Institute of Health Stroke Scale. Excellent: Potential to make a complete, or near complete, recovery of hand and arm function within 3 months. Good: Potential to be using their affected hand and arm for most activities of daily living within 3 months. Limited: Potential to regain some movement in their hand and arm within 3 months. Poor: Unlikely to regain useful movement in their hand and arm within 3 months. *Figure copied from the PRESTO homepage.*[48]

Prediction of UL use

It is often assumed that increased UL function assessed in a clinical setting equals increased UL use in daily life[19] and activity level measures recommended in clinical practice and research guidelines nearly always assess capacity, not performance.[49,50] However, several studies have shown that while UL capacity and UL performance are related, improvements in capacity, or what a person is capable of doing, are not necessarily reflected in increased performance or daily life

use.[18,21,51-54] Capacity of the affected UL often exceeds actual use,[53] and it has been shown that learned non-use of the paretic UL reduces the level of use.[20,52,55] A substantial group of stroke survivors may also perceive less function than clinical tests would suggest.[56] Also, UL use may be influenced by several other factors, e.g. motivation[20] or attention and arousal.[21]

Whereas the prediction of UL function or motor recovery has been examined in several studies, the prediction of UL use in daily life is an emerging research area and predictive factors for UL use have been only sparsely investigated. However, patients who engage in physical rehabilitation mainly seek improvement in movement performance within their daily lives.[57] Thus, from a patient perspective, prediction of UL use may be even more relevant than prediction of UL function. In a recent study, 20 chronic stroke survivors with mild to moderate UL impairments, characterized by FMA, were assessed for learned non-use with a modified version of the Actual Amount of Use Test.[21] The Actual Amount of Use Test measures the disparity between amount of use in spontaneous versus forced conditions. Patients were also assessed with measures of limb apraxia, spatial neglect, attention/arousal, and self-efficacy. The authors concluded, that FMA and attention and arousal predicted the degree of non-use.[21]

Wrist-worn accelerometry enables measurement of UL use in the unstructured environment[60] and accelerometry is a well-established method for capturing UL use in nondisabled adults and adults with stroke.[58,59] It could be assumed, that mainly the dominant UL would be engaged in daily life activities. However, in accelerometer studies of nondisabled adults bimanual UL activity makes up a significant portion of daily activity and the dominant and non-dominant UL are used to a similar degree.[61-63]

The only study found that examined potential long-term predictors of UL use after stroke was by Rand & Eng.[18] In their study, UL function was assessed early after stroke and daily life UL use was assessed with wrist-worn accelerometers 1 year

after stroke. The authors concluded, that better UL function at discharge predicted increased UL use after one year. However, even in patients with only mild UL impairments, daily life use was still reduced compared with healthy controls.

Implementation of prediction models

At the time of this PhD project, the only study identified that reported on implementation of an UL prediction model in a clinical setting, was by Stinear et al.[64] In this study, the first version of PREP2 was implemented in the clinical setting where it was developed and it was shown, that the UL predictions modified therapy content and increased rehabilitation efficiency.[64] The study implies, that PREP2 is a promising tool for clinical application, and this conclusion is further supported by a review, that recommends PREP2 for further validation.[22]

However, before commencing the present PhD project, no studies on clinical implementation of PREP2 outside the setting where it was developed were detected. It has been reported to take an average of 17 years for new evidence to become embedded into clinical practice[65] and this gap between evidence and practice denies patients the opportunity to benefit from new knowledge.[66] The lack of studies on implementation may reflect that knowledge obtained from clinical studies is not necessarily easily adopted in the clinical setting and a focus on implementation is needed if patients are to benefit from the developments.[67,68] A recent survey study confirms that at least in Denmark, UL prediction models are not yet a part of daily practice in stroke rehabilitation.[69] The study was conducted amongst Danish PTs and OTs employed in neurology or neurorehabilitation and revealed that despite therapists' considering knowledge of prognosis relevant in their clinical work, UL prediction models were not yet an integrated part of daily practice.[69]

A main obstacle for implementing PREP2 in a rehabilitation unit may be the time points of the initial assessment with SAFE and TMS, which is at day 1-3 and 3-7 after stroke, respectively. In several countries, including Denmark, patients are transferred from the acute stroke units to various subsequent neurorehabilitation services during the first days or weeks after stroke. This short stay at the acute unit leaves little time for prognostic evaluation. A recent paper by Connell et al.[67] focuses on how the implementation of PREP2 can be facilitated. The authors proposed, that future research should examine whether the time windows to obtain of SAFE and TMS can be expanded.[67]

As most recovery occurs within the first three months after stroke, it is essential that all patients are assessed at a fixed point in time after stroke.[50]
In 2018 and 2019 patients were admitted to RHN a median of ten days after stroke and around 2/3 of the patients arrived within two weeks after stroke. In the present PhD project, the predictions were made two weeks after stroke to include as many patients in the subacute phase as possible. Predictions made two weeks after stroke may be used to inform therapists about the expected recovery potential and can guide the choice of UL intervention and treatment. Patients and relatives can be informed on UL prognosis, enabling them to adjust their expectations and plan for the future.

UL predictions of function can support individual goals for rehabilitation and may result in more effective utilization of health resources.[22,23,44] If the PREP2 algorithm could be applied two weeks after stroke with satisfactory accuracy, this would facilitate its implementation.

Another important factor for a future implementation is whether the healthcare providers regard an intervention or an assessment as meaningful and useful for themselves and their patients.[70,71] To ensure successful implementation in a clinical setting, a crucial first step is identifying and describing potential barriers and facilitating factors for UL prediction algorithms.[70,72,73]

Gap of knowledge

In times of limited resources, the prediction of UL function in stroke rehabilitation is highly relevant in order to provide targeted rehabilitation.
However, existing prediction models may not be applicable in most rehabilitation settings, due to the fixed time points of the assessments very early after stroke. This PhD project set out to modify the PREP2 prediction algorithm in a way that would extend its applicability.

To be truly meaningful, improvements in UL function should be reflected in improved UL use in daily life. However, prediction of UL use is a new research field, and factors that predict UL use have received little attention. Thus, further high-quality longitudinal studies that identify predictive factors of UL use at a future time point are needed.

The clinicians responsible for UL treatment and most likely to obtain and use the PREP2 predictions are PTs and OTs. To the knowledge of the PhD fellow, it has not previously been explored how therapists in a stroke rehabilitation setting perceive UL prediction with the help of the PREP2 algorithm.

Aims and hypothesis

The overall purpose of this PhD project is to examine the topic UL prediction after stroke. The project is divided into three studies, and the specific aim for each study is outlined below.

Study I

The aim of Study I was to assess the prognostic accuracy of the PREP2 algorithm when applied in a neurorehabilitation setting two weeks after stroke.
The secondary aim was to assess if modifications of the algorithm at this point in time could improve prediction accuracy.

It was hypothesized that the prediction accuracy of PREP2 applied two weeks after stroke would be similar to its original application. Thus, an overall correct classification rate (CCR) of 75% (95% CI 65- 85%) was hypothesized.

Study II

The primary aim of Study II was to assess if UL impairment two weeks after stroke could predict real-life daily UL use three months after stroke. The secondary aims were to identify additional key predictors of UL use, and characteristics of patients who did not achieve normal UL use.

It was hypothesized that UL function two weeks after stroke was a statistically significant predictor of UL use three months after stroke and that other factors too contributed to the prediction of UL use.

Study III

The aim of Study III was to explore how therapists in a neurorehabilitation setting perceive UL prediction models in general, and the PREP2 algorithm in particular. Furthermore, to identify potential barriers to and facilitators of implementation.

Materials & methods

Design

Study I & II

A prospective, observational longitudinal study was undertaken to examine the aims of Study I & II.

Study III

This was a qualitative study using focus group interviews.

Study setting

All three studies were conducted at a Hammel Neurorehabilitation Centre and University Research Clinic (RHN), Denmark. The RHN is distributed across three physically distinct rehabilitation units. Unit 1 is the largest with app. 70 beds, units 2 has 30 beds and unit 3 has 15 beds. While adult patients with stroke attend all three units, a number of the beds at unit 1 are allocated patients with severe (traumatic) acquired brain injury. A research department is placed in connection to Unit 1.

Patients are admitted to RHN if they are considered to benefit from in-patient rehabilitation. Each year approximately 500 patients with stroke are admitted from various stroke units. A substantial number of these patients have UL impairments.

A total 67 physiotherapists (PTs) and 67 occupational therapists (OTs) are involved in the treatment of patients, and the rehabilitation is organized in teams. Some of the therapists are assigned key positions, e.g. specialist PTs or specialist OTs, and are responsible for professional development.

Study participants

Study I & II

Patients were included consecutively from June 2018 to October 2019.

The inclusion criteria were:

- First or recurrent hemorrhagic or ischemic stroke.
- Admitted within 2 weeks after stroke.
- SAFE score < 10.
- Age ≥ 18 years.
- Ability to cognitively comply with examinations, defined by a FIM cognitive score ≥ 11 in combination with the rehabilitation team considering the patient able to participate.

Exclusion criteria were:

- Subarachnoid haemorrhage.
- Prior UL impairment, e.g. from an injury or a previous stroke, as this would impede the potential for complete UL recovery.

Additional criteria to be fulfilled:

- For study I only: Prediction of UL function obtained at baseline.
- For study II only: Accelerometer data available at follow-up.

Study III

The participants for study III were OTs and PTs employed at RHN.

Procedure

Study I & II

Patients who fulfilled all eligibility criteria were invited to participate. After signing informed consent, demographic information (including age, sex, comorbidities) and stroke details (including stroke location, lesion type, Functional Independence Measure score and NIHSS score), were extracted from the medical records.

Baseline assessments

Included patients were examined with a range of different assessments at baseline, two weeks after stroke, and at follow-up, three months after stroke. Some of the assessments were used in Study I only and others in Study II only. A range of additional assessments was used to describe the study population and enable comparison with other populations. The assessments are described below. An overview of the assessments and the time line for each study is displayed in Figure 3, page 29.

- UL impairment was assessed with FMA.[45,74,75] The FMA consists of 33 sub-items divided into 4 subsections: shoulder-arm, wrist, hand, and coordination. Each sub-item is scored on an ordinal scale from 0 - 2, with a sum score of 0 - 66 points (best). The psychometric properties such as concurrent-, predictive-, content- and construct validity, reliability, and responsiveness of the FMA are well established.[45,74,75] To ensure reliability in the present PhD project a scoring manual with a detailed description of the testing procedure was used.[74]

- UL function/ capacity was assessed with ARAT.[45,46,50,76] The ARAT evaluates 19 sub-items of arm motor function, both distally and proximally. Patients can score from 0 - 57 (best). ARAT is found to be reliable and valid.[45,46,76] To further ensure reliability a scoring manual was used.[46] FMA and ARAT are internationally recommended for use in clinical trials.[50]

- Shoulder abduction and finger extension strength were scored separately from 0-5 using the medical research council grades for limb power. The two scores were added to form the SAFE score from 0 - 10 (best).[23]

- In patients with a SAFE score < 5, TMS was used to assess MEP status. The TMS procedure was conducted in line with international recommendations.[42] Screening for contraindications and establishment of MEP status were performed in accordance with the protocols from Stinear et al.[77,78] Absolute contraindications were metal implants in the head, implanted electronics, epilepsy, skull fracture or serious head injury, brain surgery and pregnancy.[42,78]

  During the TMS procedure, patients were seated with the affected UL placed in a relaxed position on a table. Electromyographic activity was recorded from the first dorsal interosseous and the extensor carpi radialis muscle. Magnetic stimulation was delivered using a 70-mm figure-of-eight coil connected to a MagStim 200 unit (Magstim Co. LtD) and consisted of monophasic pulse waveforms. The coil induced a posterior-to-anterior current flow in the ipsilesional M1 and stimulus intensity began at 50% of the maximal stimulator output. To

locate the optimal site for producing MEPs the assessor moved the coil in 1 cm steps (anterior, posterior, medial, lateral) and delivered app. 3 stimuli at each scalp location. Stimulus intensity was increased in steps of 10% until MEPs were consistently observed in one or both muscles or until 100% stimulator output was reached. If MEPs were not observed, the patient should attempt to make a firm fist with affected and also the non-affected hand as this may facilitate MEPs.[77]

The acquired data were visually inspected and stored with a custom-made LabVIEW (National Instruments, TX, USA) software (Mr. Kick, Knud Larsen, Aalborg University, Denmark). The patient was classified as MEP+ if MEPs were observed in response to a minimum of 5 consecutive stimuli with a peak-to-peak amplitude ≥ 50 µV and at a consistent latency.[42,77,79] If MEPs were not found, the patient was categorized as MEP-.77 The TMS procedure was performed by the PhD fellow and MEP status was established by a researcher who was blinded to the results of the clinical assessment. As MEP is an indication of corticospinal tract integrity, presence of MEP was assumed in patients with a SAFE score ≥ 5.

- Inferior subluxation in the glenohumoral joint was assessed by palpation of the subarchrominal space and scored 0 (no subluxation) to 5 (2½ finger widths subluxation). This method has been found reliable.[80]

- Light touch and proprioception were assessed with the Fugl-Meyer Sensory Assessment Upper Extremity.[81] Six sub items are scored on an ordinal scale from 0 - 2, the patient can score from 0 - 12 (best).

- Bilateral stimulation was assessed in the palmer surface of the hand in accordance with the Nottingham Sensory Assessment Scale[81] from 0 - 2 (best).

- Two-point discrimination (twopd) was assessed at the pulp of the index finger with a Discriminator. Discrimination thresholds ranged from 2 - 15 mm, with

lower scores indicating higher discriminative acuity. In accordance with a previous study a score of 16 was given if twopd was absent.[82] If discrimination was above the thresholds for healthy age-matched individuals, e.g. above 6 mm for a person aged 60 - 69 years, twopd was considered affected.[83]

- Pain was rated on a numerical rating scale and patients rated their UL pain from 0 - 10 (worst pain).[84]

- Neglect was assessed with the Star Cancellation Test and the Line Bisection Test, as previous studies have recommended that a combination of tests are used to diagnose the neglect syndrome.[85,86] In this PhD project, patients were classified with neglect if they had neglect on one or both neglect tests.[86,87]

  In the Star Cancellation Test, the patient was presented with a page containing 52 large stars, interspersed with letters, short words, and 56 smaller stars. The patient was instructed to cross out the small stars. To analyze presence and severity of neglect, the cancelled small stars were entered in a computer program for measuring the centre of cancellation index.[86,87] On the Star Cancelleation Test neglect was present if centre of cancellation was above 0.083 after a right hemisphere brain lesion or below -0.083 for left hemisphere brain lesion.[86,87] This was the case if number of small stars omitted were 51 or below, and the center of omission was to either the right or left of the midline. The center of cancellation not only takes into account the number of omissions, but also their specific location, resulting in one outcome measure that distinguishes spatially biased performance from inattentive performance.[86,87]
  In the Line Bisection Test, the patient was instructed to estimate the mid-point of three lines. Deviations from the actual mid-point were noted. Using a scoring-sheet the patient could score 0 - 9 (max). In the Line Bisection Test neglect was present if the score was ≤ 7.

- Walking ability was scored with the Functional Ambulation Classification.[88] Scores ranges from 0 - 5 (best).

Follow-up assessments

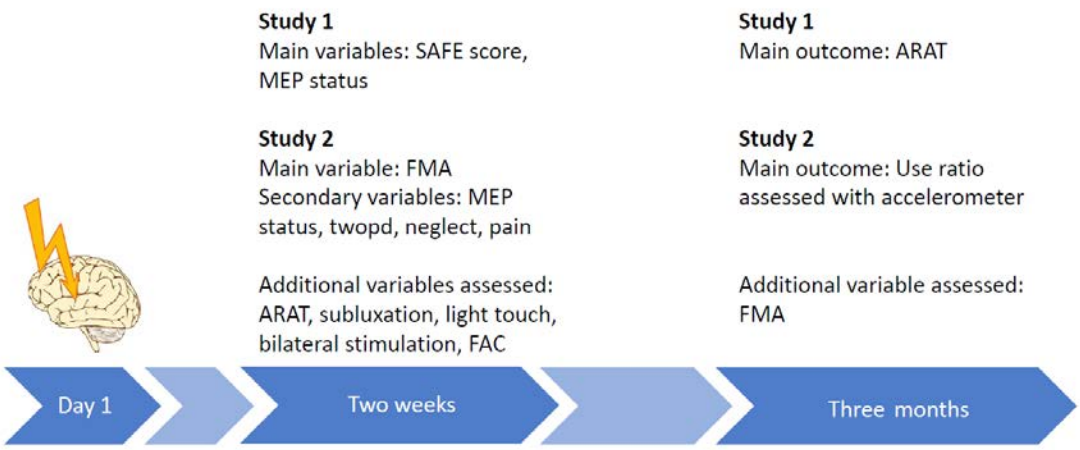
At three months after their stroke, most patients were at home. A research therapist assessed the patients and also delivered the accelerometers to the patients.

- The primary outcome in Study I was ARAT (described above).

- The primary outcome in Study II was real life use measured with wrist-worn accelerometers and expressed as the use ratio between paretic and non-paretic UL. Validity and reliability for accelerometers are well-established for measuring UL use in non-disabled adults and adults with stroke.[58,59] Accelerometers are described in more detail below the specific procedure for Study II.

- Additionally, to describe the population, FMA was assessed at follow-up.

Figure 3. Overview of Predictor Variables Used in Study I & II.



SAFE: Shoulder Abduction Finger Extension. MEP: Motor-evoked Potentials. FMA: Fugl-Meyer Motor Assessment Upper Extremity. ARAT: Action Research Arm Test.

Baseline assessments were performed by the PhD fellow, who was not involved in patient care. Follow-up assessments were performed by three experienced research therapists, blinded to baseline scores, the predicted categories (Study I only), and not involved in patient care.

Before commencing the study, all assessors were instructed in the FMA and ARAT scoring procedure. Several patients were assessed by all assessors and the results discussed until consensus was achieved. This calibration process was repeated after three months. In cases of doubt on how to score a certain item, the PhD fellow was contacted.

Inclusion in the longitudinal study did not affect patient rehabilitation or choice of UL treatment. Length of stay, constitution and intensity of training were individually arranged by the rehabilitation team, in cooperation with the patients and their relatives. The rehabilitation included 45 min of physiotherapy and 45 min of occupational therapy on weekdays and twice this amount for patients with severe brain damage. Members of the rehabilitation team were blinded to the clinical measurements and in Study I also to the baseline prediction.

Specific for Study I

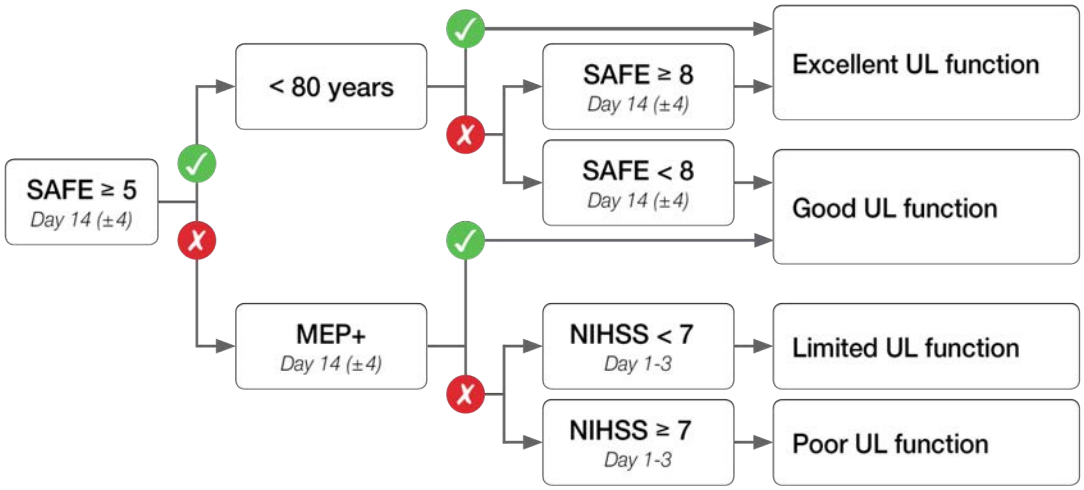
Included patients had their future UL function predicted in line with the PREP2 prediction.[23,89] (Figure 4).

In line with the PREP2 procedures, the outcome was predicted in one of four ARAT categories. The category *Excellent* comprises the ARAT scores of 51 - 57, *Good* 34 - 50, *Limited* 13 - 33, and *Poor* 0 - 12.

Originally, the SAFE score was obtained within 3 days after stroke and MEP status at day 3 - 7 after stroke.[23] In the present study, the SAFE score and MEPs were

obtained two weeks after stroke (Figure 4). Information on age and NIHSS score, or the comparable Scandinavian Stroke Scale (SSS) score, was routinely assessed within three days after stroke and could be extracted from the medical record as proposed by Stinear et al.[23] Patient with a SAFE < 5 had their MEP status established with TMS.

Figure 4. The Predict Recovery Potential Algorithm Performed Two Weeks After Stroke



SAFE: Shoulder Abduction and Finger Extension. < 80 y: Below 80 years old. MEP+: motor-evoked potentials present. NIHSS: National Institute of Health Stroke Scale. Excellent: Potential to make a complete or near complete recovery of hand and arm function within 3 months. Good: Potential to use their affected hand and arm for most activities of daily living within 3 months. Limited: Potential to regain some movement in their hand and arm within 3 months. Poor: Unlikely to regain useful movement in their hand and arm within 3 months.

Source: Replicated from Study I[89]

Specific for Study II

The primary outcome was real life use expressed as the use ratio between paretic and non-paretic UL.[89]

A research therapist instructed the patients on how and when to don the pre-programmed accelerometers. The accelerometers had Velcro straps for easy handling, but if the patient needed help, arrangements were made with either a relative or a home carer. The accelerometers had to be worn on both wrists for a 12-hour period from 08:00 to 20:00 on an average day within a week after follow-up assessment. Patients were encouraged to wear the accelerometers when pursuing their normal, daily routines, and were advised not to change their behaviour or increase their UL activity. Previous research has shown that activity levels do not increase in response to wearing accelerometers.[90] The accelerometers were returned to the research unit in a prepaid envelope.

Accelerations were recorded along three axes at 50 Hz. Accelerometry data were downloaded using ActiLife 6 software, which band-pass filtered data between frequencies of 0.25 and 2.5 Hz, used a proprietary process to remove acceleration due to gravity, down-sampled data to 1 Hz (i.e., 1 s) samples, and converted acceleration into activity counts (0.001664g/count).[61] ActiLife 6 was also used to visually inspect the accelerometer data to ensure that the accelerometers functioned properly during the recording period. The CSV files from ActiLife were imported to Matlab and the relevant 12-hour intervals were identified and exported to STATA 16. In STATA 16, activity counts were combined across the three axes to create a vector magnitude √x2 + y2 + z2 for each second of data and the following accelerometry-derived parameters were calculated, using the approach described by Bailey et al61: hours of paretic UL use, hours of non-paretic UL use, use ratio, hours of bilateral UL use, magnitude ratio, and bilateral magnitude.

Total hours of paretic and non-paretic UL use are the total time that the specific limb was active during a 12-hour period as measured by summing up the seconds with activity. The use ratio is total hours of paretic UL use divided by total hours of non-paretic use. A use ratio of 0.5 indicates that the paretic UL is active 50% of the time the non-paretic is active. In the present Study II, the use ratio was used as the primary outcome as it, compared with other accelerometry outcomes, is less dependent on varying activity levels between different people.[19]

The bilateral magnitude quantifies the intensity of activity across both ULs, and was calculated for each second of activity by summing up the vector magnitude of both ULs.[60,61] Bilateral magnitudes of 0 indicate that no activity occurred across either UL while increasing bilateral magnitudes indicate increasing activity intensity.

The magnitude ratio quantifies the contribution of each UL to activity, for every second of data. The magnitude ratio value is the natural log of the paretic UL vector magnitude divided by the vector magnitude of the non-paretic UL.[60,61] Negative magnitude ratio values represent greater use of the non-paretic UL, while positive numbers represent greater paretic UL use.

Study III

In the qualitative study, the Consolidated Framework for advancing Implementation Research (CFIR) was applied as a guiding framework to develop a semi-structured interview guide and structure data collection.[70,72,73] The CFIR is composed of five domains: intervention characteristics, outer setting, inner setting, characteristics of the individuals involved, and the process by which implementation is accomplished.[70,72,73] The CFIR domains explored in this study were intervention characteristics, inner setting and characteristics of the individuals involved. The participants' views and attitudes within these three domains were expected to be important to a future implementation. On the contrary, the structure and organi-

zation of the fourth domain, outer setting, would not be influenced by the views and attitudes of the participants and the fifth domain, implementation process, was still in a preliminary phase.

The interview guide was tested for comprehensibility in a test interview with a PT and an OT followed by pilot focus group interview with three PTs. The test interview and pilot focus group interview resulted in minor corrections: the number of questions was reduced or merged and information about prediction algorithms was simplified. The interview guide is presented in Table 1. Information posters displaying illustrations about the topic, e.g. the PREP2 algorithm, were composed in order to support explanations and facilitate discussion in the subsequent interviews.

The ward managers invited participants based on the following criteria: a mix of PTs and OTs, at least one year of clinical experience in neurorehabilitation, involved in the treatment of patients, and from different wards. The intention was to achieve maximal variation regarding profession, clinical experience, and degree of specialization.[91]

An information letter was sent to the participants, explaining the purpose of the interviews and the background for UL prediction models. The participants were instructed to perform step 1 of the PREP2, the SAFE test, on a minimum of three patients before participation in the interviews. Performance of the SAFE should ensure practical experience with the test and qualify the interview discussions.

The focus group interviews were explorative and focused on the feasibility and perceived usefulness of UL prediction models, in particular the PREP2 algorithm. Focus groups are an appropriate method to illuminate the shared experiences and different perspectives of the group and the interaction between participants was expected to stimulate discussion of beliefs, thoughts and attitudes.[92,93]

Table 1. Interview Guide

Main categories	Questions
General questions	In patients with paresis of arm and hand: Which factors do you consider relevant for future arm and hand function? (important elements) What is relevant for your own approach to treatment of the arm and hand? (write down three - four issues/ things)
Thoughts on prediction	What are your thoughts about prediction of arm and hand function at an early point in time? What are the likely consequences? Which patients/ groups of patients would benefit from knowledge of prognosis (e.g. paralyzed UL)? UL prediction models: to whom will it not make sense? Does age matter for prognosis (in general and for UL in particular)? Severity of stroke from onset is relevant for UL prognosis. Where do you seek this information (e.g. ward round, medical record, looking for particular scores as NIHSS or SSS)? Do your expectations of future UL function influence your approach to the patient and choice of UL treatment?
SAFE score	Before participation, you were asked to perform a SAFE test on at least three patients. How was it? What are your thoughts on using specific UL tests for (all) patients with reduced strength in arm and hand (e.g. SAFE, FMA) Are you aware of other hospitals focusing on UL prediction? E.g. if they use SAFE?
Knowledge of evidence	How do you update your knowledge on UL treatment? Do you have the time and opportunity to get updated on new knowledge? Exercise: I explain the PREP2 algorithm and show pictures of the elements: What are the pros and cons of the PREP2? What would it take for you to use a UL prediction model? Do you see patients for whom a prediction model would make no sense? Would use of a UL prediction model change your approach to a patient? PREP2 can predict future UL function with approximately 75% accuracy. What is your opinion on that? Transcranial magnetic stimulation (TMS) - can it be use in your clinical setting?
Summarising	What we have talked about. Do you have anything you would like to add?

Source: Replicated from Study III (unpublished)

The focus group interview was moderated by the PhD fellow, who was aware of ensuring a confident atmosphere that welcomed a diversity of opinions. A senior researcher participated in all interviews and asked clarifying questions, observed interactions between participants and provided feedback to the moderator. Immediately after ending an interview, the overall impression and any reflections were noted. The interviews were audio-recorded and transcribed verbatim by the PhD fellow.

Data analysis

Study I & II

The required number of patients to include in the longitudinal study was based on a power calculation for Study I, assuming a correct classification rate (CCR) of 75% with a CI 95% of 65- 85%. A CCR of 75% was chosen as this was in line with the accuracy found in the original PREP2 study.[23] Allowing for a 20% drop-out, it was decided to include at least 90 patients.[89]

STATA 16 was used for data analysis. Data were visually inspected with histograms, boxplots, qq-plots and dotplots to determine the distribution of normality. Continuous baseline characteristics, stroke details, baseline and follow-up scores were summarized by mean, standard deviation (SD), min, and max when normally distributed; otherwise by median, interquartile range (IQR), min, and max.

Demographic and clinical characteristics of the patients who were unavailable for the three-month follow-up were compared with those available to determine if the difference was statistically significant. The unpaired t-test or the Wilcoxon rank sum test was used for continuous data and the Chi^2 test for dichotomous data.

Specific for Study I

Improvement in UL impairment on FMA and UL function on ARAT from baseline to follow-up was examined. As FMA and ARAT are ordinal scales and data were non-normally distributed, within-group difference on the two scales from inclusion to follow-up was tested with the nonparametric Wilcoxon signed rank test.

The overall accuracy of the PREP2 was quantified by comparing the agreement between predicted and achieved ARAT categories using the CCR.[89] The CCR, along with sensitivity and specificity, were calculated for each of the four categories. Also, CCR was calculated separately for patients with a SAFE score < 5 or ≥ 5 to differentiate between patients with either severe UL impairment at baseline, who had MEP status obtained, and patients with relatively mild UL impairment at baseline, who did not need to have MEP status obtained.

To examine if prediction accuracy of PREP2 obtained two weeks after stroke could be improved, a classification and regression tree (CART) analysis was carried out.[89] CART analysis produces a decision tree without the user determining which variables to include or their order in the tree.[94,95] The CART analysis was based on the components of PREP2: SAFE score, age, NIHSS score, and MEP status. For patients with a SAFE ≥ 5, MEP+ status was assumed in the analysis.

Specific for Study II

Accelerometer data were displayed for the whole group and in line with a recent study also in three categories, each reflecting a range of scores on FMA at baseline.[62] The category "Severe" comprised the FMA scores of 0-22, "Moderate" 23-50, and the category "Mild" the scores 51-66.[62]

Prediction of use ratio

Several regression models were created. The first model, Model 1, was a linear regression model to assess the strength of the (unadjusted) association between baseline FMA score and UL use ratio at three months. In Model 2, a multiple regression model, the association between FMA at baseline and use ratio was adjusted for other secondary variables chosen a priori, based either on the results of previous studies or clinical reasoning. The independent variables and their distribution were assessed (their dispersion, frequency distributions). Moreover, the relationship between the independent variables, one at a time, was assessed.

Secondary variables chosen a priori were: MEP status (MEP present/ not present). Neglect (dichotomized into present/ not present). Dominant UL affected was included as previous research has demonstrated that dominant side affected may result in better UL stroke recovery.[18,61,96] Twopd (affected/ not affected), as previous research has shown this was a predictor for future UL function.[82] The FIM score, reflecting the need for assistance in daily life activities, was entered as a continuous variable from 18 - 126 (max). Gender, as older women use their dominant hand more in daily life compared with older men.[97] Severity of pain, a continuous score of 0 - 10.

In Model 3, the contribution of the biomarker MEP was assessed by removing MEP status from model 2 and comparing the fit of the model with and without MEP. Furthermore, the contributions of the individual predictive variables were examined. Finally, to assess the strength of each potential predictor univariate regression between each of the predictor variables and use ratio was performed.

All necessary assumptions for generalized linear models, including linearity, equality of variance, and normality of errors were visually inspected for all models and found adequate. Presence of multi-linearity was examined by the Variance Inflation Factor for each independent variable. Using a conservative approach, VIF below 3 were accepted.[98] Multi-linearity was not present.

The ability of the models to predict use ratio was assessed by the size of the adjusted R^2. The contribution of each individual predictor in the model was assessed from the significance level, size of p-value and the size of the β-coefficient including the 95% CI.[99]

To assess the ability of the models to predict future use ratio for an individual patient, the 95% prediction interval (PI) for the regression line was calculated based on the SD for the adjusted R^2 (PI = ± 1.96 * SD). The PI is an estimate of the interval in which a future observation of UL use ratio will fall, with 95% probability, given what has already been observed.

Normal and non-normal use ratio

Use ratio was dichotomized into normal and non-normal using a threshold based on an established reference value from a study with 74 community-dwelling adults.[63] In the reference population the mean use ratio was 0.95± SD 0.06, range 0.79-1.1.[63] In the present study the lower limit of the PI interval for the reference value was calculated (0.95- 1.96* 0.06=0.83) and used as a conservative threshold for normal use ratio. According to this, patients with a use ratio above or equal to 0.83 were classified as having a normal use ratio, and patients with a use ratio below 0.83 as having a non-normal use ratio.

The association between the use ratio and each of the variables FMA, MEP status, neglect, dominant UL affected, twopd and FIM were visually inspected followed by a multivariate logistic regression. To maintain adequate power for the statistical analysis the events per variable rule, which calls for at least ten outcomes for each variable in the regression model, was compiled with.[100,101] A receiver-operating curve (ROC) of the logistic model was graphically displayed, and a two-way contingence table was used to identify the cut point with the highest sensitivity and specificity values.

Study III

The interview transcripts were imported to the qualitative research software program NVivo12. The pilot focus group interview was considered to add interesting aspects to the topic and data from this interview were analysed along with data from the succeeding three interviews.

A thematic content analysis of the interviews was performed.[91,102] The analysis was both a deductive and an inductive process.[91,102] Deductive, as the CFIR framework was used as the aim was to answer specified pre-defined question regarding barriers and facilitators (theory-based coding). Inductive, as to let the material talk because attitudes towards UL prediction algorithms have not previously been explored, and knowledge of how to implement algorithms into the clinic setting is scarce (data-based coding). First, meaning units were identified and the four interviews were individually open-coded in NVivo. Second, the interviews were compared for similarities and divergences and subthemes were established. Finally, information gained from all four interviews was synthesized.

The coding and interpretation of results were continuously discussed with co-authors. This triangulation between authors with different perspectives and positions will increase the understanding of complex phenomena.[103] Several perspectives appeared repeatedly in all four interviews, indicating data saturation.

Ethical issues

Study I and II

The longitudinal study was approved by the Regional Ethics Committee (record number 628213). The study was reported to the Danish data protection agency. In line with the Helsinki Declaration, included patients provided written informed consent. Inclusion in the longitudinal study did not affect patient rehabilitation or choice of UL treatment and patients were informed that they could withdraw at any time.

Study III

Participation in Study III was voluntary and all participants signed informed consent. Anonymity was secured by changing names and identifiable situations or places. In accordance with Danish legislation on research ethics approval by the Regional Ethics Committee was not required. The study was conducted in line with the Helsinki Declaration.

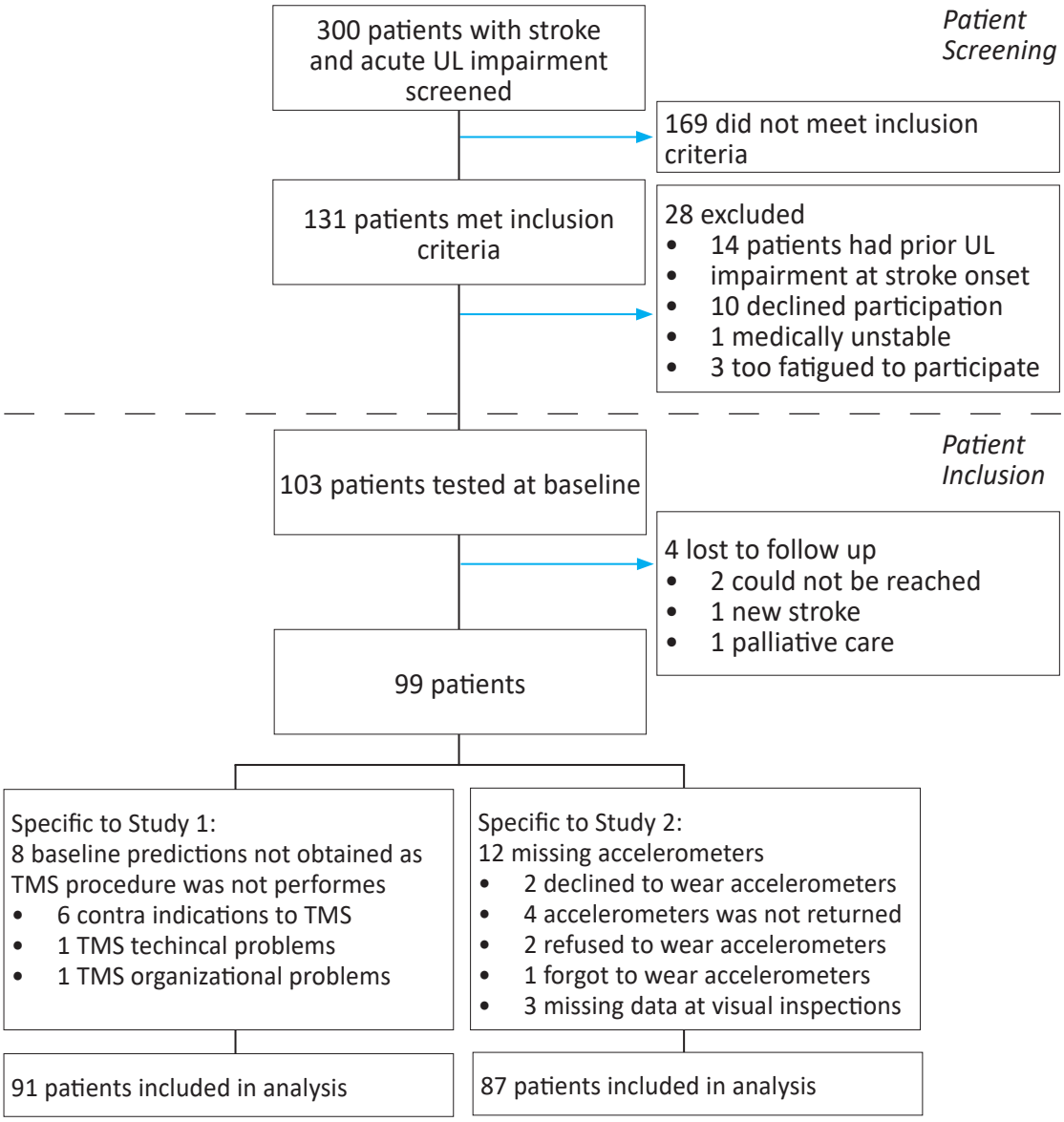
Results

Study I & II

Three hundred patients with stroke and UL impairment within the first days of stroke were screened for inclusion. The majority of the 169 patients who did not meet inclusion criteria were admitted later than 2 weeks after stroke, their UL impairment had already remitted, and/or they were cognitively not able to participate. Of the 131 patients who met the inclusion criteria 28 were excluded, mainly because of prior UL impairment impeding the potential for complete UL recovery. A total of 103 patients were included in the longitudinal study. Of these, 91 patients were available for follow-up and included in Study I and 87 were included in Study II, see Figure 5, next page.

Demographic characteristics and stroke details for all included patients are displayed in Table 2, page 45. The specific populations included in Study I & 2 differ only slightly and can be seen in Table 3, page 46 and Table 7, page 52.

Figure 5. Flowchart of Patients Included



Source: Modified from Study I[89]

Table 2. Demographic Characteristics of Included Patients (n=103)

Age, years, mean (SD, min-max)	64 (10, 44-91)
Sex, female/ male, n%	44 (42.7%) / 59 (57.3%)
Days since stroke, mean (SD, min-max)	13 (2, 10-18)
Stroke type, ischaemic/ haemorrhagic, n (%)	82 (79.6%) /21 (20.4%)
Side of paresis, left/ right n (%)	59 (57.3%) / 44 (42.7%)
Premorbid dominant hand left/ right, (n=102) n (%)	14 (13.7%) / 88 (86.3%)
Dominant UL affected, n (%)	48 (46.6%)
Stroke confirmed on imaging, n (%)	101 (98.1%)
Stroke location	
Cortical (internal capsule/ middle cerebral artery/ frontal lobe) n (%)	46 (44.7%)
Subcortical (cerebellum/ thalamus/ basal ganglia / corona radiata) n (%)	51 (49.5%)
Brainstem (pons/ medulla)	5 (4.9%)
Thrombolisis[1], n (%)	34 (41%)
Thrombectomy[1], n (%)	18 (22%)
Premorbid able to walk (+/- walking aid) (n=102), n (%)	101 (99.0%)
Premorbid living in own home, n (%)	103 (100.0%)
First stroke, n (%)	93 (90.3%)
Co-morbidity present, n (%)	75 (72.8%)
Hypertension, n (%)	51 (49.5%)
Coronary artery disease, n (%)	21 (20.4%)
Diabetes, n (%)	10 (9.7%)
Other neurological disease(s), n (%)	4 (3.9%)
Current smoker (n=89), n (%)	31 (35%)
BMI, (n=94) median (IQR, min-max)	27 (24-29, 16-46)
NIHSS score[2], (n=99) median (IQR, min-max)	8 (6-13, 1-22)
FIM score[3], (n=98) median (IQR, min-max)	73 (50- 85, 24-117)
FIM motor score, (n=98) median (IQR, min-max)	49 (32-57, 13-86)
FIM cognitive score, median (IQR, min-max)	24 (19,-30, 10-34)

For all variables, the number of participants was (n) = 103 unless otherwise stated. SD: Standard deviation. IQR: Inter quartile range. [1]Stroke thrombolysis/ thrombectomy rates were calculated for patients with ischemic stroke only. [2]NIHSS: National Institute of Health Stroke Scale. [3]FIM: Functional Independence Measure.

Table 3. Baseline Assessments of All Patients Included (n=103)	
FMA[1], median (IQR, min-max)	38 (11-53. 0-66)
ARAT[2], median (IQR, min-max)	15 (3-39. 0-57)
SAFE[3], median (IQR, min-max)	5 (2-8, 0-9)
MEP[4] not present (n=40), n (%)	12 (30%)
Shoulder subluxation present, (n=102), n (%)	19 (18.6%)
Light touch affected, n (%)	48 (46.6%)
Proprioception affected (n=101), n (%)	27 (26.7%)
Bilateral stimulation affected (n=100), n (%)	30 (30.0%)
Two point stimulation affected (n=99), n (%)	50 (51%)
UL pain present, n (%)	31 (30.1%)
UL pain intensity, median (IQR, min-max)	0 (0-4, 0-10)
Neglect present (n=100), n (%)	24 (24.0%)
FAC[5], median (IQR, min-max)	1 (0-4, 0-5)

For all variables, the number of participants was (n) = 103 unless otherwise stated. IQR: Inter quartile range. [1]FMA: Fugl-Meyer Motor Assessment Upper Extremity Score. [2]ARAT: Action Research Arm Test. [3]SAFE: Shoulder Abduction Finger Extension. [4]MEP: Motor-evoked Potentials. [5]FAC: Functional Ambulation Categories.

Specific for Study I

A total of 91 patients were eligible for Study I (see Figure 5 flow diagram for details). Characteristics at baseline and baseline assessments for the patients included in Study I are displayed in Table 4. The 12 patients not included were not statistically significant different for any of the demographic characteristics or stroke details. However, they were statistically significant different for FMA, ARAT and SAFE score at baseline. This is expected and reflects that the majority of patients not included in Study I were those with more severe UL impairments and a SAFE score below 5, excluded due to not having TMS performed, which prevented the obtainment of a baseline prediction (see Figure 5 flow diagram for details).

Table 4. Demographic Characteristics and Baseline Assessments for Study I

	Included Patients (n= 91)	Excluded Patients (n=12)	P-value
Age, years, mean (SD)	64.0 (10.6)	66.3 (7.8)	0.46
Sex, female/ male, n%	39 (43%) / 52 (57%)	5 (42%)/ 7 (58%)	0.94
Days since stroke, mean (SD)	13.4 (1.6)	13.8 (1.6)	0.35
Stroke type, ischemic/ haemorrhagic, n (%)	73 (80%) / 18 (20%)	9 (75%) / 3 (25%)	0.67
Side of paresis, left/ right n (%)	53 (58%) / 38 (42%)	6 (50%)/ 6 (50%)	0.59
Dominant UL affected, n (%)	42 (46%)	6 (50%)	0.80
Stroke confirmed on imaging, n (%)	90 (99%)	11 (92%)	0.09
Stroke location			
Cortical, n (%)	43 (47%)	3 (25%)	0.14
Subcortical, n (%)	45 (49%)	6 (50%)	0.97
Brainstem, n (%)	4 (4%)	1 (8%)	0.55
Thrombolisis[1], n (%)	32 (44%)	2 (22%)	0.21
Thrombectomy[1], n (%)	16 (22%)	2 (22%)	0.98
Premorbid able to walk (+/- walking aid), n (%)	89 (99%)	12 (100%)	0.71
Premorbid living in own home, n (%)	91 (100%)	12 (100%)	
First stroke, n (%)	83 (91%)	10 (83%)	0.39
Co-morbidity present, n (%)	65 (71%)	10 (83%)	0.38
Hypertension, n (%)	43 (47%)	8 (67%)	0.21
Coronary artery disease, n (%)	17 (19%)	4 (33%)	0.24
Diabetes, n (%)	7 (8%)	3 (25%)	0.06
Other neurological disease(s), n (%)	3 (3%)	1 (8%)	0.40
Current smoker, n (%)	30 (38%)	1 (9%)	0.06
BMI, median (IQR)	27 (24- 29)	29 (25- 30)	0.24
NIHSS[2], median (IQR)	9 (6- 13)	8 (6-10)	0.53
FIM[3], median (IQR)	74 (50- 89)	68 (50- 77)	0.32
FIM motor score, median (IQR)	49 (32- 58)	42 (33- 52)	0.37
FIM cognitive score, median (IQR)	24 (19- 30)	24 (20- 28)	0.58
FMA[4], median (IQR,)	42 (16- 53)	11 (4- 29)	0.017*
ARAT[5], median (IQR)	21 (4- 41)	3 (0- 13)	0.015*
SAFE[6], median (IQR)	5 (2- 8)	3 (0- 4)	0.013*
MEP[7] not present, n (%)	12 (13%)	0 (0%)	0.44
Shoulder subluxation present, n (%)	16 (18%)	3 (25%)	0.55
Light touch affected, n (%)	42 (46%)	6 (50%)	0.80
Proprioception affected, n (%)	24 (27%)	3 (25%)	0.89
Bilateral stimulation affected, n (%)	27 (31%)	3 (25%)	0.69
Two point stimulation affected, n (%)	49 (54%)	5 (42%)	0.43
UL pain present, n (%)	28 (31%)	3 (25%)	0.68
UL pain intensity, median (IQR)	0 (0- 4)	0 (0- 2)	0.63
Neglect present, n (%)	21 (24%)	3 (27%)	0.79
FAC[8], median (IQR)	1 (0- 4)	0 (0- 3)	0.38

*The included and excluded patients were significantly different. SD: Standard deviation. IQR: Inter quartile range. [1]Stroke thrombolysis/ thrombectomy rates were calculated for patients with ischaemic stroke only. [2]NIHSS: National Institute of Health Stroke Scale Score. The NIHSS score for 12 patients who were MEP- was a median of 13 (IQR 7-15, min 9, max 21). [3]FIM: Functional Independence Measure. [4]FMA: Fugl-Meyer Motor Assessment Upper Extremity score. [5]ARAT: Action Research Arm Test. [6]SAFE: Shoulder Abduction Finger Extension. [7]MEP: Motor-evoked potentials. MEP was assessed in 40 patients with a SAFE score below 5 and assumed present in all patients with a SAFE score ≥5. MEP was assessed a mean of 13.4 days after stroke (SD 1.7, min 11, max 18). [8]FAC: Functional Ambulation Classification. *Source: Modified from Study I[89].*

Prediction of future UL function

Information on SAFE score, MEP status, age and NIHSS score was used to predict future UL function.[89]

At baseline the mean SAFE score was 5 (SD 2.8, min 0, max 9) and was obtained 13.4 days (SD 1.6, min 10, max 18) after stroke. Twelve patients were MEP- and 26 were MEP+. For 38 of 91 patients (42%) the SAFE score was < 5 and MEP status was established.[89] Fifty patients (55%) had the prediction *Excellent UL function*, 29 (32%) had the prediction *Good UL function*, three patients (3%) had the prediction *Limited UL function*, and nine patients (10%) had the prediction *Poor UL function* (Table 5).[89]

Follow-up assessments on ARAT were obtained a mean of 91 days after stroke (SD 3.8, min 84, max 99). At follow-up the median ARAT score was 50 (IQR 33-55, min 0, max 57). The median within group improvement in ARAT scores was 17 (IQR 3 - 27, $P < 0.001$).[89]

At follow-up patients were grouped according to the actually achieved ARAT score. Forty-four patients were included in the category *Excellent UL function*, 22 in *Good*, 13 in *Limited*, and 12 patients in *Poor UL function* (Table 5). The ARAT score was a median of 55 (IQR 54-56, min 51, max 57) in the category *Excellent*, a median of 42 (IQR 39-49, min 24, max 50) in the category *Good*, a median of 31 (IQR 24-32, min 17, max 33) in the category *Limited*, and a median 0 (IQR 0-2, min 0, max 6) in the category *Poor UL function.* [89]

UL prediction accuracy

The overall CCR was 60% (95% CI 51-71).[89] Twenty-six of 91 patients (29%) did not achieve as high UL function as predicted and hence the prediction was too optimistic. On the contrary, for 10 of 91 patients (11 %) the prediction was too pessimistic and actual UL function at 3 months exceeded the predicted. Twenty-eight

of the 36 of patients (78%) for whom the prediction was inaccurate achieved an actual outcome category adjacent to the predicted category, e.g. they were predicted *Limited*, but ended up in the outcome category *Good*.

Table 5. Predicted and Actual ARAT Categories and Agreement Between Them

Predicted ARAT category at baseline	Actual ARAT outcome category at 3 months				
	Excellent	**Good**	**Limited**	**Poor**	**Total, n (%)**
Excellent	37	10	3	0	50 (55%)
Good	7	10	8	4	29 (32%)
Limited	0	1	1	1	3 (3%)
Poor	0	1	1	7	9 (10%)
Total	44 (48%)	22 (24%)	13 (14%)	12 (13%)	91 (100%)

Green: Patients for whom the outcome category was equivalent to the predicted category (n=55)
Yellow: Patients for whom the outcome category was adjacent to the predicted category (n=28)
Orange: Patients for whom the outcome category was two categories away from the predicted category (n=8)
Red: Patients for whom the outcome category was three categories away from the predicted category (n=0)
Source: Replicated from Study I[89]

Within the four categories, CCR was a maximum of 78% (95% CI 43-95%) for patients predicted *Poor UL function* followed by 74% (95% CI 60-84%) for those predicted *Excellent UL function*. CCR was 35% (95% CI 20-53%) for patients predicted *Good* and 33% (95% CI 4-85%) for those predicted *Limited UL function* (Table 6).

For the 53 patients with a SAFE ≥ 5, CCR was 74% (95% CI 62-86%). For the 38

patients with a SAFE < 5, CCR was 42% (95% CI 26-58%)(Table 6).[89] This low CCR was primarily caused by 26 patients who were predicted to have *Good UL function* based on MEP+ (Table 5, previous page). Many of these patients ended in another actual outcome category at 3 months (Table 5). In contrast to this, patients who were MEP- and predicted a *Poor UL function* generally also ended up in the predicted outcome category (Table 5).[89]

Table 6. Accuracy of the Prediction Algorithm for UL Function [89]

	CCR (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	CCR for SAFE ≥ 5 (95% CI)	CCR for SAFE < 5 (95% CI)
Overall (n= 91)	60% (50; 71)				
Excellent (n= 44)	74% (60; 84)	84% (70; 93)	72% (57; 84)	74% (62; 86)	
Good (n= 22)	35% (20; 53)	46% (24; 68)	73% (60; 83)		
Limited (n= 13)	33% (4; 85)	8% (0; 36)	97% (91; 100)		42% (26; 58)
Poor (n= 12)	78% (43; 95)	58% (28; 85)	98% (91: 100)		

CCR: Correct Classification Rate; n= number of patients in outcome category
Source: Replicated from Study I[89]

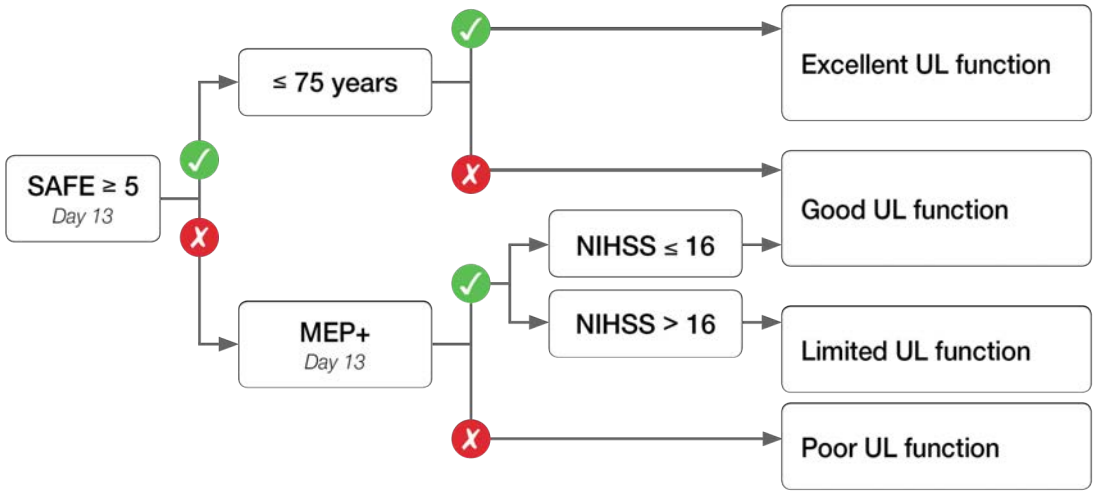
CART analysis

A CART analysis was conducted to examine if modifications of the included components could improve prediction accuracy when the PREP2 algorithm was obtained two weeks after stroke. The decision tree developed by CART analysis had an overall CCR of 66% (95% CI 56-76) (Figure 6).[89]

A total of 89 of 91 patients were included in the CART, as the NIHSS score was not available for two patients. Based on CART, the SAFE score was found to be the most important predictor. For patients with a SAFE ≥ 5 a prediction of either *Good* or *Excellent UL function* was made, based on age. For patients with a SAFE score < 5 who were MEP+, the NIHSS score was needed to differentiate between *Good* and *Limited UL function*. Patients with a SAFE < 5 who were MEP- were predicted a future *Poor UL function.*[89]

CCR for each of the four categories was 67% (95% CI 37-87%) for *Poor,* 60% (95% CI 20-90%) for *Limited,* 45% (95% CI 28-63%) for *Good,* and 80% (95% CI 66-89%) for the category *Excellent.* For patients with a SAFE ≥ 5, CCR was 78% (95% CI 67-89%); and for patients with a SAFE < 5, CCR was 50% (95% CI 34-66%).

Figure 6. CART Model for Prediction of UL Function



Available for the CART analysis were the PREP2 components: SAFE score, MEP status, age, and the NIHSS score.
Source: Replicated from Study I[89]

Table 7. Demographic Characteristics and Baseline Assessments for Study II

	Included patients (n=87)	Excluded patients (n=16)	P-value
Age, mean (SD)	64.9 (10.5)	60.8 (8.8)	0.15
Sex, female/ male, n (%)	35 (40) / 52 (60)	9 (56)/ 7 (44)	0.23
Days since stroke, mean (SD)	13.3 (1.6)	13.9 (1.5)	0.16
Stroke type, ischemic/ hemorrhagic, n (%)	70 (80) / 17 (20)	12 (75)/ 4 (25)	0.62
Side of paresis, left/right, n (%)	47 (54) / 40 (46)	12 (75)/ 4 (25)	0.12
Dominant UL affected, n (%)	43 (49)	5 (31)	0.18
Stroke confirmed on imagining, n (%)	86 (99)	15 (94)	0.17
Stroke location			
Cortical, n (%)	41 (47)	5 (31)	0.24
Subcortical, n (%)	41 (47)	10 (63)	0.26
Brainstem, n (%)	5 (6)	0 (0)	0.33
Thrombolisis[1], n (%)	29 (41)	5 (42)	0.99
Thrombectomy[1], n (%)	17 (24)	1 (8)	0.22
Premorbid able to walk (+/-walking aid), n (%)	86 (100)	15 (94)	0.02*
Premorbid living in own home, n (%)	87 (100)	16 (100)	
First stroke, n (%)	79 (91)	14 (88)	0.68
Co-morbidity present, n (%)	61 (70)	14 (88)	0.15
Hypertension, n (%)	40 (46)	11 (69)	0.09
Coronary artery disease, n (%)	16 (18)	5 (31)	0.24
Diabetes, n (%)	7 (8)	3 (19)	0.18
Other neurological disease, n (%)	3 (3)	1 (6)	0.59
Current smoker, n (%)	26 (35)	5 (33)	0.89
BMI, median (IQR),	26 (23- 29)	29 (28- 33)	0.005*
FIM[2], median (IQR)	72 (49- 85)	75 (57- 96)	0.66
FIM motor score, median (IQR)	48 (30- 57)	50 (43- 70)	0.49
FIM cognitive score, median (IQR)	24 (19- 29)	24 (15- 30)	0.68
Assessments at baseline			
FMA[3], median (IQR)	40 (14- 53)	25 (11- 49)	0.64
ARAT[4], median (IQR)	17 (3- 39)	13 (4- 39)	0.86
SAFE[5], median (IQR)	5 (2- 8)	5 (2- 8)	0.81
MEP[6] not present, n (%)	9 (11)	3 (19)	0.28
Shoulder subluxation present, n (%)	18 (21)	1 (7)	0.20
Light touch affected, n (%)	41 (47)	7 (44)	0.80
Proprioception affected, n (%)	24 (28)	3 (20)	0.52
Bilateral stimulation affected, n(%)	26 (31)	4 (27)	0.76
Two point stimulation affected, n (%)	43 (51)	8 (50)	0.83
UL pain present, n (%)	26 (30)	5 (31)	0.91
UL pain intensity, median (IQR)	0 (0- 4)	0.0 (0- 3)	0.97
Neglect present, n (%)	23 (27)	1 (7)	0.09
FAC[7], median (IQR)	1 (0- 4)	1.5 (0- 4)	0.57

*The included and excluded patients were significantly different. SD: Standard deviation. IQR: Inter quartile range. [1]Stroke thrombolysis/ thrombectomy rates were calculated for patients with ischaemic stroke only. [2]FIM: Functional Independence Measure. [3]FMA: Fugl-Meyer Motor Assessment Upper Extremity score. [4]ARAT: Action Research Arm Test. [5]SAFE: Shoulder Abduction Finger Extension. [6]MEP: Motor-evoked potentials. MEP was assessed in 40 patients with a SAFE score below 5 and assumed present in all patients with a SAFE score ≥5. [7]FAC: Functional Ambulation Classification. *Source: Modified from Study II (unpublished)*

Specific for Study II

A total of 87 patients were eligible for Study II. (see Figure 5 for details). Characteristics at baseline and baseline assessments for the patients included are displayed in Table 7. The median FMA score at baseline was 17 (IQR 14- 53, min 0 max 66), reflecting a broad range of UL impairment.

In Study II, the 16 patients not included were not statistically significant different for any of the baseline assessments displayed in Table 7. Demographic characteristics and stroke details were statistically significant different only for premorbid ability to walk and for BMI.

Upper limb use

The non-paretic unilateral UL activity was a median 2.1 hours (IQR 1.4- 2.8) and three times higher than the paretic unilateral UL activity. Bimanual UL activity was 3.0 hours (IQR 1.9- 4.0) and total UL activity was 5.8 (IQR 4.8- 7.2). The use ratio was 0.7 (IQR 0.6- 0.9) (Table 8, next page).

When accelerometer parameters were examined according to the severity of initial UL impairment, non-paretic unilateral activity decreased and paretic UL activity increased with decreasing impairment. Bimanual activity, total UL activity, use ratio and bilateral vector magnitude also increased with improving UL function. The magnitude ratio was a median of -3.8 for patients with severe UL impairment, reflecting primarily non-affected UL use, while it was -1.0 for patients with mild UL impairment, reflecting a more equal contribution of both limbs to an activity.

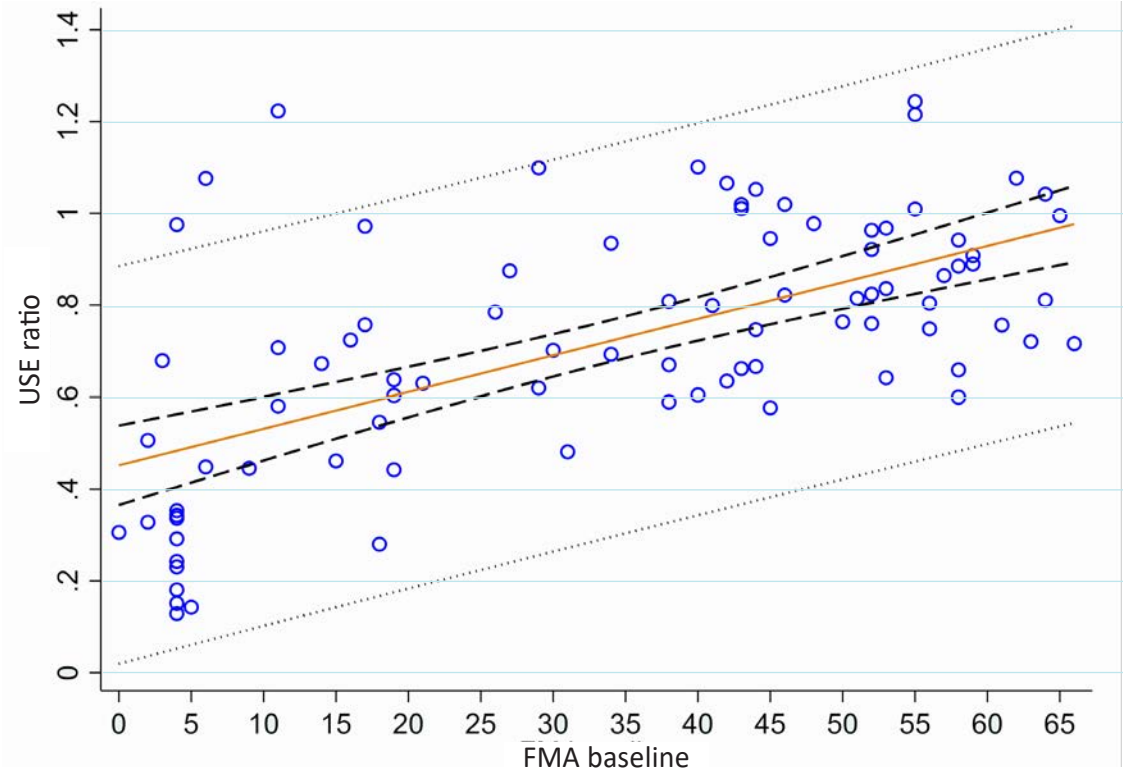
Table 8. Accelerometry Outcomes at Three Months after Stroke for all Patients and in Accordance with FMA at Baseline

	All patients (n = 87)	FMA Severe (score 0-22) (n = 32)	FMA moderate (score 23-50) (n = 28)	FMA Mild (score 51-66) (n = 27)
Non-paretic unilat. UL activity, hours, median (IQR)	2.1 (1.4; 2.8)	2.8 (2.2; 3.4)	1.7 (1.4; 2.6)	1.6 (1.2; 2.0)
Paretic unilat. UL activity, hours, median (IQR)	0.7 (0.4; 1.0)	0.4 (0.2; 0.7)	0.9 (0.5; 1.3)	0.9 (0.7; 1.4)
Bimanual UL activity, hours, median (IQR)	3.0 (1.9; 4.0)	1.7 (0.9; 3.2)	3.3 (2.5; 4.2)	3.3 (2.4; 4.3)
Total UL activity, hours, median (IQR)	5.8 (4.8; 7.2)	5.5 (4.5; 6.0)	6.4 (5.0; 7.6)	6.0 (4.7; 7.4)
Use ratio, median (IQR)	0.7 (0.6; 0.9)	0.5 (0.3; 0.7)	0.8 (0.7; 1.0)	0.9 (0.8; 1.0)
Bilateral magnitude, median (IQR)	110.7 (93.5; 127.5)	93.8 (81.8; 112.1)	116.4 (100.7; 140.6)	119.1 (107.5; 133.2)
Magnitude ratio, median (IQR)	-1.9 (-3.2; -0.4)	-3.8 (-4.7; -2.3)	-1.7 (-2.4; -0.1)	-1.0 (-1.6; -0.1)

Use ratio: total hours of paretic UL use divided by total hours of non-paretic use.
Bilateral magnitude: Intensity of activity across both ULs for each second of activity
Magnitude ratio: The natural log of the paretic UL vector magnitude divided by the vector magnitude of the non-paretic UL the natural log of the paretic UL vector magnitude divided by the vector magnitude of the non-paretic UL.
FMA: Fugl-Meyer Motor Assessment Upper Extremity
Source: Replicated from Study II (unpublished)

A linear regression (Table 9, Model 1) demonstrated, that the FMA score at baseline was a statistically significant predictor of use ratio at three months with a β of 0.008 (95% CI 0.006- 0.010), P<0.0001. FMA explained 0.38 of the variation in use ratio. The association between FMA score at baseline and use ratio at 3 months is displayed in Figure 7.

Figure 7. Association between FMA at Baseline and Use Ratio at Three Months After Stroke



The solid red line is the best-fitted prediction line of the association between FMA at baseline and use ratio at three months. The 95% confidence interval is displayed with dashed lines and the wider 95% prediction interval is displayed with the dotted lines. With 95% accuracy, the true mean use ratio for a given FMA score will fall within the 95% CI. The PI is an estimate of the interval in which a future observation of UL use ratio for an individual patient will fall, with 95% probability, given what has already been observed.
FMA: Fugl-Meyer Motor Assessment Upper Extremity
Source: Replicated from Study II (unpublished)

When all secondary variables were entered in a multiple regression model (Table 9, Model 2), data from 74 patients were included, as data for one or more variables were missing for 13 patients. R^2 improved to 0.55, an improvement of 0.17, reflecting that the model now explained a higher percentage of the use ratio. The equation line for use ratio was:

Use ratio= 0.308 +0.222 * mep -0.128 * neglect +0.077 * dominant UL affected +0.024 * twopd +0.0004 * fim -0.046 * gender -0.005 * pain +0.006 * FMA

The statistically significant predictors were FMA, MEP status and neglect. The β-slope for FMA was 0.006 (95% CI 0.003- 0.009, P=0.000*) and for every FMA score higher a patient was at baseline, use ratio would be a mean of 0.006 higher. With 95% accuracy, the true mean would be contained in the interval of 0.003-0.009. The β-coefficient for MEP status was 0.222 (95% CI 0.069- 0.376, P=0.005*), and a patient who was MEP+ at baseline achieved a use ratio that was 0.222 higher than a patient who was MEP-. The β-coefficient for neglect was -0.128 (95% CI 0.240- 0.016, P=0.025*), thus a patient who had neglect achieved a use ratio that was 0.128 lower compared to a patient without neglect. The 95% PI for the expected use ratio in model 2 was ±0.348.

In Model 3, the biomarker MEP was removed and the adjusted R^2 decreased to 0.458, which was 0.09 lower than in model 2 with MEP included. The 95% PI for the expected use ratio in model 3 was ±0.397.
The univariate linear regressions of each of the potential individual predictors showed that the strongest secondary predictor of use ratio was MEP status, followed by FIM, neglect, twopd, dominant side and gender. Pain was not a significant predictor of use ratio (Table 9).

Table 9. Regression Models to Examine Prediction of Use Ratio

	Predictors	Constant	β- coefficient	p-value	95% confidence interval	Adjusted model R^2	SD
Model 1- univariate regression (n =87)						0.376	0.213
	FMA score		0.008	0.000*	0.006; 0.010		
	Constant	0.452		0.000*	0.365; 0.539		
Model 2- multiple regression (n=74)						0.548	0.178
	FMA score		0.006	0.000*	0.003; 0.009		
	MEP +		0.222	0.005*	0.069; 0.376		
	Neglect present		-0.128	0.025*	-0.240; 0.016		
	Dominant side affected		0.070	0.108	-0.018; 0.157		
	Twopd affected		0.024	0.614	-0.071; 0.120		
	FIM score		0.000	0.736	-0.002; 0.003		
	Male		-0.046	0.306	-0.136; 0.043		
	Pain score		-0.005	0.495	-0.020; 0.010		
	Constant		0.380		0.086; 0.530		
Model 3 Multiple regression without MEP biomarker (n= 80)						0.458	0.203
	FMA score		0.007	0.000*	0.004; 0.009		
	Neglect present		-0.115	0.059	-0.234; 0.004		
	Dominant side affected		0.014	0.030*	0.010; 0.200		
	Twopd affected		-0.042	0.409	-0.144; 0.059		
	FIM score		0.000	0.635	-0.002; 0.003		
	Male		-0.062	0.216	-0.161; 0.037		
	Pain score		-0.000	0.983	-0.016; 0.016		
	Constant		0.488		0.268; 0.708		
Univariate regressionsof all secondary variables							
N=81	MEP +		0.405	0.000*	0.246; 0.565	0.235	
N=85	Neglect present		-0.222	0.001*	-0.346; 0.098	0.123	
N=87	Dominant side affected		0.130	0.025*	0.017; 0.242	0.047	
N=84	Twopd affected		-0.140	0.018*	-0.255; 0.025	0.055	
N=83	FIM score		0.005	0.000*	0.003; 0.008	0.180	
N=87	Male		-0.123	0.036*	-0.238; -0.008	0.039	
N=87	Pain score		-0.007	0.465	-0.027; 0.013	0.006	

CI: Confidence interval. SD: Standard deviation. *The β- coefficient was statistically significant. In 34 of 40 possible patients with a SAFE < 5, MEP status was established and in six patients it was not. In 57 patients with a SAFE ≥ 5, MEP+ was assumed. FMA: Fugl-Meyer Motor Assessment Upper Extremity. MEP: Motor-evoked potentials. Twopd: Two-point discrimination. FIM: Functional Independence Measure. *Source: Replicated from Study II (unpublished)*

Characteristics of patients who did not achieve normal use ratio

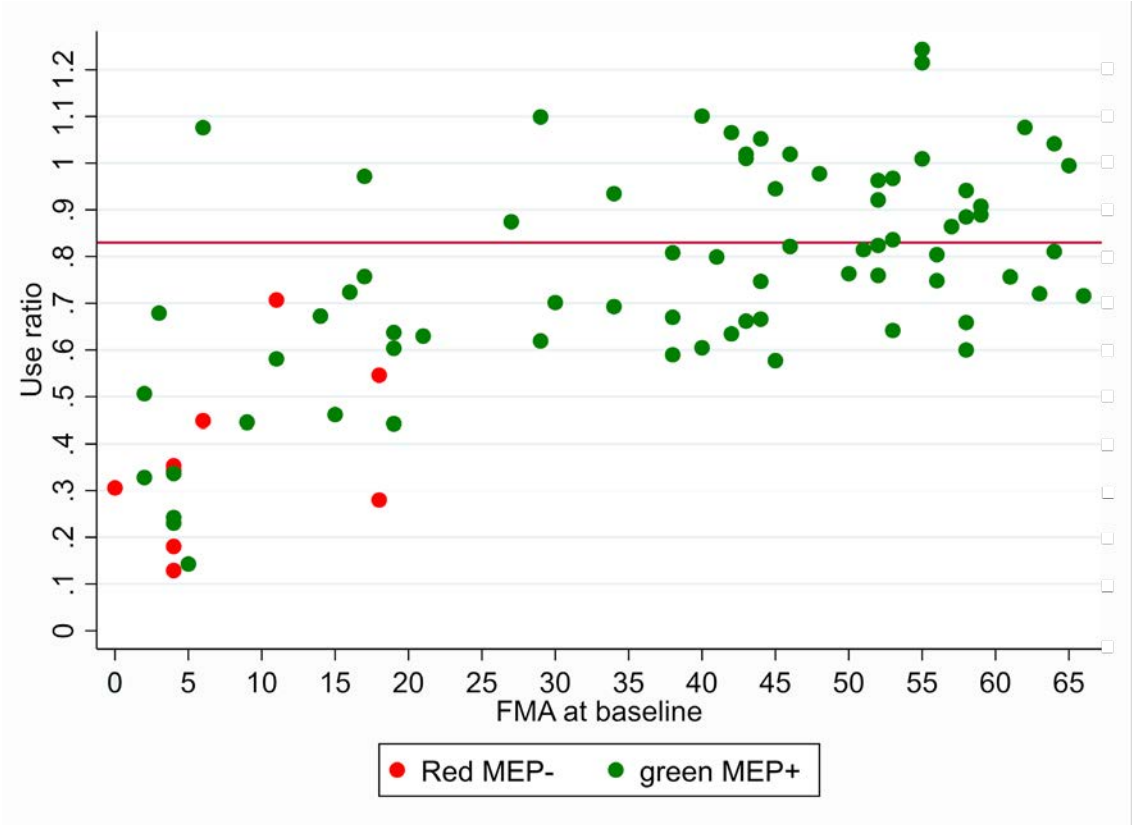
Use ratio was dichotomized at a threshold of 0.83 and patients with a use ratio ≥0.83 were classified as having a normal use ratio, and patient with a use ratio < 0.83 as having non-normal use ratio. A total of 30 (34%) patients were classified as having normal use ratio and 57 (66%) as having non-normal use ratio at three months.

Visual inspection revealed that none of the nine patients with MEP- achieved a normal use ratio (Figure 8). Accordingly, 22 of the 23 patients with neglect did not achieve normal use ratio (Figure 9, page 60). Two patients had MEP- and also neglect, seven patients had MEP- only, and 21 patients had neglect only.

For the remaining patients, all with MEP+ and without neglect, multivariate logistic regression was applied to assess how well the variables FMA, dominant side, twopd, and FIM could predict non-normal use ratio. Data from 48 of 57 possible patients were included, as nine patients had missing data for one of the variables. Significant predictors of non-normal use ratio were FMA and dominant UL affected. The β for FMA was 0.928 (95% Ci 0.890- 0.980, P= 0.007*), and β for dominant UL affected was 0.113 (95% CI 0.023- 0.570, P=0.008*). This means that the odds for not achieving a normal use ratio decreased with 0.07 (7%) for each FMA score higher at baseline. For patients who had their dominant UL affected the odds of not achieving normal use was 0.89 (89%) lower.

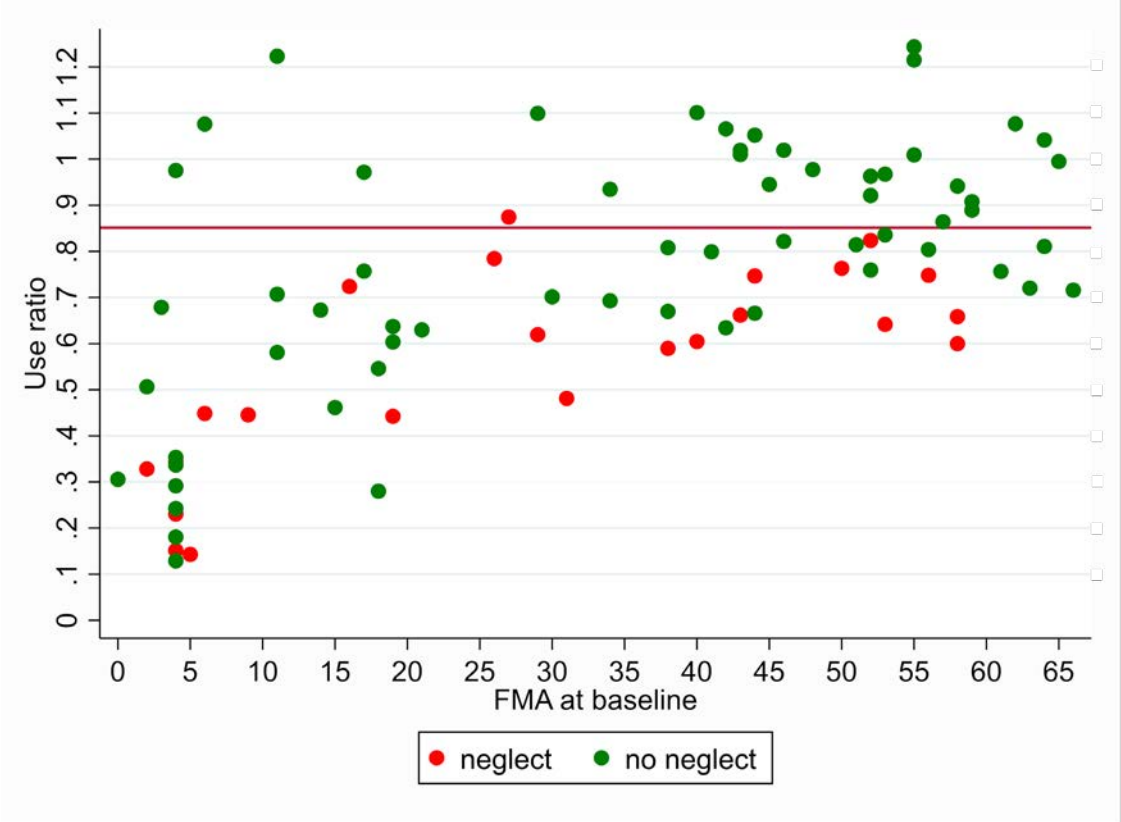
FIM and twopd did not significantly contribute to the prediction of not achieving normal use ratio (P=0.757 and P= 0.079). The ROC based on the multivariate logistic regression (Figure 10, page 61) revealed an AUC of 0.84 (95% CI 0.73-0.96). The optimal cut point for prediction of non-normal use ratio for patients with MEP and without neglect was 0.55 with a sensitivity of 0.80 (95% CI 0.61- 0.91) and a specificity of 0.83 (95% CI 0.53-0.93).

Figure 8. Association between MEP status at Baseline and Use Ratio three months After Stroke



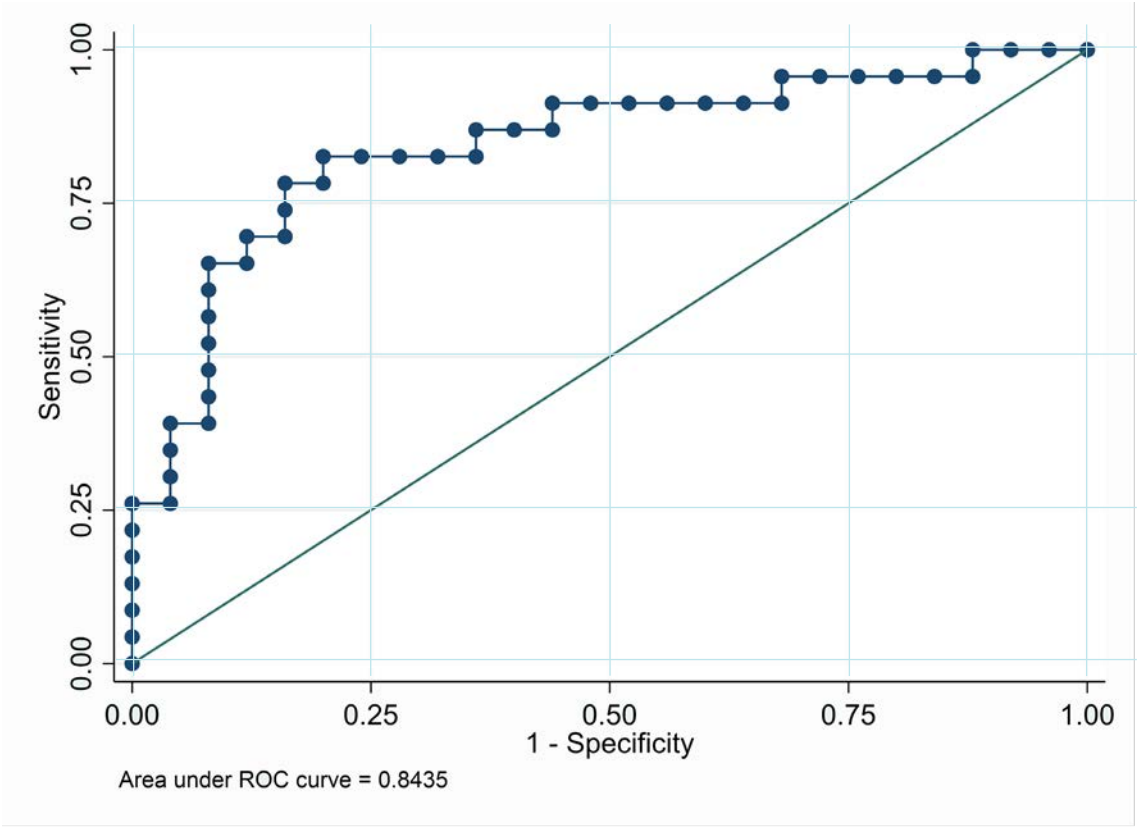
Horizontal red line: Threshold for normal use ratio. MEP status for a total of 81 patients. None of the nine patients who were MEP- achieved a normal use ratio. Of the remaining 72 patients, 44 patients did not and 28 patients did achieve a normal use ratio.
Source: Replicated from Study II (unpublished)

Figure 9. Association Between Neglect at Baseline and Use Ratio Three Months After Stroke



Neglect was examined in a total of 85 patients and found present in 23 patients.
Almost all, 22 of 23 patients with neglect did not achieve normal use ratio.
Among the 62 patients without neglect, 28 did and 34 did not achieve a normal use ratio.
Source: Replicated from Study II (unpublished)

Figure 10. ROC of Sensitivity and Specificity for Prediction of not Achieving a Normal Use Ratio



Prediction of achieving non-normal use ratio for patients who had MEP+ and were without neglect. The ROC was based on a multivariate logistic regression with the variables FMA, dominant side, twopd and FIM. The AUC was 0.84 (95% CI 0.73-0.96%). If a cut point of 0.55 was chosen, the odds of achieving a non-normal use ratio could be predicted with a sensitivity of 0.80 (95% CI 0.61- 0.91) and a specificity of 0.83 (0.63-0.93).
Source: Replicated from Study II (unpublished)

Study III

Four focus group interviews were conducted from January to April 2019 and lasted from 68 to 90 minutes. In the pilot focus group, three PTs participated. All had clinical experience in neurorehabilitation and were engaged in either a Master's or PhD study. In the succeeding three interviews, all participants were employed at neurorehabilitation wards. The number of participants in each focus group corresponded to the size of the rehabilitation unit. Characteristics of participants are displayed in Table 10.

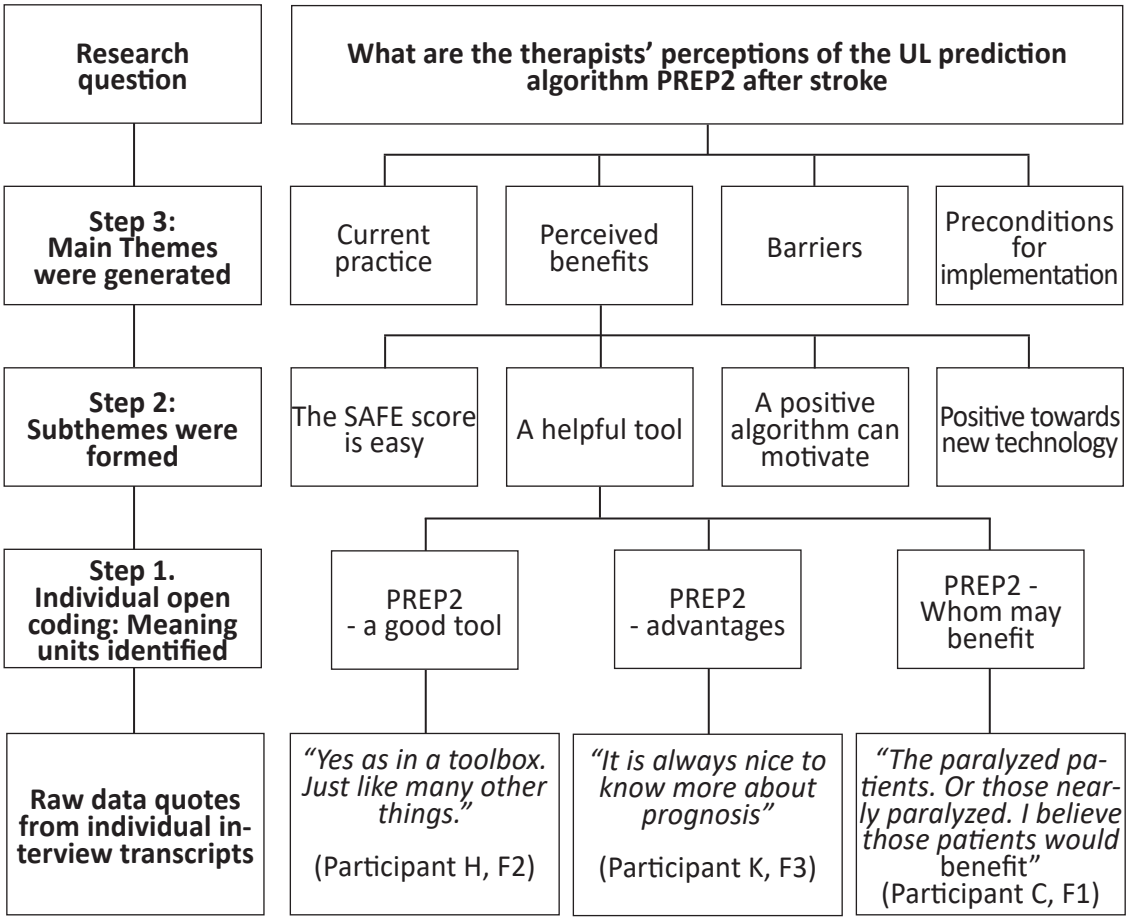
Table 10. Characteristics of Focus Group Participants

Group	Pilot focus group (F1)	Focus group 1 (F2)	Focus group 2 (F3)	Focus group 3 (F4)
Number of participants	3	6	4	3
Profession	3 PT	3 PT; 3 OT	2 PT; 2 OT	1 PT; 2 OT
Assigned position	1 specialist	2 specialists, 1 student advisor		
Educational level	2 Master; 1PhD	5 Bachelor; 1 Master	4 Bachelor	3 Bachelor
Gender	2 F; 1 M	6 F	4 F	3 F
Average years since graduation (range)	15 (12-18)	12 (5-17)	20 (13-23)	17 (9-23)
Average years of experience in neurorehabilitation (range)	11 (10-18)	10 (3-17)	17 (13-20)	12 (2-18)
Current unit of employment	Unit 1 and Acute Neurology	Unit 1	Unit 2	Unit 3
Anonymized initial of participant when quoted	A; B; C	D; E; F; G; H; I	J; K; L; M	N; O; P

Source: Replicated from Study III (unpublished)

Across the interviews four main themes, considered of great importance to the participants and relevant for implementing prediction algorithms, emerged (see Figure 11).

Figure 11. Diagram Showing Examples of Theme Formation



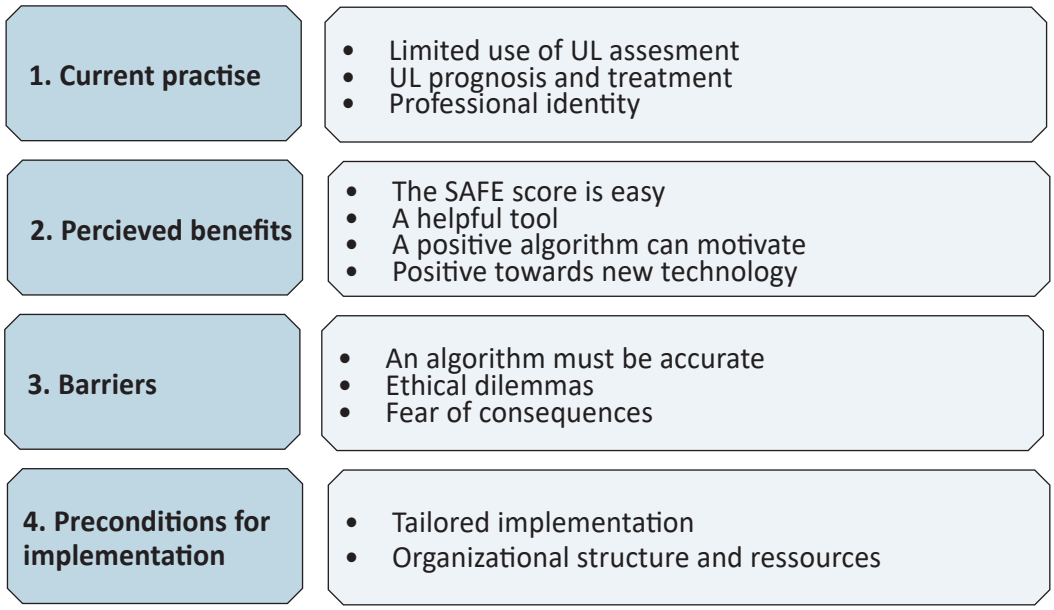
Source: Replicated from Study III (unpublished)

To document and consolidate results and increase the trustworthiness of Study III, quotations were used to display from what kind of original data the four categories were derived.[102-104] Where cited, the context was quoted in parentheses. In accordance with Table 10 participant E from focus group 2 would be quoted as (participant E, F2). To ensure credibility, a participant from each focus group read

the interview transcripts and the interpretation of the results.[91] The participants recognized themselves and provided further nuances to the results.

Results and quotations related to the four main themes: current practice, perceived benefits, barriers, and preconditions for successful implementation are presented below and an overview are seen in Figure 12.

Figure 12. Four Main Themes and Their Subthemes



Source: Replicated from Study III (unpublished)

1. Current practice

To know the current practice is a requirement for understanding the participants' considerations on barriers and perceived benefits. This first main theme comprised three subthemes: limited use of UL assessments, considerations on UL prognosis and treatment, and professional identity.

Limited use of UL assessments

UL prediction algorithms includes the performance of standardized assessments, e.g. the SAFE test, and information about the use of UL assessments was therefore relevant. Overall, the participants agreed that UL tests were used, but on a limited scale. Consensus existed that the UL test had to be clinically relevant for the specific patient and not a routine test used for all patients. In addition, the test had to be quick to perform and easy to administer:

> *"One has to prioritize the time to do it. So it has to make sense to do it." (participant B, F1)*

UL prognosis and treatment

A range of factors was considered important for UL recovery. Some, but not all, aligned with factors highlighted in the literature. Initial UL function and time since stroke were mentioned in all interviews, but not stressed by the participants as particularly important predictors:

> *"I think that having some function is important. We have a lot…I believe where the SAFE score is zero…because they are paralyzed… you cannot palpate any muscle activity. That has a huge importance for… whether they regain any function at all…" (participant G, F2)*

Other factors mentioned in the interviews as important for recovery were pain, sensory motor deficits, time since stroke, location of stroke, type of stroke, and initial medical treatment. Several participants mentioned the importance of past experiences, self-efficacy, motivation, and inner drive. Everyone agreed that cognition was vital, especially neglect and awareness of own disabilities.

The PREP2 algorithm includes information on age and initial score on stroke severity. However, age was not considered particularly important for UL prognosis, and only a few participants were aware of initial scores, e.g. NIHSS or SSS, performed in the acute units.

When planning UL treatment and choosing interventions, the participants took many of the same elements into account as when considering UL prognosis. Importantly, they found that the patients' individual goal should guide whether or not UL treatment was a main priority.

Professional identity

Participants in all interviews found that use of UL assessment and algorithms aligned more with the PT profession than the OT profession. The PTs often focused on limitations on impairment level, while the OTs centred on activity level:

> *"Well, if I have a patient I look for … because I am an OT… for activity limitations in relation to the use of arms and hands…because I am an OT." (participant H, F2)*

Most of the participants considered themselves experienced neuro-therapists. According to several participants, prediction algorithms would make the most sense for recently qualified therapists who may need a simple tool, while the more experienced could draw on years of experience:

> *"I believe this PREP2 is for more recently qualified therapists…a lot easier to access…because then you can draw on the cold facts: this is what we have to guide us. And they are more trained in that that than the rest of us." [the group agrees] (participant M, F3)*

2. Perceived benefits

The second main theme encompassed thoughts on how an algorithm may aid UL rehabilitation. Subthemes were: The SAFE test is easy; a helpful tool; a positive algorithm can motivate; and positive towards new technology.

The SAFE test is easy

Participants in the pilot focus group had some knowledge of UL prediction algorithms but across the other interviews knowledge was less profound. Some of the participants used the SAFE test and especially the physiotherapists considered the SAFE test easy to administer:

> *"The SAFE test is easy and quick and you can allow yourself to do it no matter what." (participant B, F1)*

Some participants found that the difference between score 2 (=limited range of motion without gravity) and score 3 (=full range of motion against gravity, but not resistance) was rather large. Despite this, the same participants considered the SAFE test appealing, because it was fast and could be performed without equipment.

> *"But that big gap…we actually discussed it…. Actually, for some patients, we would like to score 2½ [the group agrees]." (participant M, F3)*

A helpful tool

PREP2 was envisaged as a potentially helpful tool for considering prognosis and planning UL treatment. If used in combination with information from other sources, PREP2 could be used as a tool or an indicator to decide what way to go, e.g. whether to intensify UL training or instead start to train compensatory strategies:

> *"I believe an indicator is a good word. An indicator. Because it is not an answer to if they will achieve function or not …or how good that function will be. But it gives an indication. For this reason, we choose to go this way. But it does not mean that when the patient is discharged from RHN, we will write: The patient will never achieve any function. It is just a good tool." (participant F, F2)*

> *"Yes as in a toolbox. Just like many other things." (participant H, F2)*
> *"It is always nice to know more about prognosis." (participant K, F3)*

Different opinions existed on whether UL prediction algorithms would be a prognostic aid for all or only some patients. The predominant opinion was that it would be particularly relevant for patients with little or no UL function.

"The paralyzed patients. Or those nearly paralyzed. I believe those patients would benefit." (participant C, F1)

A positive algorithm can motivate

All participants agreed that an optimistic prediction could be used to motivate patients and therapists:

"Some indication… would be nice. It could be used to motivate when progression is slow and you think nothing is happening in an arm. If I could say: I KNOW if we do this exercise for the next four weeks every day, then it will come; that would motivate the patient. And me as a therapist." (participant P, F4)

Positive towards new technology

A positive attitude towards TMS and MEP was present in all four interviews. The participants found it appealing that information on MEP status could add information to UL prediction that could not be obtained by a clinical test. They believed this information would motivate both patient and therapist:

"But what I find really interesting is that you can have this.. MEP…? If there is a connection in the corticospinal tract. So you can have a SAFE below five and still expect a good function." (participant L, F3)

"There might be some people where you think they should have got some more…(training) because if we had that examination, TMS…" (participant G, F2)

3. Barriers

The third main theme concerns the participants' perceptions of the limitations of prediction algorithms and potential barriers to their implementation in clinical practice. Subthemes were: an algorithm must be accurate, ethical dilemmas, and fear of consequences.

An algorithm must be accurate

All participants agreed that an algorithm should be as accurate as possible:

> *"Definitely, definitely" [the group agrees]. (participant L, F3)*

> *"It must, of course, be very precise for us to use it." (participant O, F4)*

Disagreement existed on whether the 75% accuracy of the PREP2 algorithm was precise enough and many participants imagined that an accuracy of 75% could still be used by the team or individual therapist along with other indications and tools of prognosis. For some, a precision of 75% would be a barrier, and one participant stated that even if the algorithm was 100% accurate, she still might not follow it.

Ethical dilemmas

Whether or not to discuss the UL prediction with patients would be a dilemma for many participants. If a patient was predicted to have little or no function, this might depress the patient and would conflict with the participants' desire to motivate them:

> *"Yes. And what day do we tell the patient? Is it when they arrive and have been here in…? Well. I really don't know. On top of everything else?" (Participant N, F4)*

Whereas some participants were sceptical, they were still open for discussion when other participants responded, that informing the patient would make it easier to focus on other aspects of the rehabilitation where improvement seemed more obtainable:

> *"I find it difficult to shatter someone's dream. You need to dream and believe this one will gain function. For some time. Of course, not for several years." (Participant G, F2)*

> *"Well…Well it is a balance, isn´t it. We have patients who come and tell us they are sorry that they weren't told…so the most important thing is to dare tell them, to be honest…well why should we treat an arm that we are nearly 100% will never function again?" (participant H, F2)*

Fear of consequences

The general view across interviews was that UL treatment should be offered regardless of the initial level of UL impairment. All patients deserved that the therapists did their best to restore UL function. In focus groups 3 and 4, concern was expressed that the use of an algorithm would eventually result in patients with a negative prediction receiving little or no UL treatment. If so, the algorithm would serve as a self-fulfilling prophecy. The participants feared that a prediction algorithm would alter their approach to the patients:

> *"And then I might prioritize other issues instead. I am afraid so. And I hope I wouldn't. Because I believe that they need all the treatment they can get…. Because, truly, there is a chance in reality." (Participant O, F4)*

> *"If I had a diagram that could tell…your arm will never be good…. then I believe the patient should get the opportunity to prove this wrong." (Participant J, F3)*

In all interviews, it was mentioned that an algorithm could be used to stratify and prioritize which patients should receive treatment. In one interview, the participants regarded this a positive consequence because it could be used to optimize treatment in times of limited resources. On the other hand, participants from two of the units looked at algorithms in light of pressure from budget cutbacks; they feared that an algorithm would be used to accelerate rehabilitation periods:

> *"It depends how – if you can say so - our managers wish to use this tool…because we are under pressure. And will this be a tool to evaluate…which patients should be here?" (Participant K, F3)*

4. Preconditions for implementation

Preconditions for a future successful implementation was the fourth main theme. It was grouped in two subthemes: tailored implementation and organizational structure and resources.

Tailored implementation

The focus group interviews were performed at three different units. Despite being part of the same rehabilitation hospital (RHN) the prevailing culture at each unit differed. Especially at unit 1, incorporating new evidence was embedded in the culture and the participants (see Table 10, page 62) seemed particularly open to the implementation of prediction models. While been open to new ideas too, the participants at the remaining two units were more sceptical.

In all interviews, the participants discussed the importance of tailoring implementation to the specific unit, ward, and patient. If something new had to be implemented, a persistent focus on the topic was needed:

> *"I believe that you must realize that implementation is just a lot more*

time-consuming and difficult than you imagine. A single day - when you present, discuss and maybe do something practical - is just not enough." (Participant D, F2)

"I can say…in my ward…if something must be implemented, you have to take the specific patient and the patient's team to make it work. We cannot say something general about you having to…..in all upper limbs...to do so and so. It has to be specific so they can relate to that." (Participant I, F2)

"We aren't different from the patients. We, too, need a lot of repetition to implement something new and learn it." [the group agrees] (Participant M, F3)

Organizational structure and resources

In all interviews the time to get acquainted with new evidence and practice new skills was considered insufficient. There was a sense among the interview participants of being well-informed, whereas time to incorporate and practice new skills and routines was lacking:

No. we don´t even have the time to plan our daily treatments. So no, not at all. That is a real challenge [the group agrees] (Participant E, F2)
"The level of information is actually okay…it is more the time afterwards…to incorporate it." (Participant J, F3)

"Yes exactly." (Participant K, F3)

"True …to make it a routine." (Participant J, F3)

Prioritization and support from the ward manager were stressed as important for success. In all interviews, participants mentioned that weekly or monthly meet-

ings could be dedicated to specific issues. These meetings were valued when new initiatives had to be implemented. Generally, the participants felt that there was a culture of sharing the acquired knowledge with colleagues.
Members of the staff were assigned specific positions as a specialist OT or specialist PT. Specialists were considered a resource, capable of, and responsible for presenting and implementing evidence:

> *"And when some of the specialist therapists have been out in the wide world and return home and tell us about it, or some colleagues have been at a course…" (participant M, F3)*

> *"For me, it is about responsibility. Someone has to take responsibility. Because if all are responsible, nothing happens. I, as a specialist, can be the one responsible and say: Your patient, has he got an UL problem?" (participant F, F2)*

Discussion

Study I

Summary of main results

When the time window to obtain the SAFE score and the MEP status was expanded and the PREP2 was applied two weeks after stroke, the overall CCR was 60% (95% CI 50- 71%).[89] However, CCR varied for each of the four UL prediction categories. For patients predicted Poor or Excellent UL function CCR was 78% and 74%, respectively. In contrast, CCR for patients predicted Limited or Good was only 33% and 35%, respectively. The overall CCR of 60% is considered too low to be clinically relevant. However, the high accuracy for patients predicted Poor or Excellent UL function may still be valuable to clinicians and patients. Despite an overall increase in CCR to 66% by CART, the CCR for the category Poor decreased to 67% and CCR was still very low for the categories Good and Limited. Thus, the algorithm could not be improved by CART analysis.

Comparison with other studies

The overall CCR of 60% found in Study I was statistically significantly lower than the overall accuracy of 75% found by Stinear et al[23] (15% difference, 95% CI 3-27%, P< 0.001, Chi2 test). Three possible explanations for the lower CCR in Study I should be mentioned. First, differences in study populations may contribute to the

lower CCR.[105] In the original study, patients were included within days of stroke. Thus, a part of these patients would have been either too severely or too mildly affected to be referred to in-patient rehabilitation 2 weeks after stroke. Consequently, patients in the present Study I may represent a more selected sample with respect to both UL impairment as well as other factors, e.g. motivation and co-morbidity, which may affect UL prognosis. Thus, the included patients still experienced UL impairment two weeks after stroke, but on the other hand, were expected to benefit from rehabilitation.

Second, that the CCR of 75% in the original study was computed from the training data used for developing it. Despite Stinear et al using pruning and cross validation to prevent data from being over-fitted, it might still be difficult to achieve an equally high accuracy in a subsequent data set.[105,106] A third possible explanation for the overall lower accuracy, and particularly for the large number of patients for whom the algorithm was too optimistic, is the decreased room for improvement two weeks after stroke. Most spontaneous biological recovery occurs early after stroke, and while the room for increases in UL function scores is high during the first days after stroke, it declines during the course of recovery.[107,108] When the prediction is obtained at a later point in time, patients will be closer to their maximally achievable UL function.

Results from the EPOS cohort study corroborate that the chance of recovery of UL function declines the longer the severe impairment lasts.[12] In the EPOS study 188 patients with stroke were included. The return of some dexterity on ARAT (i.e., ≥10 points) at 6 months was predicted by shoulder abduction and finger extension measured within 72 hours after stroke.[12] Retesting the model on days 5 and 9 showed that the probability of regaining dexterity remained 98% for those patients who were able to extend their fingers and abduct their shoulders, whereas the probability declined from 25% to 14% for those who did not satisfy either of these criteria.[12]

In the present population, MEP status was established using a belly-to-tendon placement of electrodes. This electrode placement will generally capture MEPs from not only one, but several muscles from the hand and forearm[79,109] and ensures, that not only activity limited to specific muscles, but any activity in the corticospinal tract is more likely to be captured. Other studies have used different TMS procedures[110-112] and e.g. used the resting motor threshold and recruitment curves steepness over the primary motor hand area (slope ratio between the ipsilesional and contralesional hemisphere) as a measure of corticomotor excitability.[110,111] This alternative method may provide a more detailed measure of corticomotor excitability. Still, in the present study, the TMS protocol by Stienar et al.[77] was followed as to ease comparison of results.[89] Additionally, other studies suggest that MEP amplitudes and latencies to single TMS pulses have adequate reliability for both healthy volunteers[113] and certain patient populations.[114]

The Stroke Recovery and Rehabilitation Roundtable recommends MEP as a biomarker for use in clinical trials for stratification purposes.[2] Still, its clinical application can be debated. In a recent review, Kim and Winstein suggest that TMS should be used in combination with clinical tests to predict UL function.[37] In line with this a review by Hayward et al.[36] concluded that the presence of MEPs as indexed by TMS was the only biomarker associated with better motor outcome, though with large inter-individual variability.[36] The value of MEP status as a supportive tool for prognosis has also been suggested by other research, though not always unambiguous.[22,23,35-41,115]

The unsatisfactory overall accuracy of PREP2 as applied in the present Study I can be mainly attributed to the poor accuracy for patients predicted Good based on MEP+. Studies have shown that the presence of MEP+ day 3-7 after stroke indicates a good prognosis and improves the prediction accuracy for UL function.[2,23,29,38,39,43] In contrast, the presence of MEP+ two weeks after stroke as in Study I is not as promising. This discrepancy in findings may be attributed to the above-mentioned spontaneous biological recovery. On the other hand, MEP- in-

dicates a poor prognosis for UL function both when obtained at two weeks after stroke as in the present study, or when obtained on day 3-7. Still, even patients with MEP- may improve. When MEP- is obtained within the first days of stroke, 15-20% of the patients recover at least some UL function.[32,39,117] In Study I as well, three of the 12 MEP- patients (25%) ended up in the outcome category *Limited* instead of the predicted category *Poor.* These findings indicate that in some of the patients with very poor UL function early after stroke, at least some regeneration of the corticospinal pathways or some compensation for the loss of functionality may occur. However, the category *Limited* still indicates a very compromised UL function.

The use of TMS may be an obstacle for the clinical use of the PREP2 algorithm. Patients have to be screened for contraindications for TMS, and a substantial part may not be available. In the present Study I, 8 patients could not have the prediction obtained either due to contraindications or due to technical issues. Furthermore, TMS can be relatively expensive and it requires trained staff. On the other hand, TMS is relatively easy and fast to perform. In Study I, MEP status could be established within ½ hour, including the setup of equipment.

Limitations and strengths

A limitation of Study I is that TMS examinations could not be conducted for all relevant patients. Six patients had contraindications to TMS[42], thus, their MEP status and subsequent UL prediction could not be established.[118] For patients who had their MEP status established, only 12 were MEP-. The predictions *Poor* or *Limited UL function* were consequently based on a relatively small number of patients, reflected in wide confidence intervals for the CCR.

For patients who were MEP- the NIHSS score was used to differentiate between the categories *Poor* and *Limited UL function*. In patients where the NIHSS score

was not available, the SSS score was instead converted into a NIHSS score according to a conversion model by Gray et al.[47] Whereas SSS and NIHSS are comparable, they are not identical and the CCR may have been higher if a true NIHSS score had been available. However, a SSS converted NIHSS score was necessary for six patients only.

Other factors than those included in the PREP2 algorithm may have an impact on future UL function. It has been described that factors such as individual goals, motivation, self-efficacy, neglect, aphasia and depression influence rehabilitation outcomes and should be considered.[119-122] In a study by Winters et al., 100 stroke patients without voluntary finger extension day 8 ± 4 days after stroke were followed prospectively and 45 of these patients achieved an Action Research Arm Test score of 10 points or more at 6 months.[123] In this study, the majority of patients with paresis mainly restricted to the upper limb, no neglect, and sufficient somatosensory function showed at least some return of UL capacity at 6 months after stroke.[123] In the present cohort, many patients suffered from neglect, impaired somatosensory function, and UL pain. Moreover, for the majority of patients, their motor deficits were not restricted to the UL, since most patients were unable to walk independently. The effect of these factors on UL function was not examined; nor was the effect of UL dose, treatment modality, or length of inpatient rehabilitation.

In alignment with the consensus-based core recommendations from the stroke recovery and rehabilitation roundtable, the patients in this longitudinal study were assessed at defined time points after stroke onset to account for underlying recovery processes.[50] Also, the recommendation for a wide range of patient demographics and baseline data to be collected in clinical studies were followed and reported to describe the included study population and enable comparison with other populations.[50]

To further minimize the risk of bias, patients were assessed with FMA and ARAT,

two reliable and valid UL assessments recommended for use in clinical trials.[50] Training of assessors prior to commencing the study, standardization of training and the use of assessment manuals has been shown to reduce variance in scoring, thereby increasing reliability.[45,46,74]

Additional strengths to be mentioned are that a relatively high number of patients were included. The study had very few dropouts, blinded obtainment of MEP status, blinded assessment at follow-up as well as patients and treating therapists being unaware of the UL prediction, including the MEP status.

Conclusion

Based on Study I, the PREP2 algorithm should not be implemented in clinical practice if the time window to obtain the SAFE score and MEP status is expanded to two weeks after stroke.

However, components from the PREP2 may be used to predict future UL function for certain patients. Patients with a SAFE score of 5 or above who are below the age of 80 years most likely achieve an excellent UL function three months after stroke. In patients with a SAFE score below 5, information on MEP- can be used to confirm that no useful UL function can be expected.

Study II

Summary of main results

In Study II, FMA predicted 38% of the variance in UL use ratio at three months after stroke. A multivariate regression model with FMA in combination with the

variables MEP status, neglect, dominant side affected, twopd, FIM score, gender and pain predicted 55%. In the multivariate model, the statistically significant predictors of use ratio were FMA, MEP status and neglect.
In contrast to what was found in the multivariate models, all potential predictor variables except pain were statistically significant independent predictors of UL use ratio in univariate regressions analyses. This is not surprising, as all variables are expressions of the same underlying phenomenon (stroke) and significant univariate predictors may become non-significant in the presence of other independent variables.[99]

Most prediction studies focus on how well the chosen predictors explain variation in outcome, reporting R^2 or adjusted R^2. Generally, little attention has been paid to the 95% CI or 95% PI of the prediction line. However, both of these estimates are informative. In the present study, the 95% CI gives information on the interval in which the true mean use ratio for a given FMA score will fall, with 95% accuracy. Thus, the 95% CI gives valuable information on prediction at a group level. In contrast, the 95% PI is an estimate at an individual level and displays the interval in which a future observation of UL use ratio for an individual patient will fall with 95% probability, given what has already been observed. In Study II the 95% PI of the regression lines were wide, reflecting that individual prediction of future UL use at a patient level in a clinical setting is difficult.

When use ratio was dichotomised in normal and non-normal, non-normal use could be predicted with very high accuracy in patients who were MEP- and/or had neglect. For the remaining patients, with MEP+ and without neglect, a logistic regression revealed that not achieving normal use could be predicted with a sensitivity of 0.80 and a specificity of 0.83.

Comparison with other studies

In line with the few previous studies, the most significant individual predictor of UL use in Study II was UL function at baseline.[18,21] Rand and Eng assessed patients at discharge from rehabilitation and one year after stroke.[18] They found that ARAT and grip strength combined with age were significant predictors of affected UL use assessed with accelerometers and Motor Activity Log. Contrary to Rand and Eng, statistically significant univariate prediction of gender, dominant UL affected, along with a range of other individual predictors were found in the present Study II. In a recent study by Buxbaum et al., the authors found that FMA and attention/ arousal predicted non-use.[21] However, they did not predict future UL use, but assessed the association between FMA and use at the same point in time. Moreover, UL use was not assessed in an unstructured environment but by means of observing UL movements during a clinical test. The mentioned studies, including the present Study II, indicates that while UL function is a prerequisite to UL use, UL use is not an imperative consequence of good UL function. This disparity between UL function and use has been confirmed by other studies.[54,56,124]

It was not possible to establish a threshold on FMA for normal use in Study II. This is in contrast to a study by Schweighofer et al. who found, that above a functional threshold, UL use improves.[125] However, their study was based on a very selected sample of patients, included 3-9 months after stroke, eligible for CIMT training in the EXCITE trial.[126] Moreover, their study did not assess use during daily life activities.

In the present study, the prediction accuracy of UL use ratio could be substantially increased by adding information on MEP status to the multiple regression analysis. This emphasizes the importance of corticospinal tract integrity and resembles findings from studies on prediction of UL function.[2,23,29,38,39] In Study I, the absence of MEPs in patients with severe paresis reliably resulted in poor function 3 months after stroke.[89] As suggested by Stinear et al.[23], information on corticospinal tract

integrity seems to be an indispensable component of prediction models for patients with severely impaired UL function and hence also for UL use.

Neglect was the third most important negative predictor of use ratio. Even in patients with only mild UL impairments at baseline, the presence of neglect was a major contributor to not achieving normal UL use. A recent review indicated that neglect is associated with poor UL motor recovery.[127] This may not be surprising, nevertheless, neglect is hardly ever explicitly addressed in motor rehabilitation programs. Neglect is a frequent phenomenon after stroke and was found present in roughly a quarter of patients in the present cohort, and according to a recent review in 30 % of stroke patients in general.[128] The fact that neglect is a major obstacle for daily life activities emphasizes the need for better assessment and treatment strategies, particularly in patients with motor potential.

The main outcome in Study II was the use ratio between affected and unaffected UL. No gold standard exists for which accelerometer outcome best expresses UL use[19] and other accelerometer outcomes too may provide valuable insights into UL use. Still, the use ratio is independent of varying activity levels between different people and has been recommended based on the clear clinical relevance in stroke rehabilitation populations with asymmetric effects on the limbs.[19]

Within stroke rehabilitation, the differentiation between true recovery of function and compensatory movements is a topic of focus.[129-131] For the moment, 3D kinematics represent state of the art for measurement of quality of movement or compensatory movement.[129,130] However, unlike accelerometers, 3D kinematics is not feasible for the measurement of real-life daily use. Still, a limitation of accelerometers may be that they do not capture the quality of movement and degree of compensation. Thus, a distinction between true recovery of function and compensatory movements cannot be made. However, a recent study by Barth et al. showed, that accelerometry, while mainly measuring movement quantity, could also reflect the use of general compensatory movement.[130] In this study, with 78

chronic stroke survivors, it was shown that individuals who move their UL more in daily life with respect to time and variability, tend to move with less compensatory movement and with a more typical movement pattern.[130] Thus, the patients in Study II would be expected to move with less compensatory movement of the affected UL as the use ratio approaches normal.

Whereas the Actigraph is the type of accelerometer most commonly used by researchers, it is rarely used in clinical practice.[19] If UL accelerometers are to be used at a larger scale in clinical practice, the output should be visible for the patient and used as a feedback mechanism.[19] For the lower limb, performance tracking at an individual level has been effective at increasing daily steps, physical activity, and reducing sedentary time in research studies of healthy populations.[132] In patients with stroke, monitoring of performance has been effective at improving daily walking activity[133,134] and walking endurance.[135] For the time being, there are major barriers to widespread clinical adoption of wearable sensor technology to measure UL use and most consumer-friendly device systems have questionable accuracy in rehabilitation populations.[19]

Limitations and strengths

MEP status was an important predictor in Study II. However, MEP was only examined in 34 patients with SAFE < 5. In patients with SAFE ≥ 5, MEP was assumed to be present. Despite MEP status not being explicitly examined in these patients, it nevertheless seems reasonable to assume that the corticospinal tract is at least partly intact in patients with active movement of the paretic UL. Previous research supports strong correlations between corticospinal tract integrity and motor function[136] and in a study of the PREP2 algorithm, Stinear et al. presumed MEP to be present in patients with SAFE ≥ 5.[23]

While the first regression model was based on data from 87 patients the following regressions were based on less participants as some had missing data on one or

more variables. Thus, some of the results are founded on a relatively small number of patients.

Currently, there is no universally accepted method for establishing neglect.[137] In the present cohort, neglect was established with a combination of two conventional tests. However, some patients may be able to compensate for their deficits during conventional testing, that require concentration for only a short period of time, but still have difficulties in daily life activities.[137] Thus, the presence of especially mild cases of neglect might be underestimated in the present cohort. Direct observation of patients' performance during ADL could have secured a focus on the patients' functional ability and impairment in real-world situations. As a consequence, more patients with neglect might have been identified.[137] Still, the use of not only one but two conventional neglect tests reduced the risk of over-looking patients with neglect. Further, the use of the centre of cancellation for the star cancellation test not only takes into account the number of omissions, but also their specific location. Thereby, spatially biased performance can be distinguished from inattentive performance.[86,87]

In the present study, accelerometry is assumed to reflect daily life UL use. However, voluntary functional movements are not completely identical to the movements captured by accelerometers and functional movements based on activity counts could be overestimated.[138] Yet, this overestimation concerns both limbs. A ratio, which is the relative duration of activity of one limb versus the other, will be less vulnerable to this bias compared to unilateral expressions of use. An increasing body of research supports the use of accelerometry as a valid and reliable tool to assess real-life use.[58,62] To summarize, for the time being, accelerometry and particularly the use ratio, appears as the closest resemblance to real-life use available.

Other limitations concern the practicalities of wrist-worn accelerometers. Some patients might have been wearing the accelerometers for less time than the

requested 12-hour period. However, data were visually inspected and excluded if activity was insufficient. Furthermore, compared to other accelerometer outputs, the use ratio is less likely to be affected by wearing time.

Wearing visible accelerometers may have encouraged the patients to increase their UL use. However, a recent study has shown that patients do not increase their physical activity while wearing accelerometers.[90] Nor does it matter what day of the week the patients used the accelerometers, as physical activity levels are very similar on weekends compared to weekdays.[90] Still, if a patient chose a day with high levels of UL training or UL activity, the use ratio might be slightly overestimated.[62]

Use ratio was predicted three months after stroke. Despite the majority of recovery occurring within this time span, patients with stroke may still experience recovery of UL beyond three months. Hence, prediction of use at a later point in time would be interesting. Especially patients with severe stroke and severe UL impairments may recover at a slower speed and improvements may only be captured if the timespan for prediction is expanded.

A strength to be highlighted is the prospective longitudinal study design with the predictor variables collected at an earlier point in time than the outcome variable.[39] Predictor variables were chosen a priory and not based on univariate analysis of their association with exposure and outcome. This theory founded selection of variables reduces the risk of including variables that are statistically significant by chance. Also, it reduces the risk of discarding variables that may be statistically significant in a larger sample, as the size of the p-value depends on the sample size.[99] By selecting variables a priori, the risk of including variables that are highly correlated or are surrogate measures of outcome was reduced.

Broad inclusion criteria were employed for the longitudinal study (Study I & II) and a substantial number of patients with a broad range of UL limitations were in-

cluded. Additionally, patients with co-morbidities or previous stroke were not excluded. Thus, the results are considered generalizable to the majority of patients with stroke and UL impairment admitted for in-patient rehabilitation. Still, being in need of in-patient rehabilitation implies that the impairments of the included patients were complex and not restricted to the UL. Thus, the results may not be generalizable to all patients at two weeks after stroke.

Conclusion

Predictors of UL use are relatively unexplored. Study II contributes with new knowledge on characteristics of patients who do not achieve normal UL use. It was shown that better function of the paretic UL at baseline predicted increased use of the arm and hand in daily life. Wide variation in the achievement of UL use existed and even patients with only mild UL impairment at two weeks poststroke may not achieve normal UL use at three months. Individual predictions were difficult due to this large variation in outcome. However, not achieving normal UL use could be predicted reliably based on the absence of MEPs and/ or the presence of neglect.

Study III

Summary of main results

In the qualitative study III perceptions of UL prediction models were explored and four main themes were identified: *current practice, perceived benefits, barriers,* and *preconditions for implementation*. While the majority of participants knew of UL prediction models, only some elements were applied in clinical practice and only by a few therapists. The PREP2 algorithm was seen as a potentially helpful

tool when planning treatment and setting goals. The perceived benefits encompassed the information derived from the SAFE score and the use of TMS. The main barriers were concern about the accuracy of the algorithm and having to confront patients with a negative prognosis. Preconditions for implementation encompassed having sufficient time, tailoring the implementation to a specific unit, and being part of an organization that supports implementation.

Comparison with other studies

Current practice was characterized by limited knowledge and use of UL measurements and UL prediction models. This is unsurprising, as the Danish Stroke Guidelines do not recommend the use of any particular UL measurement or UL prediction model.[8,9,139] The participants' skeptical attitudes towards measurements are in line with those expressed in a Danish study by Jaeger Pedersen et al. who showed that, despite being positive towards outcome measurements, OTs and PTs have reservations about standardization of the rehabilitation practice.[140] The participants in Study III proposed, that the PREP2 algorithm would be particularly helpful for recently graduated therapists. However, previous research indicates that even among experienced therapists prediction of UL function based on clinical expertise alone is less accurate than prediction models.[141]

In all interviews, the participants were positive towards the SAFE test and the use of TMS.[23] For approximately 2/3 of patients only the SAFE test is needed to perform the PREP2 prediction. According to Connell et al. simple tools as the SAFE test are more likely to be implemented.[67] The participants welcomed the use of TMS if this could help to distinguish patients who regained function from those who remained paralyzed. However, while a favourable prediction was considered motivating for both therapists and patients, most participants found a negative prediction demotivating. Individualized prediction is a new field for therapists and they may need assistance in delivering negative predictions.[67] Two homepages

provide suggestions on how to deliver negative PREP2 predictions.[48,142]

Despite all three units being part of the same rehabilitation hospital, different cultures existed, e.g. expressed in participants from unit 1 having a more welcoming attitude towards UL prediction models. A reason for this might be a greater focus on UL prediction models at unit 1 prior to the interviews. Furthermore, some of the differences may be attributed to the characteristics of the participants and the site. Differences in culture stress the importance of tailoring a future implementation to the particular setting in which it is intended to be used.[70,72,73]

Limitations and strengths

A limitation of Study III is that therapists with an interest in UL prediction may have been over-represented in the interviews. If so, the expressed perceptions may be more positive than amongst therapists in general. Another limitation is that the participants did not try to perform the complete PREP2 before attending the interviews. PREP2 is comprised of the SAFE test, NIHSS score, information on age and MEP status obtained by TMS. Performance of TMS requires a period of training, which was not feasible for the present study. As a consequence, the participants' considerations on TMS were merely theoretical. Still, before attending the interviews the participants practiced the most essential part of the PREP2, the SAFE test.

The emphasis in Study III was on the participants' perspectives. This approach was chosen as the beliefs of healthcare staff about an intervention are often more influential for implementation than other factors such as the strength of evidence supporting the intervention.[71-73]
A checklist for focus group interviews was used to assure that important considerations concerning the research team, methods, context of the study, analysis and interpretations were addressed.[104] The scientific trustworthiness was evaluated

using the concepts credibility, confirmability, and transferability.[91,102,103] To ensure credibility several researchers with different positions and perspectives were involved in the analysis.[91,103] In addition, a participant from each interview reviewed focus group transcripts and findings. Confirmability was assured by not letting the pre-understandings of the researchers involved in the study influence the interpretation of results.[91] Being aware of own preconceptions enabled the PhD fellow and the involved researchers to perform the analysis with an open mind.[102,103]

The transferability or generalizability concerns the application of the study findings in another context or setting.[103] Perceived barriers and facilitators for implementation will differ between sites, depending on the characteristics of the clinical setting and the people involved.[67,73] As a consequence, findings from the current study will not necessarily be generalizable to other settings. Nevertheless, the systematic use of the implementation research framework CFIR can be transferred to other contexts.

Conclusion

Study III revealed, that the perceptions of the participants only partly align with current scientific evidence, reflecting a lack of translation from evidence to applied knowledge. Thus, the evidence behind UL prediction models should be presented and discussed in detail with the therapists prior to implementation. Performance of UL tests and the SAFE score aligned more with the physiotherapy profession than the occupational therapy profession. If a prediction model is to be implemented, PTs may be the ones responsible for performing the prediction. A future implementation strategy should address how to support therapists in handling and delivering predictions, especially if they are negative.

Perspectives

The present PhD project origins from a desire to implement and use UL prediction models in a rehabilitation setting. Some implications for clinical practice could be derived. Most importantly, the PREP2 algorithm should not be implemented if the time window to obtain the PREP2 prediction is expanded to two weeks post stroke. Still, prospective longitudinal data from Study I and II showed that better function of the paretic UL at baseline predicted increased function and use of the arm and hand in daily life. These findings confirm results from previous studies and underscore the value of structured assessment of UL impairment. Valid and reliable tests uch as the FMA should be applied in clinical practice.

Results from the PhD project emphasize the prognostic value of MEPs, particularly for patients with severe paresis. The absence of MEPs was predictive of both poor UL function three months after stroke and of not achieving normal UL use. In line with findings from other studies, the use of TMS and the biomarker MEP contributes to UL prediction with knowledge that cannot be obtained by clinical tests alone. As a consequence, the use of TMS to obtain a MEP status should be considered in clinical practice to gain insight into UL prognosis. In patients without MEP, UL rehabilitation could focus on preventing pain and learning compensatory strategies.

Still, the need for special equipment and training might provide an obstacle for the use of TMS. If it is not feasible to perform TMS to establish MEP status, knowledge of other individual predictors found in Study II may be obtained to predict UL use, e.g. dominant hand affected or two-point discrimination.

The fact that neglect is a major obstacle to future use of the paretic UL in daily life emphasizes the need for better assessment and treatment strategies for neglect, particularly in patients with motor potential. In the present study, patients were screened for neglect with a combination of conventional tests. However, in a clinical setting, it may be more valuable to assess how neglect affects the patients in daily life activities, and a test like the KF-NAP may be more informative.

Even patients with only mild to moderate UL impairment at baseline did frequently not achieve a normal UL use. This confirms earlier findings, that improved UL function does not necessarily translate into increased UL use. Better and more effective training strategies are needed for patients with a potential for improvement. Furthermore, patients should be encouraged to use their affected UL, e.g. with the help of wrist-worn devices, which encourage UL movements and provide feedback. Such motivational aids could help to overcome learned non-use and increase the intensity of training, which is related with motor recovery.

For the time being, prediction models for the UL are still mainly used in clinical trials and at a group level. While reliable prediction on an individual level is a prerequisite for targeted rehabilitation, existing models are of limited value. More flexible UL prediction models are needed, with the possibilities of obtainment of predictions at later time point than within days of stroke. Future models could be based on multiple data entries, thereby increasing their precision. Also, the optimal UL prediction model should have high accuracy, not only for patients with mild or moderate UL function, but also for those with limited or no function. If prediction models are to be implemented in the clinical setting, they should be supported and facilitated by existing computer systems. Optimally, the required information for obtainment of prediction could be automatically extracted from the medical record. Future research should focus on improving prediction models and make them applicable for most clinical settings.

Acknowledgements

I would like to express my sincere gratitude to everyone who has contributed to this PhD project at Hammel Neurorehabilitation Centre and Research Clinic.

First and foremost, I wish to thank all the patients who consented to participate, despite being in a difficult life situation. Without these consents, it would not have been possible to conduct the PhD project, for the benefit of future patients.

I feel fortunate to have had an always supporting main supervisor, Iris Brunner, who has provided me guidance throughout the PhD process. Thanks for your positive and constructive feedback and for always finding the time to discuss the big and small topics related to the project. A respectful thanks go to my co-supervisor Jørgen Feldbæk Nielsen for his critical appraisal throughout the process, including methodological considerations and clarifying questions. A special thanks to my co-supervisor Tine Tjørnhøj-Thomsen for introducing me to implementation research and for providing an outsiders view on Neurorehabilitation, thus bringing me out of my comfort zone.

I am truly grateful that Hanne Pallesen has patiently shared her knowledge within qualitative research. My sincere thanks to Federico Gabriel Arguissain for setting up the TMS equipment and for analysis of TMS data. Thanks to Asger Roer Pedersen for statistical support and valuable advice.

The past year has been marked by Covid-19 and my planned research stay abroad was cancelled. Still, thank you to professor Cathy Stinear for responding to all my questions on the PREP2 algorithm and for giving me advice.

I have enjoyed being part of a very diverse research unit at RHN with colleagues who possess a wide range of competences, are curious and asks questions. I have had the pleasure to discuss important aspects of my PhD with many of you, for which I am very grateful. Thanks to Gritt Bennedsen for proofreading of the thesis.

My gratitude goes to the board of directors at RHN and to the ward managers and the clinical staff at the three rehabilitation units that constitute RHN: Hammel, Lemvig and Skive. A special thanks to the therapists who participated in the focus group interviews and to the therapists and secretaries who helped with practicalities and patient recruitment, making it possible to run a project at nine different wards distributed at three physically distinct rehabilitation units. Thank you very much to Sedsel Kristine Stage Pedersen and Ulla Kjærulf for assessment of patients.
Additionally, thanks to the therapists at the acute neurological wards in Århus and Holstebro for helping with data collection, data that I will be looking into in a near future.

I am grateful for the support from my close friends and a special thanks to Karina Poulsen for valuable discussions. Thanks to my parents for always believing in me and thanks to my daughters Martine and Helene for laughs and love. And finally, to my husband Jess, thank you for your encouragement and support.

References

1. Definition of an algorithm. Cambridge Dictionary. https://dictionary.cambridge.org/dictionary/english/algorithm. Accessed February 02, 2021.
2. Boyd LA, Hayward KS, Ward NS, et al. Biomarkers of Stroke Recovery: Consensus-Based Core Recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Neurorehabil Neural Repair.* 2017;31(10-11):864-876.
3. Heilman KM, Valenstein E. Mechanisms underlying hemispatial neglect. *Ann Neurol.* 1979;5(2): 166-170.
4. Definition of prediction. Cambridge Dictionary. https://dictionary.cambridge.org/dictionary/english/prediction. Accessed February 02, 2021.
5. Benjamin EJ, Virani SS, Callaway CW, et al. Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. *Circulation.* 2018;137(12):67-492.
6. Feigin VL, Krishnamurthi RV, Parmar P, et al. Update on the Global Burden of Ischemic and Hemorrhagic Stroke in 1990-2013: The GBD 2013 Study. *Neuroepidemiology.* 2015;45(3):161-176.
7. Flachs EM, Eriksen L, Koch MB, Ryd JT, Dibba E, Skov-Ettrup L, Juel K. Statens Institut for Folkesundhed, Syddansk Universitet. Sygdomsbyrden i Danmark –sygdomme. København: Sundhedsstyrelsen. 2015.
8. Sundhedsstyrelsen. National klinisk retningslinje for fysioterapi og ergoterapi til voksne med nedsat funktionsevne som følge af erhvervet hjerneskade, herunder apopleksi. 2014.
9. Sundhedsstyrelsen. Anbefalinger for tværsektorielle forløb for voksne med erhvervet hjerneskade-apopleksi og transitorisk cerebral iskæmi (TCI) - trauma, infektion, tumor, subarachnoidalblødning og encephalopati. 2020.

10. Vestergaard SV, Rasmussen TB, Stallknecht S, et al. Occurrence, mortality and cost of brain disorders in Denmark: a population-based cohort study. *BMJ Open.* 2020;10(11):e037564.
11. Persson HC, Parziali M, Danielsson A, Sunnerhagen KS. Outcome and upper extremity function within 72 hours after first occasion of stroke in an unselected population at a stroke unit. A part of the SALGOT study. *BMC neurology*. 2012;12:162-2377-2312-2162.
12. Nijland RH, van Wegen EE, Harmeling-van der Wel BC, Kwakkel G, Investigators E. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery: early prediction of functional outcome after stroke: the EPOS cohort study. *Stroke*. 2010;41(4):745-750.
13. Kwakkel G, Kollen BJ, van der Grond J, Prevo AJ. Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. *Stroke*. 2003;34(9):2181-2186.
14. Faria-Fortini I, Michaelsen SM, Cassiano JG, Teixeira-Salmela LF. Upper extremity function in stroke subjects: relationships between the international classification of functioning, disability, and health domains. *J Hand Ther.* 2011;24(3):257-264.
15. Veerbeek JM, Kwakkel G, van Wegen EE, Ket JC, Heymans MW. Early prediction of outcome of activities of daily living after stroke: a systematic review. *Stroke*. 2011;42(5):1482-1488.
16. Morris JH, van Wijck F, Joice S, Donaghy M. Predicting health related quality of life 6 months after stroke: the role of anxiety and upper limb dysfunction. *Disabil Rehabil.* 2013;35(4):291-299.
17. WHO. Towards A Common Language for Functioning, Disability, and Health: ICF. 2002.
18. Rand D, Eng JJ. Predicting daily use of the affected upper extremity 1 year after stroke. *J. Stroke Cerebrovasc. Dis.* 2015;24(2):274-283.
19. Lang CE, Barth J, Holleran CL, Konrad JD, Bland MD. Implementation of Wearable Sensing Technology for Movement: Pushing Forward into the Routine Physical Rehabilitation Care Field. *Sensors*. 2020;20(20).

20. Varghese R, Kutch JJ, Schweighofer N, Winstein CJ. The probability of choosing both hands depends on an interaction between motor capacity and limb-specific control in chronic stroke. *Exp Brain Res.* 2020;238(11):2569-2579.
21. Buxbaum LJ, Varghese R, Stoll H, Winstein CJ. Predictors of Arm Nonuse in Chronic Stroke: A Preliminary Investigation. *Neurorehabil Neural Repair.* 2020;34(6):512-522.
22. Kwah LK, Herbert RD. Prediction of Walking and Arm Recovery after Stroke: A Critical Review. *Brain sciences.* 2016;6(4):E53.
23. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA. PREP2: A biomarker-based algorithm for predicting upper limb function after stroke. *Ann Clin Transl Neurol.* 2017;4(11):811-820.
24. Winters C, van Wegen EE, Daffertshofer A, Kwakkel G. Generalizability of the Proportional Recovery Model for the Upper Extremity After an Ischemic Stroke. *Neurorehabil Neural Repair.* 2015;29(7):614-622.
25. Prabhakaran S, Zarahn E, Riley C, et al. Inter-individual variability in the capacity for motor recovery after ischemic stroke. *Neurorehabil Neural Repair.* 2008;22(1):64-71.
26. Katrak P, Bowring G, Conroy P, Chilvers M, Poulos R, McNeil D. Predicting upper limb recovery after stroke: the place of early shoulder and hand movement. *Arch Phys Med Rehabil.* 1998;79(7):758-761.
27. Smania N, Paolucci S, Tinazzi M, et al. Active finger extension: a simple movement predicting recovery of arm function in patients with acute stroke. *Stroke*. 2007;38(3):1088-1090.
28. Buch ER, Rizk S, Nicolo P, Cohen LG, Schnider A, Guggisberg AG. Predicting motor improvement after stroke with clinical assessment and diffusion tensor imaging. *Neurology*. 2016;86(20):1924-1925.
29. Byblow WD, Stinear CM, Barber PA, Petoe MA, Ackerley SJ. Proportional recovery after stroke depends on corticomotor integrity. *Annals of Neurology.* 2015;78(6):848-859.

30. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA. Proportional Motor Recovery After Stroke: Implications for Trial Design. *Stroke*. 2017;48(3):795-798.
31. Gebruers N, Truijen S, Engelborghs S, De Deyn PP. Prediction of upper limb recovery, general disability, and rehabilitation status by activity measurements assessed by accelerometers or the Fugl-Meyer score in acute stroke. *Am J Phys Med Rehabil.* 2014;93(3):245-252.
32. Zarahn E, Alon L, Ryan SL, et al. Prediction of motor recovery using initial impairment and fMRI 48 h poststroke. *Cereb Cortex.* 2011;21(12):2712-2721.
33. Stinear CM, Barber PA, Petoe M, Anwar S, Byblow WD. The PREP algorithm predicts potential for upper limb recovery after stroke. *Brain*. 2012;135(Pt 8):2527-2535.
34. Boccuni L, Meyer S, D'Cruz N, et al. Premotor dorsal white matter integrity for the prediction of upper limb motor impairment after stroke. *Sci Rep.* 2019;9(1):19712.
35. Stinear CM. Prediction of motor recovery after stroke: advances in biomarkers. *The Lancet Neurol.* 2017;16(10):826-836.
36. Hayward KS, Schmidt J, Lohse KR, et al. Are we armed with the right data? Pooled individual data review of biomarkers in people with severe upper limb impairment after stroke. *NeuroImageClinical*. 2016;13:310-319.
37. Kim B, Winstein C. Can Neurological Biomarkers of Brain Impairment Be Used to Predict Poststroke Motor Recovery? A Systematic Review. *Neurorehabil Neural Repair*. 2017;31(1):3-24.
38. Pizzi A, Carrai R, Falsini C, Martini M, Verdesca S, Grippo A. Prognostic value of motor evoked potentials in motor function recovery of upper limb after stroke. *J Rehabil Med.* 2009;41(8):654-660.
39. van Kuijk AA, Pasman JW, Hendricks HT, Zwarts MJ, Geurts AC. Predicting hand motor recovery in severe stroke: the role of motor evoked potentials in relation to early clinical assessment. *Neurorehabil Neural Repair.* 2009;23(1):45-51.
40. Rosso C, Lamy JC. Does Resting Motor Threshold Predict Motor Hand Recovery After Stroke? *Front Neurol.* 2018;9:1020.

41. Smith MC, Ackerley SJ, Barber PA, Byblow WD, Stinear CM. PREP2 Algorithm Predictions Are Correct at 2 Years Poststroke for Most Patients. *Neurorehabil Neural Repair.* 2019;33(8):635-642.
42. Groppa S, Oliviero A, Eisen A, et al. A practical guide to diagnostic transcranial magnetic stimulation: report of an IFCN committee. *Clin Neurophysiol.* 2012;123(5):858-882.
43. Bembenek JP, Kurczych K, Karli Nski M, Czlonkowska A. The prognostic value of motor-evoked potentials in motor recovery and functional outcome after stroke - a systematic review of the literature. *Funct Neurol.* 2012;27(2):79-84.
44. Stinear CM, Byblow WD. Predicting and accelerating motor recovery after stroke. *Curr. Opin. Neurol.* 2014;27(6):624-630.
45. Platz T, Pinkowski C, van Wijck F, Kim IH, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. *Clin Rehabil.* 2005;19(4):404-411.
46. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair.* 2008;22(1):78-90.
47. Gray LJ, Ali M, Lyden PD, Bath PM, Virtual International Stroke Trials Archive C. Interconversion of the National Institutes of Health Stroke Scale and Scandinavian Stroke Scale in acute stroke. *J Stroke Cerebrovasc Dis.* 2009;18(6):466-468.
48. PRESTO PREP2. https://presto.auckland.ac.nz/prep2-overview/. Accessed September 09, 2020.
49. Moore JL, Potter K, Blankshain K, Kaplan SL, O'Dwyer LC, Sullivan JE. A Core Set of Outcome Measures for Adults With Neurologic Conditions Undergoing Rehabilitation: A CLINICAL PRACTICE GUIDELINE. J *Neurol Phys Ther.* 2018;42(3):174-220.
50. Kwakkel G, Lannin NA, Borschmann K, et al. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Int J Stroke.* 2017;12(5):451-461.

51. Waddell KJ, Strube MJ, Bailey RR, et al. Does Task-Specific Training Improve Upper Limb Performance in Daily Life Poststroke? *Neurorehabil Neural Repair.* 2017;31(3):290-300.
52. Han CE, Kim S, Chen S, et al. Quantifying arm nonuse in individuals post-stroke. *Neurorehabil Neural Repair.* 2013;27(5):439-447.
53. Stewart JC, Cramer SC. Patient-reported measures provide unique insights into motor function after stroke. *Stroke*. 2013;44(4):1111-1116.
54. Doman CA, Waddell KJ, Bailey RR, Moore JL, Lang CE. Changes in Upper-Extremity Functional Capacity and Daily Performance During Outpatient Occupational Therapy for People With Stroke. *Am. J. Occup. Therap.* 2016;70(3):7003290040p7003290041-7003290040p7003290011.
55. Sterr A, Freivogel S, Schmalohr D. Neurobehavioral aspects of recovery: assessment of the learned nonuse phenomenon in hemiparetic adolescents. *Arch Phys Med Rehabil.* 2002;83(12):1726-1731.
56. Essers B, Meyer S, De Bruyn N, et al. Mismatch between observed and perceived upper limb function: an eye-catching phenomenon after stroke. *Disabil Rehabil.* 2019;41(13):1545-1551.
57. Waddell KJ, Birkenmeier RL, Bland MD, Lang CE. An exploratory analysis of the self-reported goals of individuals with chronic upper-extremity paresis following stroke. *Disabil Rehabil.* 2016;38(9):853-857.
58. Uswatte G, Giuliani C, Winstein C, Zeringue A, Hobbs L, Wolf SL. Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial. *Arch Phys Med Rehabil.* 2006;87(10):1340-1345.
59. de Niet M, Bussmann JB, Ribbers GM, Stam HJ. The stroke upper-limb activity monitor: its sensitivity to measure hemiplegic upper-limb activity during daily life. *Arch Phys Med Rehabil.* 2007;88(9):1121-1126.
60. Lang CE, Waddell KJ, Klaesner JW, Bland MD. A Method for Quantifying Upper Limb Performance in Daily Life Using Accelerometers. *JoVE.* 2017;(122). doi(122):10.3791/55673.

61. Bailey RR, Klaesner JW, Lang CE. Quantifying Real-World Upper-Limb Activity in Nondisabled Adults and Adults With Chronic Stroke. *Neurorehabil Neural Repair.* 2015;29(10):969-978.
62. Chin LF, Hayward KS, Brauer S. Upper limb use differs among people with varied upper limb impairment levels early post-stroke: a single-site, cross-sectional, observational study. *Top Stroke Rehabil*. 2020;27(3):224-235.
63. Bailey RR, Lang CE. Upper-limb activity in adults: referent values using accelerometry. *J Rehabil Res Dev.* 2013;50(9):1213-1222.
64. Stinear CM, Byblow WD, Ackerley SJ, Barber PA, Smith MC. Predicting Recovery Potential for Individual Stroke Patients Increases Rehabilitation Efficiency. *Stroke*. 2017;48(4):1011-1019.
65. Balas EA, Boren SA. Managing Clinical Knowledge for Health Care Improvement. *Yearb Med Inform.* 2000(1):65-70.
66. Eng JJ, Bird ML, Godecke E, et al. Moving stroke rehabilitation research evidence into clinical practice: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Int J Stroke.* 2019;14(8):766-773.
67. Connell LA, Smith MC, Byblow WD, Stinear CM. Implementing biomarkers to predict motor recovery after stroke. *NeuroRehabilitation*. 2018;43(1):41-50.
68. Lynch EA, Chesworth BM, Connell LA. Implementation-The Missing Link in the Research Translation Pipeline: Is It Any Wonder No One Ever Implements Evidence-Based Practice? *Neurorehabil Neural Repair.* 2018:1545968318777844.
69. Kiaer C, Lundquist CB, Brunner I. Knowledge and application of upper limb prediction models and attitude toward prognosis among physiotherapists and occupational therapists in the clinical stroke setting. *Top Stroke Rehabil.* 2020:1-7.
70. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the Consolidated Framework for Implementation Research to Identify Barriers and Facilitators for the Implementation of an Internet-Based Patient-Provider Communication Service in Five Settings: A Qualitative Study. *J. Med. Internet Res.* 2015;17(11):e262.

71. Connell LA, McMahon NE, Harris JE, Watkins CL, Eng JJ. A formative evaluation of the implementation of an upper limb stroke rehabilitation intervention in clinical practice: a qualitative interview study. *Implement Sci..* 2014;9:90-014-0090-0093.
72. May C. Towards a general theory of implementation. *Implement Sci.* 2013;8:18-5908-5908-5918.
73. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci.*. 2009;4:50-5908-5904-5950.
74. Lundquist CB, Maribo T. The Fugl-Meyer assessment of the upper extremity: reliability, responsiveness and validity of the Danish version. *Disabil Rehabil.* 2017;39(9):934-939.
75. Gladstone DJ, Danells CJ, Black SE. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabil Neural Repair.* 2002;16(3):232-240.
76. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *Int J Rehabil Res.* 1981;4(4):483-492.
77. TMS Instructions for PREP2. https://presto.auckland.ac.nz/tms-instructions-for-prep2/. Accessed August 25, 2020.
78. Safety Checklist for TMS. https://presto.auckland.ac.nz/safety-checklist-for-tms/. Accessed August 25, 2020.
79. Morita H, Baumgarten J, Petersen N, Christensen LO, Nielsen J. Recruitment of extensor-carpi-radialis motor units by transcranial magnetic stimulation and radial-nerve stimulation in human subjects. *Exp Brain Res.* 1999;128(4):557-562.
80. Hall J, Dudgeon B, Guthrie M. Validity of clinical measures of shoulder subluxation in adults with poststroke hemiplegia. The American Journal of Occupational Therapy : Official Publication of the American Occupational Therapy Association. 1995;49(6):526-533.

81. Wu CY, Chuang IC, Ma HI, Lin KC, Chen CL. Validity and Responsiveness of the Revised Nottingham Sensation Assessment for Outcome Evaluation in Stroke Rehabilitation. *Am. J. Occup. Therap.* 2016;70(2):7002290040p7002290041-7002290048.
82. Au-Yeung SS, Hui-Chan CW. Predicting recovery of dextrous hand function in acute stroke. *Disabil Rehabil.* 2009;31(5):394-401.
83. van Nes SI, Faber CG, Hamers RM, et al. Revising two-point discrimination assessment in normal aging and in patients with polyneuropathies. *J Neurol Neurosurg Psychiatry.* 2008;79(7):832-834.
84. Karcioglu O, Topacoglu H, Dikme O, Dikme O. A systematic review of the pain scales in adults: Which to use? *Am J Emerg Med.* 2018;36(4):707-714.
85. Agrell BM, Dehlin OI, Dahlgren CJ. Neglect in elderly stroke patients: a comparison of five tests. *Psychiatry Clin Neurosci.* 1997;51(5):295-300.
86. Nijboer TCW, Winters C, Kollen BJ, Kwakkel G. Impact of clinical severity of stroke on the severity and recovery of visuospatial neglect. *PLoS One.* 2018;13(7):e0198755.
87. Rorden C, Karnath HO. A simple measure of neglect severity. *Neuropsychologia.* 2010;48(9):2758-2763.
88. Mehrholz J, Wagner K, Rutte K, Meissner D, Pohl M. Predictive validity and responsiveness of the functional ambulation category in hemiparetic patients after stroke. *Arch Phys Med Rehabil.* 2007;88(10):1314-1319.
89. Lundquist CB, Nielsen JF, Arguissain FG, Brunner IC. Accuracy of the Upper Limb Prediction Algorithm PREP2 Applied 2 Weeks Poststroke: A Prospective Longitudinal Study. *Neurorehabil Neural Repair.* 2020:1545968320971763.
90. Tinlin L, Fini N, Bernhardt J, Lewis LK, Olds T, English C. Best practice guidelines for the measurement of physical activity levels in stroke survivors: a secondary analysis of an observational study. *Int J Rehabil Res.* 2018;41(1):14-19.
91. Ohman A. Qualitative methodology for rehabilitation research. *J Rehabil Med*. 2005;37(5):273-280.
92. Krueger R, Casey M. Planning the Focus Group Study. In: Focus groups: A Practical Guide for Applied Research. 5rd ed. SAGE Publications; 2015:19-37.

93. Wilkinson S. Focus groups in health research: exploring the meanings of health and illness. *J Health Psychol.* 1998;3(3):329-348.
94. Batterham M, Neale E, Martin A, Tapsell L. Data mining: Potential applications in research on nutrition and health. *Nutr Diet.* 2017;74(1):3-10.
95. Hastie T, Tibshirani, R, Friedman, J. The elements of statistical learning. Data mining, inference, and prediction. In: second ed. New York: Springer; 2009 305-317.
96. Harris JE, Eng JJ. Individuals with the dominant hand affected following stroke demonstrate less impairment than those with the nondominant hand affected. *Neurorehabil Neural Repair.* 2006;20(3):380-389.
97. Rand D, Eng JJ. Arm-hand use in healthy older adults. *Am J Occup Ther.* 2010;64(6):877-885.
98. Kutner MH, Neter J, Li W. Applied Linear Statistical Models. 5th ed. New York: McGraw-Hill; 2005.
99. Kent P, Cancelliere C, Boyle E, Cassidy JD, Kongsted A. A conceptual framework for prognostic research. *BMC Med Res Methodol.* 2020;20(1):172.
100. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol.* 1996;49(12):1373-1379.
101. Kirkwood BRS, A.C. Essential Medical Statistics 2 ed.: Blackwell Science 2003;339-340.
102. Elo S, Kyngas H. The qualitative content analysis process. *J. Adv. Nurs.* 2008;62(1):107-115.
103. Malterud K. Qualitative research: standards, challenges, and guidelines. *Lancet.* 2001;358(9280):483-488.
104. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Care.* 2007;19(6):349-357.
105. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009;338:b605.
106. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338:b604.

107. Verheyden G, Nieuwboer A, De Wit L, et al. Time course of trunk, arm, leg, and functional recovery after ischemic stroke. *Neurorehabil Neural Repair.* 2008;22(2):173-179.
108. Kwakkel G, Kollen B, Twisk J. Impact of time on improvement of outcome after stroke. *Stroke*. 2006;37(9):2348-2353.
109. Butler AJ, Wolf SL. Transcranial magnetic stimulation to assess cortical plasticity: a critical perspective for stroke rehabilitation. *J Rehabil Med.* 2003(41 Suppl):20-26.
110. Ward NS, Newton JM, Swayne OB, et al. Motor system activation after subcortical stroke depends on corticospinal system integrity. *Brain*. 2006;129(Pt 3):809-819.
111. Lotze M, Roschka S, Domin M, Platz T. Predicting Training Gain for a 3 Week Period of Arm Ability Training in the Subacute Stage After Stroke. *Front Neurol.* 2018;9:854.
112. Hayward KS, Lohse KR, Bernhardt J, Lang CE, Boyd LA. Characterising Arm Recovery in People with Severe Stroke (CARPSS): protocol for a 12-month observational study of clinical, neuroimaging and neurophysiological biomarkers. *BMJ Open.* 2018;8(11):e026435.
113. Goldsworthy MR, Hordacre B, Ridding MC. Minimum number of trials required for within- and between-session reliability of TMS measures of corticospinal excitability. *Neuroscience*. 2016;320:205-209.
114. Fried PJ, Jannati A, Davila-Perez P, Pascual-Leone A. Reproducibility of Single-Pulse, Paired-Pulse, and Intermittent Theta-Burst TMS Measures in Healthy Aging, Type-2 Diabetes, and Alzheimer's Disease. *Front Aging Neurosci.* 2017;9:263.
115. Hoonhorst MHJ, Nijland RHM, van den Berg PJS, Emmelot CH, Kollen BJ, Kwakkel G. Does Transcranial Magnetic Stimulation Have an Added Value to Clinical Assessment in Predicting Upper-Limb Function Very Early After Severe Stroke? *Neurorehabil Neural Repair.* 2018;32(8):682-690.
116. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA. PREP2: A biomarker-based algorithm for predicting upper limb function after stroke. *Ann Clin Trans Neurol.* 2017;4(11):811-820.

117. Hendricks HT, Pasman JW, Merx JL, van Limbeek J, Zwarts MJ. Analysis of recovery processes after stroke by means of transcranial magnetic stimulation. *J Clin Neurophysiol.* 2003;20(3):188-195.
118. Cantone M, Lanza G, Vinciguerra L, et al. Age, Height, and Sex on Motor Evoked Potentials: Translational Data From a Large Italian Cohort in a Clinical Environment. *Front Hum Neurosci.* 2019;13:185.
119. Winstein C, Varghese R. Been there, done that, so what's next for arm and hand rehabilitation in stroke? *NeuroRehabilitation*. 2018;43(1):3-18.
120. Jones F, Postges H, Brimicombe L. Building Bridges between healthcare professionals, patients and families: A coproduced and integrated approach to self-management support in stroke. *NeuroRehabilitation*. 2016;39(4):471-480.
121. Mayo NE, Fellows LK, Scott SC, Cameron J, Wood-Dauphinee S. A longitudinal view of apathy and its impact after stroke. *Stroke*. 2009;40(10):3299-3307.
122. Subramanian SK, Chilingaryan G, Sveistrup H, Levin MF. Depressive symptoms influence use of feedback for motor learning and recovery in chronic stroke. *Restor Neurol Neurosci.* 2015;33(5):727-740.
123. Winters C, Kwakkel G, Nijland R, van Wegen E, consortium EX-s. When Does Return of Voluntary Finger Extension Occur Post-Stroke? A Prospective Cohort Study. *PLoS One.* 2016;11(8):e0160528.
124. Franck JA, Smeets R, Seelen HAM. Changes in actual arm-hand use in stroke patients during and after clinical rehabilitation involving a well-defined arm-hand rehabilitation program: A prospective cohort study. *PLoS One.* 2019;14(4):e0214651.
125. Schweighofer N, Han CE, Wolf SL, Arbib MA, Winstein CJ. A functional threshold for long-term use of hand and arm function can be determined: predictions from a computational model and supporting data from the Extremity Constraint-Induced Therapy Evaluation (EXCITE) Trial. *Phys Ther.* 2009;89(12):1327-1336.

126. Wolf SL, Winstein CJ, Miller JP, et al. Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the EXCITE randomized clinical trial. *JAMA*. 2006;296(17):2095-2104.
127. Doron N, Rand D. Is Unilateral Spatial Neglect Associated With Motor Recovery of the Affected Upper Extremity Poststroke? A Systematic Review. *Neurorehabil Neural Repair.* 2019;33(3):179-187.
128. Esposito E, Shekhtman G, Chen P. Prevalence of spatial neglect post-stroke: a systematic review. *Ann Phys Rehabil Med.* 2020.
129. Kwakkel G, van Wegen EEH, Burridge JH, et al. Standardized Measurement of Quality of Upper Limb Movement After Stroke: Consensus-Based Core Recommendations From the Second Stroke Recovery and Rehabilitation Roundtable. *Neurorehabil Neural Repair.* 2019;33(11):951-958.
130. Barth J, Klaesner JW, Lang CE. Relationships between accelerometry and general compensatory movements of the upper limb after stroke. *J Neuroeng Rehabil.* 2020;17(1):138.
131. Levin MF, Kleim JA, Wolf SL. What do motor "recovery" and "compensation" mean in patients following stroke? *Neurorehabil Neural Repair.* 2009;23(4):313-319.
132. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. *JMIR Mhealth Uhealth.* 2019;7(4):e11819.
133. Danks KA, Roos MA, McCoy D, Reisman DS. A step activity monitoring program improves real world walking activity post stroke. *Disabil Rehabil.* 2014;36(26):2233-2236.
134. Moore JL, Roth EJ, Killian C, Hornby TG. Locomotor training improves daily stepping activity and gait efficiency in individuals poststroke who have reached a "plateau" in recovery. *Stroke*. 2010;41(1):129-135.
135. Danks KA, Pohlig R, Reisman DS. Combining Fast-Walking Training and a Step Activity Monitoring Program to Improve Daily Walking Activity After Stroke: A Preliminary Study. *Arch Phys Med Rehabil.* 2016;97(9 Suppl):S185-193.

136. Koch P, Schulz R, Hummel FC. Structural connectivity analyses in motor recovery research after stroke. *Ann Clin Transl Neurol.* 2016;3(3):233-244.
137. Pitteri M, Chen P, Passarini L, Albanese S, Meneghello F, Barrett AM. Conventional and functional assessment of spatial neglect: Clinical practice suggestions. *Neuropsychology.* 2018;32(7):835-842.
138. Lum PS, Shu L, Bochniewicz EM, et al. Improving Accelerometry-Based Measurement of Functional Use of the Upper Extremity After Stroke: Machine Learning Versus Counts Threshold Method. *Neurorehabil Neural Repair.* 2020;34(12):1078-1087.
139. Sundhedsstyrelsen. Anbefalinger til nationale redskaber til vurdering af funktionsevne - hos voksne med erhvervet hjerneskade. 2020.
140. Jaeger Pedersen T, Kaae Kristensen H. A critical discourse analysis of the attitudes of occupational therapists and physiotherapists towards the systematic use of standardised outcome measurement. *Disabil Rehabil.* 2016;38(16):1592-1602.
141. Nijland RH, van Wegen EE, Harmeling-van der Wel BC, Kwakkel G, Early Prediction of Functional Outcome After Stroke I. Accuracy of physical therapists' early predictions of upper-limb function in hospital stroke units: the EPOS Study. *Phys Ther.* 2013;93(4):460-469.
142. PREP2 training. https://preptraining.auckland.ac.nz/. Accessed January 01 2021.

Appendix

Papers 1 - 3

Declaration of co-author ship for paper 1 - 3

Paper 1

Original Research Article

Neurorehabilitation and Neural Repair
1–11
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1545968320971763
journals.sagepub.com/home/nnr
SAGE

Accuracy of the Upper Limb Prediction Algorithm PREP2 Applied 2 Weeks Poststroke: A Prospective Longitudinal Study

Camilla Biering Lundquist, MSc, PT[1], Jørgen Feldbæk Nielsen, MDsc, MD[1], Federico Gabriel Arguissain, PhD[1,2], and Iris Charlotte Brunner, PhD, PT[1,3]

Abstract

Background. The Predict Recovery Potential algorithm (PREP2) was developed to predict upper limb (UL) function early after stroke. However, assessment in the acute phase is not always possible. *Objective*. To assess the prognostic accuracy of the PREP2 when applied in a subacute neurorehabilitation setting. *Methods*. This prospective longitudinal study included patients ≥18 years old with UL impairment following stroke. Patients were assessed in accordance with the PREP2 approach. However, 2 main components, the shoulder abduction finger extension (SAFE) score and motor-evoked potentials (MEPs) were obtained 2 weeks poststroke. UL function at 3 months was predicted in 1 of 4 categories and compared with the actual outcome at 3 months as assessed by the Action Research Arm Test. The prediction accuracy of the PREP2 was quantified using the correct classification rate (CCR). *Results*. Ninety-one patients were included. Overall CCR of the PREP2 was 60% (95% CI 50%-71%). Within the 4 categories, CCR ranged from the lowest value at 33% (95% CI 4%-85%) for the category Limited to the highest value at 78% (95% CI 43%-95%) for the category Poor. In the present study, the overall CCR was significantly lower ($P < .001$) than the 75% reported by the PREP2 developers. *Conclusions*. The low overall CCR makes PREP2 obtained 2 weeks poststroke unsuited for clinical implementation. However, PREP2 may be used to predict either excellent UL function in already well-recovered patients or poor UL function in patients with persistent severe UL paresis.

Keywords

stroke, rehabilitation, upper extremity, algorithms, PREP2, prediction

Background

Stroke is a leading cause of death and disability worldwide,[1] and upper limb (UL) impairment has been reported in 48% of stroke survivors in the acute phase[2] and in 30% to 66% of stroke survivors in the chronic phase.[3,4] UL impairment is related to subsequent functional limitations affecting activities of daily living.[5,6] Accurate prediction of UL function can provide patients and therapists with realistic expectations for UL prognosis, help set individual goals for rehabilitation, and may result in more effective utilization of health resources.[7-9]

In 5 prospective longitudinal studies published in 2014-2017, researchers measured UL motor impairment within 2 weeks of stroke and at 3 or 6 months after stroke.[10-14] These studies showed that most patients recover 70% to 80% of their maximum possible UL motor function within 3 to 6 months after stroke.[10-14] However, increased scores on clinical assessments demonstrated at a group level have little individual prognostic value due to great variation between individuals[15] and which patients will regain UL function cannot be safely predicted from clinical measures alone.[11,12]

[1]Hammel Neurorehabilitation Centre and University Research Clinic, Hammel, Denmark
[2]Aalborg University, Aalborg East, North Jutland, Denmark
[3]Aarhus University, Aarhus, Denmark

Supplementary material for this article is available on the Neurorehabilitation & Neural Repair website along with the online version of this article.

Corresponding Author:
Camilla Biering Lundquist, Research Department, Hammel Neurorehabilitation Centre and University Research Clinic, Voldbyvej 15, Hammel, 8450, Denmark.
Email: camilla.lundquist@midt.rm.dk

In patients with severe UL impairment, the use of a biomarker may improve prediction accuracy for motor recovery.[8,16,17] A biomarker is an indicator of disease state that can be used clinically to reflect underlying molecular events and/or predict outcomes associated with recovery from stroke.[18] A biomarker widely used to assess corticospinal excitability is motor-evoked potentials (MEPs). MEPs are motor contractions elicited by pulses of transcranial magnetic stimulation (TMS).[19] Patients in whom TMS elicits MEPs in muscles of the paretic limb generally achieve better and faster motor recovery than patients without MEPs.[8,12,20] In a recent review, MEPs at rest was the only biomarker predicting motor outcome in individuals with severe UL impairment following stroke.[16]

Several prediction algorithms for UL function have been proposed and evaluated in clinical trials.[8,9,11,21-23] However, the majority of these algorithms are most accurate for predicting recovery in patients with mild to moderate UL impairment.[9,12,24] The Predict Recovery Potential (PREP2) algorithm stands out as its overall predictive value is reported to be 75%.[25] Especially in patients with severe paresis its accuracy exceeds that of previous prediction algorithms.[7-9] The first version of the PREP2 algorithm increased therapist confidence and rehabilitation efficacy.[26] Hence, research indicates that PREP2 is a promising tool for clinical application. The PREP2 combines clinical assessment with the shoulder abduction finger extension (SAFE) test with information about MEP status. For some patients, additional information on either age or the National Institutes of Health Stroke Scale (NIHSS) score is included.

A recent study by Kier et al[27] revealed that knowledge of prognosis seems to be relevant for most therapists in their clinical work. At the same time, prediction models for UL function after stroke are not yet a part of daily practice in Danish stroke rehabilitation.[27] A main obstacle for implementing PREP2 in clinical practice is the fixed time points of the initial assessment with SAFE and TMS, which are days 1 to 3 and 3 to 7, respectively. In several countries, including Denmark, patients needing inpatient rehabilitation are transferred from the acute stroke units to various subsequent rehabilitation services during the first days or weeks poststroke. The stay at the acute stroke unit is usually short, which leaves little time for prognostic evaluation. As most recovery occurs within the first 3 months after stroke, it is essential that all patients are assessed at a fixed point in time after stroke.[28] Based on the clinical reality we experience in our health system we decided to make the prediction 2 weeks poststroke to include as many patients in the subacute phase as possible who had not been available for earlier assessments. Furthermore, this point in time was considered relevant for targeted rehabilitation. Predictions made 2 weeks poststroke may support the choice of adequate therapeutic approaches, can be used inform clinicians and patients about future potential UL function and may influence length of stay. The knowledge obtained may be used to guide choice of UL intervention and treatment. If the PREP2 algorithm could be applied 2 weeks poststroke with satisfactory accuracy, this would facilitate its implementation. The aim of this study was to assess the prognostic accuracy of PREP2 when applied in a subacute neurorehabilitation setting 2 weeks poststroke.

Methods

Study Design

This was a prospective longitudinal study. We followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting observational data and the recommendations for standardized measurement of sensorimotor recovery in stroke trials.[28,29]

Setting and Patients

The study took place at a neurorehabilitation hospital in Denmark. Approximately 500 adult patients with stroke are admitted annually from various stroke units. Patients are admitted if they are considered to benefit from inpatient neurorehabilitation and approximately two-thirds arrive within 14 days poststroke. For the present study patients were included consecutively from June 2018 to October 2019.

Patient inclusion criteria were first or recurrent ischemic or hemorrhagic stroke, admitted to the rehabilitation hospital within 2 weeks poststroke, level of UL function defined as a SAFE score <10, age ≥ 18 years, and ability to cognitively comply with examinations defined by a Functional Independence Measure cognitive score ≥ 11 in combination with the rehabilitation team considering the patient able to participate. Exclusion criteria were subarachnoid hemorrhage or prior UL impairment, for example, from a previous stroke or injury, as that would impede the potential for full recovery. In addition, patients were excluded if the information necessary for prediction could not be obtained at baseline.

Procedure

Included patients were examined according to the PREP2 (Figure 1), and UL function in 1 of 4 categories was predicted.

The first step in the PREP2 is a calculation of the SAFE score by scoring shoulder abduction and finger extension strength separately between a minimum of 0 and a maximum of 5 (best)[8]. The scores are added to form the SAFE score ranging from 0 to 10 (best). The second step depends on the SAFE score. For patients with SAFE score ≥ 5, information on age is used, and a prediction of either Excellent or Good UL function is made. For patients with a

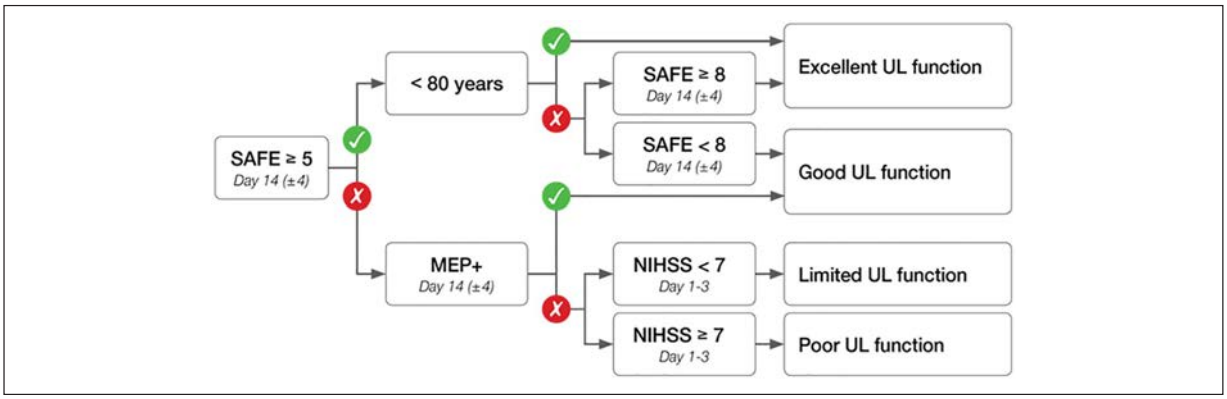


Figure 1. The Predict Recovery Potential algorithm performed 2 weeks poststroke: SAFE, Shoulder Abduction and Finger Extension; <80 y, less than 80 years old; MEP+, motor-evoked potentials present; NIHSS, National Institutes of Health Stroke Scale. Excellent: Potential to make a complete or near complete recovery of hand and arm function within 3 months. Good: Potential to use their affected hand and arm for most activities of daily living within 3 months. Limited: Potential to regain some movement in their hand and arm within 3 months. Poor: Unlikely to regain useful movement in their hand and arm within 3 months.

SAFE score <5, TMS is used to establish MEP status. If MEPs are present (MEP+), the patient is predicted a Good UL function. If MEPs cannot be elicited (MEP−), then a measure of stroke severity, the patient's NIHSS score, is used; and UL function will be predicted as either Limited or Poor, if the NIHSS score is <7 or ≥7, respectively.

UL function was measured with the Action Research Arm Test (ARAT).[28,30-32] The ARAT reflects a broad range of arm and hand activities. Patients can score from a minimum of 0 to a maximum of 57 (best). In line with the PREP2 procedures, the outcome was predicted in 1 of 4 categories, each reflecting a range of scores on ARAT. The category "Excellent" comprises the ARAT scores of 51 to 57, "Good" 34 to 50, "Limited" 13 to 33, and "Poor" 0 to 12.

In the study by Stinear et al,[8] the SAFE score was obtained within 3 days and MEP from 3 to 7 days poststroke. Information on age and NIHSS scores within 3 days poststroke was obtained from medical records. In the present study, both the SAFE score and MEPs were obtained 2 weeks poststroke (see Figure 1). Information on age and NIHSS score or the comparable Scandinavian Stroke Scale (SSS) score was routinely recorded in the acute units within 3 days poststroke and could be collected from the medical record as originally proposed by Stinear et al.[8]

The TMS procedure was performed in line with international recommendations[19] and screening for contraindications and application of TMS were in accordance with the protocol from Stinear et al.[33,34] Absolute contraindications were epilepsy, metal implants in the head, implanted electronics (cardiac pacemaker, defibrillator, cochlear implant, medication pump), skull fracture or serious head injury, brain surgery, and pregnancy.[19]

Patients were seated with the affected UL resting on a table in a relaxed position with elbow flexion. Electromyographic activity was recorded from the first dorsal interosseous and the extensor carpi radialis muscle of the affected UL, using standard surface electrodes (Neuroline 720, Ambu A/S). Recording electrodes were placed in a belly-tendon montage, while the reference electrode was placed over the lateral epicondyle of the humerus. Signals were sampled at 4 kHz, amplified (150 V/V gain), band-pass filtered (10-500 Hz), and acquired with a 16-bit data acquisition board (USB-6341, National Instruments) for offline analysis. The acquired data were visually inspected and stored with a custom-made LabVIEW (National Instruments) software (Mr. Kick, Knud Larsen, Aalborg University, Denmark). Magnetic stimuli consisted of monophasic pulse waveforms that were delivered using a 70-mm figure-of-eight coil connected to a MagStim 200 unit (Magstim Co LtD). The coil was oriented to induce posterior-to-anterior current flow in the ipsilesional M1. Stimulus intensity began at 50% of the maximal stimulator output (MSO), and was increased in 10% MSO steps, delivering approximately 3 to 5 stimuli at each intensity and scalp location. The experimenter moved the coil in approximately 1-cm steps (anterior, posterior, medial, lateral) to find the optimal location for producing MEPs. Stimulus intensity was increased until MEPs were consistently observed in one or both muscles or until 100% MSO was reached. If 100% MSO was reached and no MEPs were observed, the patient was asked to make a firm fist with the nonparetic hand and to attempt to do so with the paretic hand as this may facilitate MEPs.

The patient was classified as MEP+ if either passive or active MEPs were observed with a peak-to-peak amplitude ≥50 μV at consistent latency in response to at least 5 consecutive stimuli. The expected latencies for the MEPs in the first dorsal interosseous muscle were ≈20 to 30 ms, and extensor carpi radialis muscle ≈15 to 25 ms.[19,33,35] If this

criterion was not met with stimuli delivered at 100% MSO intensity, the patient was categorized as MEP−.[33]

Electromyographic recordings were further evaluated offline and MEP status was established by one of the authors (FGA), who was blinded to the results of the clinical assessment.

Supplementary Assessments

To describe the population and enable comparison with other stroke populations a range of supplementary assessments were performed. UL impairment was assessed with the Fugl-Meyer motor assessment upper extremity (FM) at baseline and follow-up.[30,36] FM is found to be reliable and valid.[30,36] Moreover, UL limb pain, light touch, proprioception, neglect, and walking ability were assessed at baseline.

Follow-up Assessment

Patients were tested at 3 months poststroke by experienced research therapists trained in assessment procedures and blinded to both baseline scores and predicted categories. The research therapists were not involved in patient care.

The primary outcome was the achieved UL function in 1 of the 4 categories based on the ARAT scores. ARAT is found to be reliable and valid, and is internationally widely applied and recommended.[28,30] To ensure reliability, the assessors received a thorough introduction on how to administer ARAT and a comprehensive manual was provided based on previous research.[32] Before commencing the study several patients were assessed by all assessors and their results discussed until agreement was achieved. After 3 months, this calibration process was repeated. In cases of doubt on how to score a certain item, the principal investigator was contacted.

Inclusion in the present study did not affect patient rehabilitation or choice of UL treatment. Length of stay, contents and intensity of the rehabilitation were individually organized and determined by the rehabilitation team, patient and relatives. Amount of standard rehabilitation included 45 minutes of physiotherapy and 45 minutes of occupational therapy 5 days a week. For patients with severe brain damage the amount was double. The team members were blinded to predictions and clinical assessments.

Statistical Analysis

The required number of patients in the study was 73 assessed by a power calculation assuming a correct classification rate (CCR) of 75% with a CI 95% of 65% to 85%. A CCR of 75% was chosen as this was in line with the accuracy found in the original PREP2 study.[8] Allowing for a 20% dropout, it was decided to include at least 90 patients.

Data were analyzed with STATA 16. Data were visually inspected to determine the distribution of normality. Continuous baseline characteristics, stroke details, baseline and follow-up scores were summarized by mean, standard deviation (SD), minimum (min), and maximum (max) when normally distributed; otherwise by median, interquartile range (IQR), min, and max. As ARAT is an ordinal scale and data were nonnormally distributed, within group difference from inclusion to follow-up was tested with the nonparametric Wilcoxon signed rank test.

The overall accuracy of the PREP2 was quantified by comparing predicted and actual ARAT categories using the CCR. In addition, CCR, sensitivity and specificity were calculated for each of the four categories. To differentiate between patients who would need additional information on MEP status, CCR was calculated for those with a SAFE score <5 or ≥ 5.

To examine if CCR of the PREP2 obtained 2 weeks poststroke could be improved, a classification and regression tree (CART) analysis was carried out using pruning and cross-validation according to Hastie et al.[37] CART analysis produces a decision tree without the user determining which variable to include or their order in the tree.[37,38] Available for the CART analysis were the components of PREP2: SAFE score, age, NIHSS score, and MEP status. For patients with a SAFE score >5, MEP+ status was assumed in the analysis.

Ethical Considerations

The study was reported to the Danish data protection agency and approved by the regional ethics committee for the Central Denmark Region with the number 628213. All participants provided written informed consent in accordance with the Declaration of Helsinki.

Results

The inclusion criteria were fulfilled by 131 patients of whom 36 were excluded, leaving 95 patients for whom a baseline prediction of UL function could be obtained. Of these, 91 patients were available for follow-up and included in the analysis (Figure 2). Patients' demographic and clinical characteristics are reported in Table 1.

Baseline Algorithm Measures

The SAFE score was obtained 13.4 days after stroke (SD 1.6, min 10, max 18). At baseline, the mean SAFE score was 5 (SD 2.8, min 0, max 9).

Corticospinal tract integrity was examined in 38 of 91 patients (42%), a mean of 13.4 days after stroke (SD 1.7, min 11, max 18). Twenty-six patients were MEP+ and 12 were MEP-. For the latter, the NIHSS scores were included with a median of 13 (IQR 7-15, min 9, max 21).

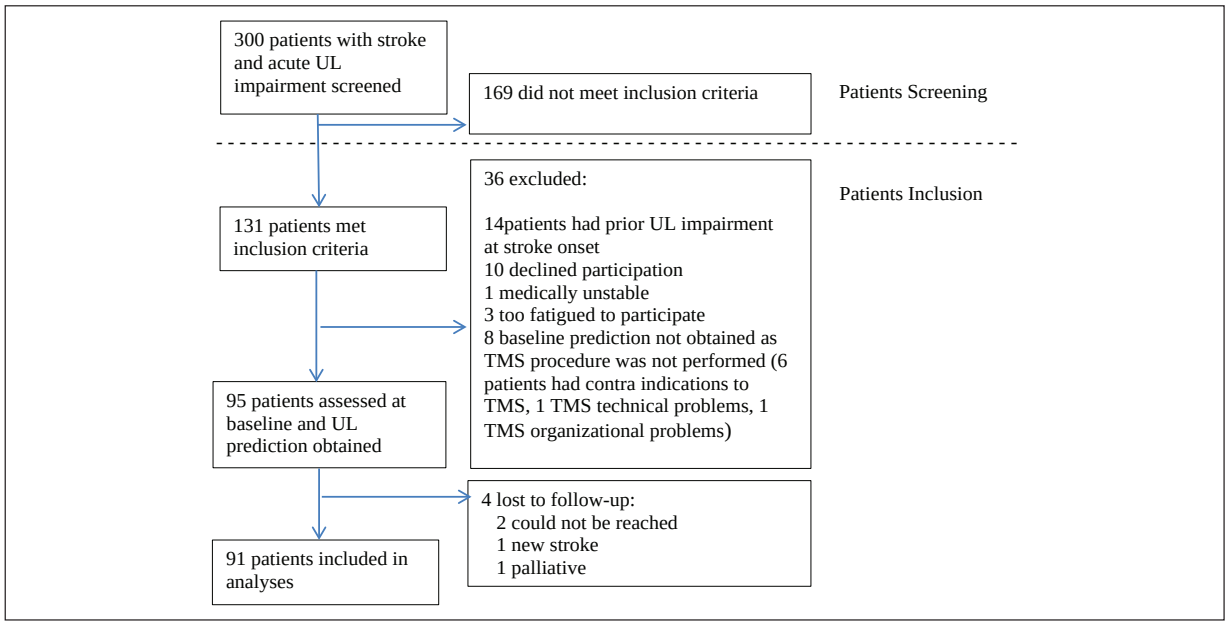


Figure 2. Flowchart of patients included.

At baseline, 9 patients (10%) were predicted Poor UL function at 3 months, 3 patients (3%) were predicted Limited UL function, 29 (32%) were predicted Good UL function, and 50 (55%) were predicted Excellent UL function (Table 2).

Outcome 3 Months After Stroke

Follow-up assessments were performed a mean of 91 days after stroke (SD 3.8, min 84, max 99). The median ARAT score at follow-up was 50 (IQR 33-55, min 0, max 57). The within group increase in ARAT scores from baseline to follow-up was 17 (IQR 3-27, min −4, max 57) and statistically significant ($P < .001$).

Based on the actual ARAT score at follow-up, 12 patients ended in the category Poor, 13 in Limited, 22 in Good, and 44 in Excellent (Table 2). In the category Poor, the median ARAT score was 0 (IQR 0-2, min 0, max 6), in Limited 31 (IQR 24-32, min 17, max 33) in Good 42 (IQR 39-49, min 24, max 50) and in Excellent 55 (IQR 54-56, min 51, max 57).

UL Prediction Accuracy

Overall, CCR was 60% (95% CI 51-71) (Table 3). In 26 of 91 patients (29%), the prediction was too optimistic, and the patients did not achieve the predicted UL function. In 10 of 91 patients (11 %), the prediction was too pessimistic and the actual UL function at 3 months exceeded the predicted function. Most patients (n = 28, 31%) for whom the prediction was inaccurate achieved an actual outcome category adjacent to the predicted category, for example, predicted as Good, but ending up in the outcome category Limited (Table 2).

For each of the 4 categories, CCR was a highest 78% (95% CI 43%-95%) for patients with a prediction of Poor UL function followed by 74% (95% CI 60%-84%) for those predicted Excellent UL function. For patients predicted Good UL function, CCR was 35% (95% CI 20%-53%); and for those predicted Limited UL function, CCR was 33% (95% CI 4%-85%) (Table 3).

For the 53 patients with a SAFE score ≥5, CCR was 74% (95% CI 62%-86%). For the 38 patients with a SAFE score <5, CCR was 42% (95% CI 26%-58%) (Table 3). The low CCR for patients with a SAFE score <5 was mainly due to the 26 patients who were MEP+ and hence predicted Good UL function (Figure 2). In this category, there was considerable variability in outcome categories (Table 2). On the contrary, patients who were MEP− and predicted a Poor UL function generally achieved the outcome category Poor (Table 2).

CART Analysis

The CART analysis produced a decision tree with an overall CCR of 66% (95% CI 56-76) for PREP2 obtained 2 weeks poststroke (Figure 3). The SAFE score was found to be the most important predictor. Patients with a SAFE score ≥5 were predicted either Excellent or Good UL function based on age ≥75 years. For patients with a SAFE score <5, the

Table 1. Participants' Characteristics and Stroke Details (n = 91).[a]

Age, years, mean (SD, min-max)	64 (10.6, 44-91)
Sex, female/male, n (%)	39 (43) / 52 (57)
Stroke type, ischemic/hemorrhagic, n (%)	73 (80) / 18 (20)
Hemisphere of stroke, left/right, n (%)	53 (58) / 38 (42)
Days since stroke, mean (SD, min-max)	13.4 (1.6, 10-18)
Stroke confirmed on imagining, n (%)	90 (99)
Cortical (internal capsule/middle cerebral artery/frontal lobe), n (%)	43 (47)
Subcortical (cerebellum/thalamus/basal ganglia/corona radiata), n (%)	45 (49)
Brain stem (pons/medulla), n (%)	4 (4)
Thrombolysis/thrombectomy,[b] n (%)	32 (35) / 16 (18)
NIHSS score[c] (n = 89), median (IQR, min-max)	9 (6-13, 1-22)
FIM score (n = 86), median (IQR, min-max)	74 (50-89, 24-117)
Premorbid able to walk (±walking aid) (n = 90), n (%)	89 (98)
Premorbid living in own home, n (%)	91 (100)
Premorbid dominant hand right (n = 90), n (%)	76 (84)
First stroke, n (%)	83 (91)
Hypertension, n (%)	43 (47)
Coronary artery disaese, n (%)	17 (19)
Diabetes, n (%)	7 (8)
Other neurological disease(s), n (%)	3 (3)
Current smoker (n = 78), n (%)	30 (38)
BMI, median (IQR, min-max) (n = 83)	27 (24-29, 16-46)
Baseline SAFE[d] score within outcome categories	
Excellent category (n = 44), median (IQR, min-max)	8 (5-9, 0-9)
Good category (n = 22), median (IQR, min-max)	5 (4-6, 1-9)
Limited category (n = 13), median (IQR, min-max)	3 (2-4, 0-6)
Poor category (n = 12), median (IQR, min-max)	1 (0-2, 0-4)
Assessments at baseline	
ARAT score, median (IQR, min-max)	21 (4-41, 0-57)
FM score, median (IQR, min-max)	42 (16-53, 0-66)
Upper extremity pain present[e], n (%)	28 (31)
Upper extremity light touch affected[f], n (%)	42 (46)
Upper extremity proprioception affected[f] (n = 89), n (%)	24 (27)
Visuospatial neglect present[g] (n = 89), n (%)	21 (24)
Independent walking ability at inclusion[h], n (%)	25 (27)
Assesssments at follow-up	
ARAT score, median (IQR, min-max)[i]	50 (33-55, 0-57)
FM score, median (IQR, min- max)[j]	53 (38-62, 0-66)

Abbreviations: IQR, interquartile range; BMI, body mass index; FM, Fugl-Meyer Assessment upper extremity score; NIHSS, National Institute of Health Stroke Scale; FIM, Functional Independence Measure; SAFE, shoulder abduction finger extension; ARAT, Action Research Arm Test.
[a]For all variables, the number of participants was (n) = 91 unless otherwise stated.
[b]Stroke thrombolysis/thrombectomy rates were calculated for patients with ischemic stroke only.
[c]NIHSS performed a mean of 1.4 days (SD 1.2) poststroke. Where the NIHSS was not available, the Scandinavian Stroke Scale (SSS) score was obtained instead and converted into a NIHSS score in accordance with a model developed and validated by Gray et al.[39]
[d]SAFE reported for each of the actual (not predicted) outcome categories, based on the ARAT score at 3 months.
[e]Assessed by a verbal rating scale.
[f]Assessed with Fugl-Meyer sensory assessment upper extremity scale.[40]
[g]Assessed with Star Cancellation Test and Line Bisection Test. Neglect was present if the Line Bisection Test score was ≤7 and/or the Center of Cancellation was >0.083 on the Star Cancellation Test.[41,42]
[h]Assessed with the Functional Ambulation Categories.
[i]The within-group increase in ARAT from baseline to follow-up was a median of 17 and statistically significant ($P < .001$).
[j]The within-group increase in FM from baseline to follow-up was a median of 9 and statistically significant ($P < .001$).

NIHSS score with a cut point of 16 was needed for patients who were MEP+. Patient who were MEP− were predicted Poor UL function. The CART analysis was based on 89 of the 91 patients, as 2 patients did not have a NIHSS score.

For each of the 4 categories, CCR was 80% (95% CI 66%-89%) for the category Excellent, 45% (95% CI 28%-63%) for Good, 60% (95% CI 20%-90%) for Limited, and 67% (95% CI 37%-87%) for Poor. For patients with a SAFE

Table 2. Predicted and Actual Action Research Arm Test (ARAT) Categories and Agreement Between Them.[a]

Predicted ARAT category at baseline	Actual ARAT outcome category at 3 months				
	Excellent	Good	Limited	Poor	Total, n (%)
Excellent	37	10	3	0	50 (55)
Good	7	10	8	4	29 (32)
Limited	0	1	1	1	3 (3)
Poor	0	1	1	7	9 (10)
Total, n (%)	44 (48)	22 (24)	13 (14)	12 (13)	91 (100)

[a]Patients for whom the outcome category was equivalent to the predicted category (n = 55). Patients for whom the outcome category was adjacent to the predicted category (n = 28). Patients for whom the outcome category was two categories away from the predicted category (n = 8). Patients for whom the outcome category was three categories away from the predicted category (n = 0).

Table 3. Accuracy of the Prediction Algorithm.

	CCR, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	CCR for SAFE ≥5, % (95% CI)	CCR for SAFE <5, % (95% CI)
Overall (n = 91)	60 (50-71)				
Excellent (n = 44)	74 (60-84)	84 (70-93)	72 (57-84)	74 (62-86)	
Good (n = 22)	35 (20-53)	46 (24-68)	73 (60-83)		
Limited (n = 13)	33 (4-85)	8 (0-36)	97 (91-100)		42 (26-58)
Poor (n = 12)	78 (43-95)	58 (28-85)	98 (91-100)		

CCR, correct classification rate; SAFE, shoulder abduction finger extension; n, number of patients in outcome category.

score ≥5, CCR was 78% (95% CI 67%-89%); and for patients with a SAFE score <5, CCR was 50% (95% CI 34%-66%).

Discussion

When PREP2 was applied in a subacute rehabilitation setting and the SAFE score and MEP status were obtained day 13 (SD 1.6), the overall CCR was 60% (95% CI 50%-71%). However, CCR differed greatly for each of the 4 UL prediction categories. CCR was low for patients predicted Limited (33%; 95% CI 4%-85%) or Good (35%; 95% CI 20%-53%) UL function. In contrast, CCR was high for patients predicted Poor (78%; 95% CI 43%-95%) or Excellent (74%; 95% CI 60%-84%) UL function. The overall CCR of 60% seems too low to be clinically relevant. However, the high accuracy for patients predicted Poor or Excellent UL function might still be valuable to clinicians and patients.

A CART analysis revealed that by changing the sequence of the algorithm and the cut-points for age and NIHSS, the CCR could be increased to 66%. In the CART model, the cut-point for age was lowered to 75 years and the cut-point for NIHSS score was set to 16 for patients who were MEP+. Despite the overall increase in accuracy, the CCR produced by CART was 67% or lower for 3 of the 4 categories and the CART model did not improve PREP2 used 2 weeks poststroke.

Comparison With Previous Findings

Stinear et al[8] found an overall accuracy of the original PREP2 algorithm of 75%. The overall CCR of 60% (95% CI 50%-71%) found in the present study was significantly lower (15% difference, 95% CI 3%-27%, $P < .001$, chi-square test).

There could be several possible explanations for the lower CCR in the present study compared to the study by Stinear et al.[8] First, although Stinear et al developed the PREP2 algorithm using pruning and cross validation to avoid overfitting, the CCR of 75% was still computed from the training data used for developing it and may hence be difficult to obtain in other data sets.[43,44] Second, the population in the present study differs from the population in the original study who were in the acute to early subacute stage of stroke. Some of these patients would have been either too low- or too high-functioning to be referred to an in-patient rehabilitation ward 2 weeks poststroke. This difference in populations may contribute to the lower CCR.[43] A third explanation for the overall lower CCR in the present study, especially the large number of patients for whom the algorithm was too optimistic, is that the PREP2 algorithm was designed to convert 3- to 7-day information into 3-month predictions. The fact that most spontaneous biological recovery occurs early after stroke, may have resulted in patients closer to their maximally achievable UL function when assessed at day 13 than when assessed day 3 to day 7 poststroke. Consequently, the room for increases in UL function scores declines during the course

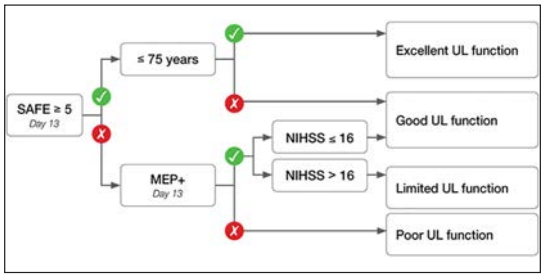


Figure 3. CART model for prediction of UL function. Available for the CART analysis were the PREP2 components: SAFE score, MEP status, age, and the NIHSS score. CART, classification and regression tree; UL, upper limb; PREP2, Predict Recovery Potential algorithm; SAFE, shoulder abduction finger extension; MEP, motor-evoked potential; NIHSS, National Institutes of Health Stroke Scale.

of recovery. In other words, a SAFE score of, for example, 6 on day 3 is more promising than the same score on day 13, and the chance of achieving a Good or Excellent outcome will be greater.

Results from the EPOS cohort study support that timing of the initial assessment can influence the accuracy of UL prediction.[3] In this study, it was shown that UL function could be predicted by shoulder abduction and finger extension measured within 72 hours after stroke.[3] Retesting the model on days 5 and 9 showed that the probability of regaining dexterity remained 98% for those with some finger extension and shoulder abduction, whereas the probability decreased from 25% to 14% for those without this voluntary control.[3] Similarly, our study with assessments 2 weeks poststroke indicates that the longer the severe impairment lasts, the lesser is the chance to regain UL function.

An essential part of PREP2 is the use of MEP as a biomarker. A number studies have shown that the presence of MEP+ days 3 to 7 after stroke indicates a good prognosis and improves the prediction accuracy for UL function.[8,12] However, this was not the case in the present study, where patients with MEP+ on day 13 may or may not improve their UL function. An explanation for this discrepancy in findings may be the aforementioned spontaneous biological recovery. On the other hand, MEP− obtained on day 13 seems just as informative as MEP- obtained on days 3 to 7. When MEP− is obtained within the first days of stroke, 15% to 20% of the patients go on to experience a certain degree of recovery.[45-47] In the present study, 3 of the 12 MEP− patients (25%) had an outcome category exceeding the predicted outcome. These findings suggest that at least some regeneration of the corticospinal tract or some compensation for the loss of its functionality may occur.

The use of TMS to assess the state of corticomotor pathways can provide objective information that cannot be obtained by clinical measures alone.[19] However, reproducibility of TMS should be considered. In the present study, MEP presence was established using a belly-to-tendon placement of electrodes. Generally, this configuration will not only capture MEPs from only one specific muscle, but from several muscles from the forearm.[35,48] Nonetheless, recording muscle activity for the UL in general rather than from a specific hand flexor is in line with the purpose of the present study and was not regarded a limitation. Also, using a setup similar to the one used in this study, a number of studies suggest that MEP amplitudes and latencies to single TMS pulses have adequate reliability for both healthy volunteers[49] and certain patient populations.[50] Nonetheless, it is reported that MEPs are highly susceptible to numerous sources of variability, for example, individual physical features as age and height[51] and MEPs cannot be obtained if patients have contraindications to TMS.[19] In the present study, 6 patients had contraindications to TMS, which prevented to obtain their MEP status and subsequent UL prediction.[51] Finally, TMS can be relatively expensive, and it requires trained staff.

Strengths and Limitations of the Present Study

A strength of the present study is that patients were assessed at set time points in relation to stroke onset to account for underlying recovery processes.[28] To further minimize risk of bias, patients were assessed with FM and ARAT, 2 reliable and valid UL assessments recommended for use in clinical trials.[28] Additional strengths to be noted are that the present study was based on a relatively high number of patients, had very few dropouts, blinded obtainment of MEP status, blinded assessment at follow-up as well as patients and treating therapists being unaware of the UL prediction.

A limitation of the present study was that only 12 patients were MEP− and the predictions Poor or Limited UL function were consequently based on a small number of patients. Longitudinal assessments of corticospinal tract integrity, preferably for the entire group, could have provided more information on potential changes in cortico-motor pathways during the course of recovery, but were beyond the scope of this study. Another limitation of the present study concerns the NIHSS score used to differentiate between the categories Poor and Limited UL function. The NIHSS is not used universally in acute units in Denmark. The NIHSS is mainly used for patients with ischemic stroke, while the SSS is used routinely for all stroke patients. According to Grey et al. the total scores for NIHSS and SSS may be interconverted with good precision.[39] In the present study, the SSS score was converted if the NIHSS score was unavailable. However, these 2 instruments only partly measure the same items.[39] Consequently, the CCR of PREP2 in the present study may have been higher if a true NIHSS score could have been used. However, only 6 patients obtained their prediction based on a converted SSS score.

The patients in the present study were all in need of inpatient rehabilitation, which implies that their impairments were complex and not restricted to the UL. Thus,

the results may not be generalizable to all patients at 2 weeks poststroke.

Other factors than those included in the PREP2 algorithm may influence future UL function. Winters et al[52] found, that patients with paresis mainly restricted to the upper limb, no visuospatial neglect, and sufficient somatosensory function were likely to show at least some return of upper limb capacity at 6 months poststroke. In the present study, many patients had somatosensory deficits, visuospatial neglect, and UL pain. Moreover, the majority were not able to walk independently due to an affected lower limb. The influence of these factors on UL function was not examined in the present study; nor was the influence of length of inpatient rehabilitation, UL treatment, or dose, or treatment modality.

Clinical Implications

Based on the present study, the PREP2 algorithm should not be implemented in the clinic when the SAFE score and MEP status are obtained 2 weeks poststroke. However, components from the PREP2 may be used for certain patients. Patients with a SAFE score of at least 5 who are younger than 80 years will most probably achieve an excellent UL function at 3 months. Furthermore, in patients with a SAFE score <5 who are MEP−, no useful UL function can be expected and UL rehabilitation may focus on preventing pain and learning compensatory strategies. Other factors that may improve prediction of UL function 2 weeks after stroke should be examined.

Acknowledgments

We thank the patients and staff at Hammel Neurorehabilitation Centre for participation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Camilla Biering Lundquist https://orcid.org/0000-0001-6836-2233

References

1. Krishnamurthi RV, Ikeda T, Feigin VL. Global, regional and country-specific burden of ischaemic stroke, intracerebral haemorrhage and subarachnoid haemorrhage: a systematic analysis of the Global Burden of Disease Study 2017. *Neuroepidemiology*. 2020;54:171-179.
2. Persson HC, Parziali M, Danielsson A, Sunnerhagen KS. Outcome and upper extremity function within 72 hours after first occasion of stroke in an unselected population at a stroke unit. A part of the SALGOT study. *BMC Neurol*. 2012;12:162.
3. Nijland RH, van Wegen EE, Harmeling-van der Wel BC, Kwakkel G; EPOS Investigators. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery: early prediction of functional outcome after stroke: the EPOS cohort study. *Stroke*. 2010;41:745-750.
4. Kwakkel G, Kollen BJ, van der Grond J, Prevo AJH. Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. *Stroke*. 2003;34:2181-2186.
5. Faria-Fortini I, Michaelsen SM, Cassiano JG, Teixeira-Salmela LF. Upper extremity function in stroke subjects: relationships between the international classification of functioning, disability, and health domains. *J Hand Ther*. 2011;24:257-265.
6. Veerbeek JM, Kwakkel G, van Wegen EE, Ket JCF, Heymans MW. Early prediction of outcome of activities of daily living after stroke: a systematic review. *Stroke*. 2011;42:1482-1488.
7. Stinear CM, Byblow WD. Predicting and accelerating motor recovery after stroke. *Curr Opin Neurol*. 2014;27:624-630.
8. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA. $PREP_2$: a biomarker-based algorithm for predicting upper limb function after stroke. *Ann Clin Transl Neurol*. 2017;4:811-820.
9. Kwah LK, Herbert RD. Prediction of walking and arm recovery after stroke: a critical review. *Brain Sci*. 2016;6:E53.
10. Buch ER, Rizk S, Nicolo P, Cohen LG, Schnider A, Guggisberg AG. Predicting motor improvement after stroke with clinical assessment and diffusion tensor imaging. *Neurology*. 2016;86:1924-1925.
11. Winters C, van Wegen EE, Daffertshofer A, Kwakkel G. Generalizability of the proportional recovery model for the upper extremity after an ischemic stroke. *Neurorehabil Neural Repair*. 2015;29:614-622.
12. Byblow WD, Stinear CM, Barber PA, Petoe MA, Ackerley SJ. Proportional recovery after stroke depends on corticomotor integrity. *AnnNeurol*. 2015;78:848-859.
13. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA. Proportional motor recovery after stroke: implications for trial design. *Stroke*. 2017;48:795-798.
14. Gebruers N, Truijen S, Engelborghs S, De Deyn PP. Prediction of upper limb recovery, general disability, and rehabilitation status by activity measurements assessed by accelerometers or the Fugl-Meyer score in acute stroke. *Am J Phys Med Rehabil*. 2014;93:245-252.
15. Stinear CM. Prediction of motor recovery after stroke: advances in biomarkers. *Lancet Neurol*. 2017;16:826-836.
16. Hayward KS, Schmidt J, Lohse KR, et al. Are we armed with the right data? Pooled individual data review of biomarkers in people with severe upper limb impairment after stroke. *Neuroimage Clin*. 2016;13:310-319.

17. Kim B, Winstein C. Can Neurological biomarkers of brain impairment be used to predict poststroke motor recovery? A systematic review. *Neurorehabil Neural Repair*. 2017; 31:3-24.
18. Bernhardt J, Borschmann K, Boyd L, et al. Moving rehabilitation research forward: Developing consensus statements for rehabilitation and recovery research. *Int J Stroke*. 2016;11:454-458.
19. Groppa S, Oliviero A, Eisen A, et al. A practical guide to diagnostic transcranial magnetic stimulation: report of an IFCN committee. *Clin Neurophysiol*. 2012;123: 858-882.
20. Boyd LA, Hayward KS, Ward NS, et al. Biomarkers of stroke recovery: consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. *Neurorehabil Neural Repair*. 2017;31:864-876.
21. Prabhakaran S, Zarahn E, Riley C, et al. Inter-individual variability in the capacity for motor recovery after ischemic stroke. *Neurorehabil Neural Repair*. 2008;22:64-71.
22. Katrak P, Bowring G, Conroy P, Chilvers M, Poulos R, McNeil D. Predicting upper limb recovery after stroke: the place of early shoulder and hand movement. *Arch Phys Med Rehabil*. 1998;79:758-761.
23. Smania N, Paolucci S, Tinazzi M, et al. Active finger extension: a simple movement predicting recovery of arm function in patients with acute stroke. *Stroke*. 2007;38: 1088-1090.
24. Byblow WD, Stinear CM. It is difficult to make predictions, especially about the future. *Stroke*. 2017;48:3187-3188.
25. Stinear CM, Barber PA, Petoe M, Anwar S, Byblow WD. The PREP algorithm predicts potential for upper limb recovery after stroke. *Brain*. 2012;135(pt 8):2527-2535.
26. Stinear CM, Byblow WD, Ackerley SJ, Barber PA, Smith MC. Predicting recovery potential for individual stroke patients increases rehabilitation efficiency. *Stroke*. 2017;48:1011-1019.
27. Kiaer C, Lundquist CB, Brunner I. Knowledge and application of upper limb prediction models and attitude toward prognosis among physiotherapists and occupational therapists in the clinical stroke setting. *Top Stroke Rehabil*. Published online June 25, 2020. doi:10.1080/10749357.20 20.1783915
28. Kwakkel G, Lannin NA, Borschmann K, et al. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Int J Stroke*. 2017; 12:451-461.
29. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12:1495-1499.
30. Platz T, Pinkowski C, van Wijck F, Kim IH, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. *Clin Rehabil*. 2005;19:404-411.
31. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *Int J Rehabil Res*. 1981;4:483-492.
32. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair*. 2008;22:78-90.
33. PRESTO. TMS instructions for PREP2. Accessed August 25, 2020. https://presto.auckland.ac.nz/tms-instructions-for-prep2/
34. PRESTO. Safety checklist for TMS. Accessed August 25, 2020. https://presto.auckland.ac.nz/safety-checklist-for-tms/
35. Morita H, Baumgarten J, Petersen N, Christensen LO, Nielsen J. Recruitment of extensor-carpi-radialis motor units by transcranial magnetic stimulation and radial-nerve stimulation in human subjects. *Exp Brain Res*. 1999;128:557-562.
36. Lundquist CB, Maribo T. The Fugl-Meyer assessment of the upper extremity: reliability, responsiveness and validity of the Danish version. *Disabil Rehabil*. 2017;39:934-939.
37. Hastie T, Tibshirani R, & Friedman JH, eds. Tree-based methods. In: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer; 2009: 305-317.
38. Batterham M, Neale E, Martin A, Tapsell L. Data mining: Potential applications in research on nutrition and health. *Nutr Diet*. 2017;74:3-10.
39. Gray LJ, Ali M, Lyden PD, Bath PM; Virtual International Stroke Trials Archive Collaboration. Interconversion of the National Institutes of Health Stroke Scale and Scandinavian Stroke Scale in acute stroke. *J Stroke Cerebrovasc Dis*. 2009;18:466-468.
40. Sullivan KJ, Tilson JK, Cen SY, et al. Fugl-Meyer assessment of sensorimotor function after stroke: standardized training procedure for clinical practice and clinical trials. *Stroke*. 2011;42:427-432.
41. Rorden C, Karnath HO. A simple measure of neglect severity. *Neuropsychologia*. 2010;48:2758-2763.
42. Nijboer TCW, Winters C, Kollen BJ, Kwakkel G. Impact of clinical severity of stroke on the severity and recovery of visuospatial neglect. *PLoS One*. 2018;13:e0198755.
43. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
44. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ*. 2009;338:b604.
45. Zarahn E, Alon L, Ryan SL, et al. Prediction of motor recovery using initial impairment and fMRI 48 h poststroke. *Cereb Cortex*. 2011;21:2712-2721.
46. van Kuijk AA, Pasman JW, Hendricks HT, Zwarts MJ, Geurts ACH. Predicting hand motor recovery in severe stroke: the role of motor evoked potentials in relation to early clinical assessment. *Neurorehabil Neural Repair*. 2009;23:45-51.
47. Hendricks HT, Pasman JW, Merx JL, van Limbeek J, Zwarts MJ. Analysis of recovery processes after stroke by means of transcranial magnetic stimulation. *J Clin Neurophysiol*. 2003;20:188-195.

48. Butler AJ, Wolf SL. Transcranial magnetic stimulation to assess cortical plasticity: a critical perspective for stroke rehabilitation. *J Rehabil Med*. 2003(41 suppl):20-26.
49. Goldsworthy MR, Hordacre B, Ridding MC. Minimum number of trials required for within- and between-session reliability of TMS measures of corticospinal excitability. *Neuroscience*. 2016;320:205-209.
50. Fried PJ, Jannati A, Davila-Perez P, Pascual-Leone A. Reproducibility of single-pulse, paired-pulse, and intermittent θ-burst TMS measures in healthy aging, type-2 diabetes, and Alzheimer's disease. *Front Aging Neurosci*. 2017;9:263.
51. Cantone M, Lanza G, Vinciguerra L, et al. Age, height, and sex on motor evoked potentials: translational data from a large Italian cohort in a clinical environment. *Front Hum Neurosci*. 2019;13:185.
52. Winters C, Kwakkel G, Nijland R, van Wegen E; EXPLICIT-Stroke Consortium. When does return of voluntary finger extension occur post-stroke? A prospective cohort study. *PLoS One*. 2016;11:e0160528.

Paper 2

Prediction of upper limb use three months after stroke: A prospective longitudinal study

ABSTRACT

Purpose

To examine if UL impairment two weeks after stroke can predict real-life UL use at three months. Furthermore, to identify additional predictors of UL use and characteristics of patients who does not achieve normal UL use.

Methods

This study included patients with stroke ≥ 18 years. UL impairment was assessed by Fugl-Meyer upper extremity motor assessment (FM). Use ratio was assessed with accelerometers at three months. The association between FM score and UL use ratio was investigated with linear regression models and adjusted for secondary variables. Non-normal use was assessed by logistic regression.

Results

Eighty-seven patients were included. FM score predicted 38% of the variance in UL use ratio. An mulivariate regression model predicted 55%, and the significant predictors were FM, motor-evoked potential (MEP) status and neglect. Non-normal use could be predicted with a high accuracy based on MEP and/or neglect. For patients with MEPs and without neglect, non-normal use could be predicted with a sensitivity of 0.80 and a specificity of 0.83.

Conclusion

Better baseline function of the paretic UL predicted increased use of the arm and hand in daily life. Non-normal UL use could be predicted reliably based on the absence of MEPs and/or presence of neglect.

KEYWORDS: stroke, rehabilitation, upper extremity, accelerometers, prediction, biomarker, neglect, prognosis

Introduction

A major goal of upper limb rehabilitation after stroke is to facilitate use of the paretic arm in daily life activities. To be truly meaningful, improvements in paretic upper limb (UL) impairment should be translated into increased UL use in daily life and resemble pre-stroke levels as closely as possible [1,2].

The International Classification of Functioning (ICF) framework distinguishes between the capacity for activity measured in a structured environment with clinical tests and performance of activity in daily life, i.e., what a person actually does in an unstructured environment [3]. Several studies have shown that whereas UL capacity and UL performance are related, UL performance is not exclusively a function of UL capacity; it may be influenced by several other factors, e.g. motivation [1], attention or arousal [2]. Moreover, learned non-use of the paretic arm can reduce the level of use [1,2,4].

During the past decade, several models for prediction of UL function have been proposed [5-11]. Five prospective longitudinal studies showed that most patients recover 70-80% of their maximum possible UL motor function within 3-6 months after stroke [5-8,11]. The use of transcranial magnetic stimulation (TMS), contributes to prediction accuracy in patients with severe paresis [10,12-15]. Patients in whom TMS elicits motor-evoked potentials (MEPs) in muscles of the paretic limb generally achieve better and faster motor recovery than patients without MEPs [7,10,16]. In a recent review, MEPs at rest was the only biomarker predicting motor outcome in individuals with severe UL impairment following stroke [12].

Whereas the association between UL function and UL use has been examined in several studies [17-20], predictive factors for UL use have been only sparsely investigated. In a study by Buxbaum et al., 20 chronic stroke survivors with mild to moderate UL impairments characterized by Fugl-Meyer Motor Assessment (FM) were assessed for learned non-use using a modified version of the Actual Amount of Use Test (AAUT). The AAUT measures the disparity between the amount of use in spontaneous versus forced conditions. It was shown that FM scores and non-

lateralized attention and arousal predicted the degree of non-use [2]. The only study identified that explored long-term predictors of UL use shortly after stroke was by Rand & Eng [21]. The authors assessed real-life UL use one year after stroke in subjects who used wrist-worn accelerometers. Their study revealed that better UL function at discharge predicted increased UL use after one year. However, UL use was still reduced compared with healthy controls, even in patients with only mild impairments.

Wrist-worn accelerometry is the method of choice to assess real-life UL in non-disabled adults and adults with stroke [22-24]. Previous accelerometer studies have shown that in non-disabled adults, dominant and non-dominant ULs are active to a similar degree, and most activities are performed bimanually [18,25,26].

The dual aim of this study was, first, to examine if UL impairment assessed by FM two weeks after stroke can predict real-life UL use three months after stroke; second, to identify potential additional predictors of UL use, and establish characteristics of patients who did not achieve normal UL use.

Method

Study design

This was an observational prospective cohort study. We followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting observational data and the recommendations for standardized measurement of sensorimotor recovery in stroke trials [27,28]. Data from the same cohort included in a previous study were used [29].

Setting and patients

Patients were included at a neurorehabilitation hospital in Denmark. The inclusion criteria were first or recurrent stroke, admission within two weeks after stroke, impaired UL function, age ≥18 years and ability to cognitively comply with

examinations. The exclusion criteria were subarachnoid haemorrhage or prior UL impairment, e.g. from a previous stroke, which would impede the potential for full UL recovery. In addition, patients were excluded if accelerometer data were unavailable. Inclusion in the present study did not affect patient rehabilitation or choice of UL treatment.

Procedure

Patients' demographics and medical information was extracted from the patients' medical records. This included information on sex, age and Functional Independence Measure (FIM) score at inclusion. Baseline assessments were performed two weeks after stroke and follow-up assessments three months after stroke.

Baseline assessments

Impairment of the paretic UL was assessed with the FM.[30,31] The FM contains 33 items, each scored on a three-point ordinal scale from 0-2, yielding a maximal total score of 66 points. The clinimetric properties of FM are well established. To ensure reliability, a scoring manual was used [30-32]. Patients were examined by the first author, who was not involved in patient care.

UL function was assessed with the Action Research Arm Test (ARAT) [28,30,33,34]. The ARAT reflects a broad range of arm and hand activities, and scores range from a minimum of 0 to a maximum of 57 (best). The FM and ARAT are internationally recommended for research studies [28]. The Shoulder Abduction Finger Extension (SAFE) score was used to score shoulder abduction and finger extension strength separately using the medical research council grades for limb power. The two sub-scores were added to form the SAFE score ranging from 0 to 10 (best) [10].

In patients with a SAFE score < 5, cortico-spinal tract integrity was examined. Using TMS with the first dorsal interosseous and the extensor carpi radialis muscle of the affected UL as target muscles, we established whether MEP was present. Procedure details have been described previously [29]. As voluntary finger movements reflect at least some cortico-spinal tract integrity, MEP was not assessed but assumed to be present in patients with a SAFE ≥ 5.

Light touch and proprioception were assessed by the Fugl-Meyer Sensory Assessment Scale Upper Limb [35], and bilateral stimulation was examined in the palmar surface of the hand in accordance with the Nottingham Sensory Assessment Scale [35]. Two-point discrimination (twopd) was measured with a Discriminator at the pulp of the index finger from 2-15 mm, with higher scores indicating a lower discriminative acuity. Twopd was considered affected if the discrimination ability was above the thresholds found for healthy age-matched controls.[36] In line with a previous study, a score of 16 was given if twopd was absent [37]. Pain was rated by the patients from 0- 10 (worst pain) on a numerical rating scale. Neglect was assessed with the Star Cancellation Test and the Line Bisection Test. Neglect was present if the Line Bisection Test score was ≤ 7 and/ or the centre of cancellation was above 0.083 on the Star Cancellation Test [38,39]. Inferior subluxation in the glenohumeral joint was assessed by palpation of the subacromial space [40], and walking ability was assessed with the Functional Ambulation Classification [41].

Follow-up assessments

The primary outcome was real-life UL use expressed as the use ratio between paretic and non-paretic UL measured with accelerometers (ActiGraph GT3X+ Activity Monitors). The validity and reliability for wrist-worn accelerometry are well-established [23,42].

At three months after their stroke, most patients were at home. A research therapist delivered the accelerometers to the patients and provided instructions in how to don the accelerometers. The accelerometers had Velcro fastenings for easy handling, but if the patient could not don the accelerometer without help, arrangements were made with either a relative or a home carer. The accelerometers should be worn on both wrists from 08:00 to 20:00 on a typical day within a week after follow-up assessment. The patients were requested not to change their behaviour or try to increase their UL activity but simply wear the accelerometers while they went about their normal daily routines. Previous research has shown that activity levels do not increase in response to wearing accelerometers [43]. The accelerometers were returned to the research lab in a prepaid envelope.

Accelerations were recorded along three axes at 50 Hz and converted into activity counts (0.001664g/count) in accordance with previous studies [25]. ActiLife 6 was used to visually inspect data to ensure that the accelerometers functioned properly during the recording period. The relevant 12-hour intervals were isolated in Matlab and exported to STATA16. The following parameters were calculated using the approach described by Bailey et al. [25]: hours of paretic UL and non-paretic UL use, use ratio, hours of bilateral UL use, magnitude ratio and bilateral magnitude.

Activity counts were combined across the three axes to create a vector magnitude $\sqrt{x2 + y2 + z2}$ for each second of data. Total hours of paretic and non-paretic UL use is the total time in hours that the specific limb was active during a 12-hour period. The use ratio was calculated by dividing total hours of paretic UL use by total hours of non-paretic use. A use ratio of 0.5 indicates that the paretic UL is active 50% of the time the non-paretic was active. The use ratio was used as the primary outcome as it is independent on varying activity levels between different people [24]. The bilateral magnitude quantifies the intensity of activity across both ULs, and the magnitude ratio quantifies the contribution of each UL to activity for every second of data [22,25].

Data analysis

Data were analysed with STATA 16. Demographic characteristics, clinical measures and accelerometer outcomes were summarized by mean and standard deviation (SD) when normally distributed; otherwise by median and interquartile range (IQR). Demographic and clinical characteristics of the patients who were unavailable for the three-month follow-up were compared with those available to determine if the difference was statistically significant. The unpaired t-test or the Wilcoxon rank sum test was used for continuous data and the chi2 test for dichotomous data. Accelerometer data were displayed for the whole group and in line with a recent study in three categories, each reflecting a range of scores on FM at baseline.[18] The category "Severe" comprised the FM scores of 0-22, "Moderate" 23-50 and "Mild" 51-66.

Prediction of use ratio

Several regression models were created. The first model, model 1, was a linear regression model to assess the strength of the (univariate) association between baseline FM score and UL use ratio at three months. In model 2, the association between FM at baseline and use ratio was adjusted for other secondary variables chosen a priori, based either on the results of previous studies or on clinical reasoning. The independent variables and their distribution were assessed and the relationship between them was assessed one at a time.

The following secondary variables were chosen a priori: MEP status (MEP present/ absent). Neglect (present /absent). Dominant UL affected as previous research has demonstrated that dominant side affected may be associated with a better UL stroke recovery [21,25,44]. Twopd (affected/not affected) as previous research has shown that this was a predictor for future UL function [37]. The FIM score, reflecting the need for assistance in daily life activities, was entered as a continuous variable from min 18 to max 126. Gender was included as older women use their dominant hand in daily life more than older men [44]. Lastly, severity of pain was included as a continuous score of 0-10.

In model 3, the contribution of the biomarker MEP was assessed by removing MEP status from model 2 and comparing the fit of the model with and without MEP. Finally, to assess the strength of each potential predictor, univariate regression between each of the predictor variables and use ratio was performed. All necessary assumptions for generalized linear models, including linearity, equality of variance and normality of errors, were visually inspected for all models and found adequate. Presence of multi-linearity was examined by the Variance Inflation Factor for each independent variable. Using a conservative approach, VIF below 3 were accepted [45]. Multi-linearity was not present.

The ability of the models to predict use ratio was assessed by the size of the adjusted R^2. The contribution of each individual predictor in the model was assessed from the significance level, size of p-value and the size of the β-coefficient including the 95% confidence interval (CI) [46].

To assess the ability of the models to predict future use ratio for an individual patient, the 95% prediction interval (PI) for the regression line was calculated based on the SD for the adjusted R^2 (PI = ± 1.96 * SD). The PI is an estimate of the interval in which a future observation of UL use ratio will fall, with 95% probability, given what has already been observed.

Normal and non-normal use ratio

Use ratio was dichotomized into normal and non-normal using a threshold based on an established reference value from a study with 74 community-dwelling adults [26]. In the reference population, the mean use ratio was 0.95± SD 0.06, range 0.79-1.1 [26]. In the present study, the lower limit of the PI interval for the reference value was calculated (0.95-1.96* 0.06=0.83) and used as a conservative threshold for normal use ratio.

A multivariate logistic regression with the outcome use ratio (normal /non-normal) and the variables FM, MEP status, neglect, dominant UL affected, twopd and FIM was performed. To maintain adequate power for the statistical analysis, we complied with the events per variable rule, which calls for at least ten outcomes for each variable in the regression model [47,48]. A receiver-operating curve (ROC) of the logistic model was graphically displayed, and a two-way contingence table was used to identify the cutpoint with the highest sensitivity and specificity values.

Ethical considerations

All patients provided written informed consent in accordance with the Declaration of Helsinki. The study was reported to the Danish data protection agency and approved by the Regional Ethics Committee for the Central Denmark Region (record. no. 628213).

Results

From June 2018 to October 2019, a total of 103 patients met the inclusion criteria and 87 patients were eligible for the final analysis (see figure 1 flow diagram for details).

Insert figure 1. Flow Chart of Patients Included around here

Patients' demographic and clinical characteristics are reported in table 1. The median FM score at baseline was 17 (IQR 14- 53, min 0 max 66), reflecting a broad range of UL impairment. The 16 patients not included in the data analysis were not statistically significantly different on any baseline characteristics or baseline assessment (see table 1).

Insert table 1. around here

Upper limb use

The use ratio was 0.7 (IQR 0.6- 0.9) (see table 2). The median non-paretic unilateral UL activity was 2.1 hours (IQR 1.4- 2.8) and three times as high as the unilateral paretic UL activity. Bimanual UL activity was 3.0 hours (IQR 1.9- 4.0), and total UL activity was 5.8 (IQR 4.8- 7.2).

When accelerometer parameters were examined according to the severity of initial UL impairment, non-paretic unilateral activity decreased and paretic UL activity increased with decreasing impairment. Bimanual activity, total UL activity, use ratio and bilateral vector magnitude also increased with improving UL function. The magnitude ratio was a median of -3.8 for patients with severe UL impairment, reflecting primarily non-affected UL use, whereas it was -1.0 for patients with mild UL impairment, reflecting a more equal contribution of both limbs to an activity.

Insert table 2 around here

Insert figure 2 Association between FM at Baseline and Use Ratio at Three Months After Stroke around her

A linear regression (table 3, Model 1) demonstrated that the FM score at baseline was a statistically significant predictor of use ratio at three months with a β of 0.008 (95% CI 0.006- 0.010), P<0.0001. FM explained 0.38 of the variation in use ratio. The association between FM scores at baseline and use ratio at three months is displayed in figure 2.

When secondary variables were entered into a multiple regression model (Model 2), data from 74 patients were included as data for one or more variables were missing for 13 patients. In model 2, R^2 improved to 0.55, an improvement of 0.17, reflecting that the model now explained a higher percentage of the use ratio.

The statistically significant predictors were FM, MEP status and neglect. The β-slope for FM was 0.006 (95% CI 0.003-0.009, P=0.000*); and for every FM score higher a patient was at baseline, use ratio would be a mean 0.006 higher. With 95% accuracy, the true mean would fall in the 0.003-0.009 interval. The β-coefficient for MEP status was 0.222 (95% CI 0.069- 0.376, P=0.005*), and a patient who was MEP+ at baseline achieved a use ratio that was 0.222 higher than a person who was MEP-. The β-coefficient for neglect was -0.128 (95% CI 0.240-0.016, P=0.025*), and a patient who had neglect achieved a use ratio that was 0.128 lower than a person without neglect. The 95% PI for the expected use ratio was ±0.348.

In Model 3, the biomarker MEP was removed and the adjusted R^2 decreased to 0.458, which was 0.09 lower than in model 2 with MEP included. The 95% PI for the expected use ratio in model 3 was ±0.397. The univariate linear regressions of each of the potential individual predictors showed that all secondary variables except pain were independent predictors (table 3).

Insert table 3 around here

Characteristics of patients who did not achieve normal use ratio

When use ratio was dichotomized at a threshold of 0.83, a total of 30 (34%) patients were classified as having a "normal use ratio" and 57 (66%) as having a "non-normal use ratio" at three months.

Visual inspection revealed that none of the nine patients who had MEP- achieved a normal use ratio (figure 3a). Accordingly, 22 of the 23 patients with neglect did not achieve a normal use ratio (figure 3b). Two patients had both MEP- and neglect, seven patient had MEP- only, and 21 patients had neglect only.

Insert figure 3a. Association between MEP status at Baseline and Use Ratio at Three Months After Stroke around her

Insert figure 3b. Association between neglect at Baseline and Use Ratio at Three Months After Stroke around her

For the remaining patients, all with MEP+ and without neglect, multivariate logistic regression was conducted to assess how well the variables FM, dominant side, twopd and FIM could predict non-normal use ratio. Data from 48 of 57 possible patients were included as nine patients had missing data for one of the variables.
Significant predictors of non-normal use ratio were FM and dominant UL affected. The β for FM was 0.928 (95% CI 0.890-0.980, P=0.007*), and β for dominant UL affected was 0.113 (95% CI 0.023-0.570, P=0.008*). This means that the odds for achieving a non-normal use ratio decreased by 0.07 (7%) for each FM score higher at baseline. For patients whose dominant UL was affected, the odds of achieving non-normal use was 0.89 (89%) lower.

FIM and twopd did not significantly contribute to the prediction of non-normal use ratio (P=0.757 and P=0.079). The ROC based on the multivariate logistic regression (figure 4) revealed an AUC of 0.84 (95% CI 0.73- 0.96). The optimal cut point for prediction of non-normal use ratio for patients with MEP and without neglect was 0.55 with a sensitivity of 0.80 (95% CI 0.61- 0.91) and a specificity of 0.83 (95% CI 0.63- 0.93).

Insert figure 4. Roc of Sensitivity and Specificity for Prediction of not Achieving a Normal Use Ratio around here

Discussion

In the present study, the FM score predicted 38% of the variance in UL use ratio at three months. An multivariate regression model predicted 55%. Statistically significant predictors of use ratio were FM, MEP status and neglect.

Even though only some variables were statistically significant in the multivariate regression models, univariate regressions analysis showed that all predictor variables except pain were independent predictors of UL use ratio. This is unsurprising as a significant univariate predictor may become non-significant, in the presence of other independent variables [46].

Although a significant amount of use ratio can be explained by the three regression models, the 95% PI of the regression lines was wide. Due to the wide prediction intervals, all models seem unsuited for individual prediction of future UL use. However, non-normal use could be predicted with high accuracy based on MEP- and/or neglect. For patients with MEPs and without neglect, a logistic regression revealed that a non-normal use ratio could be predicted with a sensitivity of 0.80 and a specificity of 0.83. Consequently, we can - with some certainty - tell will not achieve a normal UL use ratio.

Our results on UL use resemble those reported in studies on prediction of UL function. Our own work based on the same dataset indicated that UL function was difficult to predict two weeks after stroke, but the absence of MEPs in patients with severe paresis reliably resulted in poor motor function after three months [29]. Patients with severe paresis or paralysis have been found to not recover proportionally [6,7]. This emphasizes the importance of cortico-spinal tract integrity, which is also reflected in our results for UL use. Thus, prediction accuracy could be substantially increased

by adding MEP status to the regression analysis. As suggested by Stinear et al., information on cortico-spinal integrity seems to be an indispensable component for prediction of UL function in patients with severely impaired UL function [10] and hence also for prediction of UL use. The value of MEP status as a supportive prognostic tool has been supported by other research, though support is not always unambiguous [7,10,12-15,49,50].

The third most important negative predictor was neglect, suggesting that neglect is a major contributor to not using the affected UL. A recent review indicated that neglect is associated with poor UL recovery [51]. This is hardly astonishing; nevertheless, it is rarely explicitly addressed in motor rehabilitation programs. Neglect is rather frequent, presenting in roughly a quarter of the patients in our sample, and in 30% according to a recent review [52].

In line with the few previous studies, the most significant individual predictor of UL use in our study was UL function at baseline [2,21]. Rand et al. assessed patients at discharge from rehabilitation and one year after stroke [21]. They found that ARAT and grip strength combined with age were significant predictors of affected UL use assessed with accelerometers and Motor Activity Log. Compared with all regression models presented by Rand et al., our model 2 explained more of the variation in use ratio. Contrary to Rand et al., we found statistically significant univariate prediction of gender and dominant UL affected along with a range of other individual predictors. In a recent study by Buxbaum et al., the authors found that FM and attention/ arousal predicted non-use [2]. However, they did not predict future UL use but assessed the association between FM and use at the same point in time. Moreover, the assessment was not performed in a real-life setting but by means of a clinical test. The mentioned studies, including our own, indicate that UL function is a prerequisite to UL use; however, UL use is not an imperative consequence of good UL function. This disparity has also been confirmed by other studies in which the association of UL function and use was examined [17,53,54].

In the present study, it was not possible to establish a FM threshold for normal use.

This runs counter to a study by Schweighofer et al. who found that above a functional threshold, UL use improves [55]. However, their study was based on data from the EXCITE trial [56] in which a selected group of participants were included 3-9 months after stroke. Moreover, the study did not assess daily use in an unstructured environment.

The main outcome in the present study was the use ratio between affected and unaffected UL. A growing body of research supports the use of accelerometry as a valid and reliable tool to assess real-life use [23]. There is no gold standard for which accelerometer outcome best expresses UL use [24]. Other parameters, such as unilateral paretic activity and bimanual UL activity, provide valuable insights into UL use. However, the use ratio is independent of varying activity levels between different people and has been recommended owing to its clear clinical relevance in stroke rehabilitation populations with asymmetric effects on the limbs [24]. Still, it has to be taken into account that movements registered with accelerometers are not completely identical to voluntary functional movements. A recent study by Lum et al. found that the amount of functional movements based on activity counts tends to be overestimated [57]. Nevertheless, for the time being, accelerometry seems to be the closest approximation to real-life use available.

The strengths of the present study encompass the inclusion of a substantial number of patients with a broad array of UL impairments and a prospective longitudinal study design with the predictor variables collected at an earlier point in time than the outcome variable [46]. The predictor variables were chosen a priory to secure a theory-founded selection of variables. The choice of predictor variables was therefore not based on univariate analysis of their association with exposure and outcome. This reduces the risk of including variables that are statistically significant by change or of discarding variables that may be statistically significant in a larger sample, as the size of the p-value depends on the sample size [46].

Some limitations of this study should be mentioned. MEP status was an important predictor but was only examined in 34 patients with SAFE < 5. MEP+ was assumed to

be present in all patients with SAFE ≥ 5. Thus, we cannot definitely know if MEP was present. Nevertheless, it seems reasonable to assume that the corticospinal tract is at least partly intact in patients with active movement of the paretic UL as presumed in a prediction algorithm by Stinear et al. [10]. Earlier research supports strong correlations between CST integrity and motor function [58]. The first regression model was based on data from 87 patients. However, in the following regressions, fewer participants were included as some had missing data on one or more variables. Thus, some of the regressions are based on a limited number of patients.

Other limitations concern the practical issues of accelerometry measures. The patients had a relatively short wearing time of only 12 hours, and had to don and doff the accelerometer by themselves. Some might have been wearing the accelerometers for less time than requested. However, data were visually inspected and excluded if activity was deemed to be insufficient. Furthermore, compared with other accelerometer outputs, the use ratio is less likely to be affected by wearing time. Wearing visible accelerometers may have encouraged our patients to increase UL use. According to a recent study, patients do not increase their physical activity in response to wearing accelerometers [43]. Nor does it matter what day of the week they chose to wear the accelerometers as physical activity levels do not to differ appreciably between weekends and weekdays [43]. Still, if a patient with severe paresis chose a day with high therapeutic activity, the use ratio might be slightly overestimated [18].
Use ratio was measured three months after stroke as the majority of recovery occurs within the first weeks and months following the stroke. However, patients may still experience recovery of UL use and prediction of use at an even later point in time would be interesting.

Conclusion

In line with other studies, the present study showed that better function of the paretic UL at baseline predicted increased use of the arm and hand in daily life. This emphasises the value of structured assessment of UL impairment with valid and reliable tests like the FM in clinical settings. The fact that neglect is a major obstacle

for activities of daily life emphasizes the need for better assessment and treatment strategies, particularly in patients with motor potential.

For patients with severe UL impairment, the use of TMS to obtain a MEP status should be considered in clinical practice to gain insight into UL prognosis. If it is not feasible to perform TMS, knowledge of other individual predictors found in the present study may be obtained. In the present study, even patients with only a mild-moderate impairment at baseline might not achieve a normal use ratio. Patients with a potential for improvement should be encouraged to use their affected UL.

Acknowledgements

We thank patients and staff at Hammel Neurorehabilitation Centre for their participation.

Declaration of Interests

The authors report no conflicts of interest.

Funding

The authors received no financial support for the research, authorship or publication of this article.

References

1. Varghese R, Kutch JJ, Schweighofer N, et al. The probability of choosing both hands depends on an interaction between motor capacity and limb-specific control in chronic stroke. Exp Brain Res. 2020 Nov;238(11):2569-2579.
2. Buxbaum LJ, Varghese R, Stoll H, et al. Predictors of Arm Nonuse in Chronic Stroke: A Preliminary Investigation. Neurorehabil Neural Repair. 2020 Jun;34(6):512-522.
3. WHO. Towards A Common Language for Functioning, Disability, and Health: ICF 2002.

4. Sterr A, Freivogel S, Schmalohr D. Neurobehavioral aspects of recovery: assessment of the learned nonuse phenomenon in hemiparetic adolescents. Arch Phys Med Rehabil. 2002 Dec;83(12):1726-31.
5. Buch ER, Rizk S, Nicolo P, et al. Predicting motor improvement after stroke with clinical assessment and diffusion tensor imaging. Neurology. 2016 May 17;86(20):1924-
6. Winters C, van Wegen EE, Daffertshofer A, et al. Generalizability of the Proportional Recovery Model for the Upper Extremity After an Ischemic Stroke. Neurorehabilitation and neural repair. 2015;29(7):614-622.
7. Byblow WD, Stinear CM, Barber PA, et al. Proportional recovery after stroke depends on corticomotor integrity. Annals of Neurology. 2015;78(6):848-859.
8. Gebruers N, Truijen S, Engelborghs S, et al. Prediction of upper limb recovery, general disability, and rehabilitation status by activity measurements assessed by accelerometers or the Fugl-Meyer score in acute stroke. Am J Phys Med Rehabil. 2014 Mar;93(3):245-52.
9. Prabhakaran S, Zarahn E, Riley C, et al. Inter-individual variability in the capacity for motor recovery after ischemic stroke. Neurorehabil Neural Repair. 2008;22(1):64-71.
10. Stinear CM, Byblow WD, Ackerley SJ, et al. PREP2: A biomarker-based algorithm for predicting upper limb function after stroke. Ann Clin Transl Neurol. 2017 Nov;4(11):811-820.
11. Stinear CM, Byblow WD, Ackerley SJ, et al. Proportional Motor Recovery After Stroke: Implications for Trial Design. Stroke. 2017;48(3):795-798.
12. Hayward KS, Schmidt J, Lohse KR, et al. Are we armed with the right data? Pooled individual data review of biomarkers in people with severe upper limb impairment after stroke. NeuroImageClinical. 2016;13:310-319.
13. Kim B, Winstein C. Can Neurological Biomarkers of Brain Impairment Be Used to Predict Poststroke Motor Recovery? A Systematic Review. Neurorehabil Neural Repair. 2017 Jan;31(1):3-24.
14. van Kuijk AA, Pasman JW, Hendricks HT, et al. Predicting hand motor recovery in severe stroke: the role of motor evoked potentials in relation to early clinical assessment. Neurorehabil Neural Repair. 2009 Jan;23(1):45-51.

15. Pizzi A, Carrai R, Falsini C, et al. Prognostic value of motor evoked potentials in motor function recovery of upper limb after stroke. J Rehabil Med. 2009 Jul;41(8):654-60.
16. Boyd LA, Hayward KS, Ward NS, et al. Biomarkers of Stroke Recovery: Consensus-Based Core Recommendations from the Stroke Recovery and Rehabilitation Roundtable. Neurorehabil Neural Repair. 2017 Oct-Nov;31(10-11):864-876.
17. Franck JA, Smeets R, Seelen HAM. Changes in actual arm-hand use in stroke patients during and after clinical rehabilitation involving a well-defined arm-hand rehabilitation program: A prospective cohort study. PLoS One. 2019;14(4):e0214651.
18. Chin LF, Hayward KS, Brauer S. Upper limb use differs among people with varied upper limb impairment levels early post-stroke: a single-site, cross-sectional, observational study. Top Stroke Rehabil. 2020 Apr;27(3):224-235.
19. Chin LF, Hayward K, Brauer SG. Factors influencing paretic upper limb use during first four-weeks post-stroke: a cross-sectional accelerometry study. Am J Phys Med Rehabil. 2020 Jul 22.
20. Barth J, Geed S, Mitchell A, et al. Characterizing upper extremity motor behavior in the first week after stroke. PLoS One. 2020;15(8):e0221668.
21. Rand D, Eng JJ. Predicting daily use of the affected upper extremity 1 year after stroke. J. Stroke Cerebrovasc Dis. 2015;24(2):274-283.
22. Lang CE, Waddell KJ, Klaesner JW, et al. A Method for Quantifying Upper Limb Performance in Daily Life Using Accelerometers. JoVE. 2017;(122). doi(122):10.3791/55673.
23. Uswatte G, Giuliani C, Winstein C, et al. Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial. Arch Phys Med Rehabil. 2006;87(10):1340-1345.
24. Lang CE, Barth J, Holleran CL, et al. Implementation of Wearable Sensing Technology for Movement: Pushing Forward into the Routine Physical Rehabilitation Care Field. Sensors. 2020 Oct 10;20(20).
25. Bailey RR, Klaesner JW, Lang CE. Quantifying Real-World Upper-Limb Activity in Nondisabled Adults and Adults With Chronic Stroke. Neurorehabil Neural Repair. 2015;29(10):969-978.

26. Bailey RR, Lang CE. Upper-limb activity in adults: referent values using accelerometry. J Rehabil Res Dev. 2013;50(9):1213-22.
27. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014 Dec;12(12):1495-9.
28. Kwakkel G, Lannin NA, Borschmann K, et al. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. Int J Stroke. 2017;12(5):451-461.
29. Lundquist CB, Nielsen JF, Arguissain FG, et al. Accuracy of the Upper Limb Prediction Algorithm PREP2 Applied 2 Weeks Poststroke: A Prospective Longitudinal Study. Neurorehabil Neural Repair. 2020 Nov 20:1545968320971763.
30. Platz T, Pinkowski C, van Wijck F, et al. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. Clin Rehabil. 2005;19(4):404-411.
31. Lundquist CB, Maribo T. The Fugl-Meyer assessment of the upper extremity: reliability, responsiveness and validity of the Danish version. Disabil Rehabil. 2017;39(9):934-939.
32. Gladstone DJ, Danells CJ, Black SE. The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair. 2002;16(3):232-240.
33. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int J Rehabil Res. 1981;4(4):483-92.
34. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. Neurorehabil Neural Repair. 2008 Jan-Feb;22(1):78-90.
35. Wu CY, Chuang IC, Ma HI, et al. Validity and Responsiveness of the Revised Nottingham Sensation Assessment for Outcome Evaluation in Stroke Rehabilitation. Am J Occup Ther. 2016;70(2):7002290040p1-8.
36. van Nes SI, Faber CG, Hamers RM, et al. Revising two-point discrimination assessment in normal aging and in patients with polyneuropathies. J Neurol Neurosurg Psychiatry. 2008 Jul;79(7):832-4.

37. Au-Yeung SS, Hui-Chan CW. Predicting recovery of dextrous hand function in acute stroke. Disabil Rehabil. 2009;31(5):394-401.
38. Rorden C, Karnath HO. A simple measure of neglect severity. Neuropsychologia. 2010 Jul;48(9):2758-63.
39. Nijboer TCW, Winters C, Kollen BJ, et al. Impact of clinical severity of stroke on the severity and recovery of visuospatial neglect. PLoS One. 2018;13(7):e0198755.
40. Staples MP, Forbes A, Green S, et al. Shoulder-specific disability measures showed acceptable construct validity and responsiveness. 2010;63(2):163-170.
41. Mehrholz J, Wagner K, Rutte K, et al. Predictive validity and responsiveness of the functional ambulation category in hemiparetic patients after stroke. Arch Phys Med Rehabil. 2007 Oct;88(10):1314-9.
42. de Niet M, Bussmann JB, Ribbers GM, et al. The stroke upper-limb activity monitor: its sensitivity to measure hemiplegic upper-limb activity during daily life. Arch Phys Med Rehabil. 2007 Sep;88(9):1121-6.
43. Tinlin L, Fini N, Bernhardt J, et al. Best practice guidelines for the measurement of physical activity levels in stroke survivors: a secondary analysis of an observational study. Int J Rehabil Res. 2018 Mar;41(1):14-19.
44. Harris JE, Eng JJ. Individuals with the dominant hand affected following stroke demonstrate less impairment than those with the nondominant hand affected. Neurorehabil Neural Repair. 2006 Sep;20(3):380-9.
45. Kutner MH, Neter J, Li W. Applied Linear Statistical Models. Vol. 5th ed. New York: McGraw-Hill; 2005.
46. Kent P, Cancelliere C, Boyle E, et al. A conceptual framework for prognostic research. BMC Med Res Methodol. 2020 Jun 29;20(1):172.
47. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996 Dec;49(12):1373-9.
48. Kirkwood BRS, A.C. Essential Medical Statistics 2ed: Blackwell Science 2003. p. 339-340.
49. Kwah LK, Herbert RD. Prediction of Walking and Arm Recovery after Stroke: A Critical Review. Brain sciences. 2016;6(4):E53.

50. Hoonhorst MHJ, Nijland RHM, van den Berg PJS, et al. Does Transcranial Magnetic Stimulation Have an Added Value to Clinical Assessment in Predicting Upper-Limb Function Very Early After Severe Stroke? Neurorehabil Neural Repair. 2018 Aug;32(8):682-690.
51. Doron N, Rand D. Is Unilateral Spatial Neglect Associated With Motor Recovery of the Affected Upper Extremity Poststroke? A Systematic Review. Neurorehabil Neural Repair. 2019 Mar;33(3):179-187.
52. Esposito E, Shekhtman G, Chen P. Prevalence of spatial neglect post-stroke: a systematic review. Ann Phys Rehabil Med. 2020 Nov 24.
53. Doman CA, Waddell KJ, Bailey RR, et al. Changes in Upper-Extremity Functional Capacity and Daily Performance During Outpatient Occupational Therapy for People With Stroke. Am J Occup Ther. 2016;70(3):7003290040p1-7003290040p11.
54. Essers B, Meyer S, De Bruyn N, et al. Mismatch between observed and perceived upper limb function: an eye-catching phenomenon after stroke. Disabil Rehabil. 2019 Jun;41(13):1545-1551.
55. Schweighofer N, Han CE, Wolf SL, et al. A functional threshold for long-term use of hand and arm function can be determined: predictions from a computational model and supporting data from the Extremity Constraint-Induced Therapy Evaluation (EXCITE) Trial. Phys Ther. 2009 Dec;89(12):1327-36.
56. Wolf SL, Winstein CJ, Miller JP, et al. Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the EXCITE randomized clinical trial. JAMA. 2006 Nov 1;296(17):2095-104.
57. Lum PS, Shu L, Bochniewicz EM, et al. Improving Accelerometry-Based Measurement of Functional Use of the Upper Extremity After Stroke: Machine Learning Versus Counts Threshold Method. Neurorehabil Neural Repair. 2020 Dec;34(12):1078-1087.
58. Koch P, Schulz R, Hummel FC. Structural connectivity analyses in motor recovery research after stroke. Ann Clin Transl Neurol. 2016 Mar;3(3):233-44.

Table 1 Participant Characteristics and Stroke Detail (n=87)

Age, mean (SD)	64.9 (10.5)
Sex, female/ male, n (%)	35 (40) / 52 (60)
Days since stroke, mean (SD)	13.3 (1.6)
Stroke type, ischemic/ haemorrhagic, n (%)	70 (80) / 17 (20)
Side of paresis, left/right, n (%)	47 (54) / 40 (46)
Dominant UL affected, n (%)	43 (49)
Stroke confirmed on imagining, n (%)	86 (99)
Stroke location	
Cortical, n (%)	41 (47)
Subcortical, n (%)	41 (47)
Brainstem, n (%)	5 (6)
Thrombolisis[1], n (%)	29 (41)
Thrombectomy[1], n (%)	17 (24)
Premorbid living in own home, n (%)	87 (100)
First stroke, n (%)	79 (91)
Co-morbidity present, n (%)	61 (70)
Hypertension, n (%)	40 (46)
Coronary artery disease, n (%)	16 (18)
Diabetes, n (%)	7 (8)
Other neurological disease, n (%)	3 (3)
Current smoker (n=74), n (%)	26 (35)
FIM[2] (18-126), median (IQR)	72 (49-85)
FIM motor (13-91), median (IQR)	48 (30-57)
FIM cognitive (5-35), median (IQR)	24 (19-29)
Assessments at baseline	
FM[3] (0-66), median (IQR)	40 (14-53)
ARAT[4] (0-57), median (IQR)	17 (3-39)
SAFE score[5] (0-10), median (IQR)	5 (2-8)
MEP[6] not present[6] (n=81), n (%)	9 (11)
Shoulder subluxation present, n (%)	18 (21)
Light touch affected, n (%)	41 (47)
Proprioception affected (n=86), n (%)	24 (28)
Bilateral stimulation affected (n=85), n(%)	26 (31)
Two-point stimulation affected (n=84), n (%)	43 (51)
UL pain present, n (%)	26 (30)
UL pain intensity (0-10), median (IQR)	0 (0-4)
Neglect present (n=85), n (%)	23 (27)
FAC[7] (0-5), median (IQR)	1 (0-4)

For all variables, the number of participants was (n) = 87 unless otherwise stated.*The included and excluded patients were significantly different. SD: Standard deviation. IQR: Interquartile range. [1]Stroke thrombolysis/ thrombectomy rates were calculated for patients with ischaemic stroke only. [2]FIM: Functional Independence Measure. [3]FM: Fugl-Meyer Motor Assessment Upper Extremity Score.[4]ARAT: Action Research Arm Test. [5]SAFE: Shoulder Abduction Finger Extension. [6]MEP: Motor-evoked potentials. MEP was assessed in patients with a SAFE score below 5 and assumed present in patients with a SAFE score ≥5. [7]Functional Ambulation Classification.

Table 2 Accelerometry Outcomes at Three Months After Stroke for All Patients and in Accordance With FM Score at Baseline

	All patients (n = 87)	FM Severe (score 0-22) (n = 32)	FM Moderate (score 23-50) (n = 28)	FM Mild (score 51-66) (n = 27)
Non-paretic unilat UL activity, hours, median (IQR)	2.1 (1.4- 2.8)	2.8 (2.2- 3.4)	1.7 (1.4- 2.6)	1.6 (1.2- 2.0)
Paretic unilat UL activity, hours, median (IQR)	0.7 (0.4- 1.0)	0.4 (0.2- 0.7)	0.9 (0.5- 1.3)	0.9 (0.7- 1.4)
Bimanual UL activity, hours, median (IQR)	3.0 (1.9- 4.0)	1.7 (0.9- 3.2)	3.3 (2.5- 4.2)	3.3 (2.4- 4.3)
Total UL activity, hours, median (IQR)	5.8 (4.8- 7.2)	5.5 (4.5- 6.0)	6.4 (5.0- 7.6)	6.0 (4.7- 7.4)
Use ratio, median (IQR)	0.7 (0.6- 0.9)	0.5 (0.3- 0.7)	0.8 (0.7 1.0)	0.9 (0.8- 1.0)
Bilateral magnitude, median (IQR)	110.7 (93.5- 127.5)	93.8 (81.8- 112.1)	116.4 (100.7- 140.6)	119.1 (107.5- 133.2)
Magnitude ratio, median (IQR)	-1.9 (-3.2- -0.4)	-3.8 (-4.7- -2.3)	-1.7 (-2.4- -0.1)	-1.0 (-1.6- -0.1)

Use ratio: Rate of duration of paretic/ non-paretic UL use. Bilateral magnitude: Intensity of activity across both ULs for each second of activity. Magnitude ratio: The natural log of the paretic UL vector magnitude divided by the vector magnitude of the non-paretic UL.

Table 3 Regression Models to Examine Prediction of Use Ratio

	Predictors	Con-stant	β-coef ficient	p-value	95% confidence interval	Adjusted model R^2	SD
Model 1- unadjusted association (n=87)						0.376	0.213
	FM score		0.008	0.000*	0.006 to 0.010		
	Constant	0.452		0.000*	0.365 to 0.539		
Model 2- adjusted (n=74)						0.548	0.178
	FM score		0.006	0.000*	0.003 to 0.009		
	MEP +		0.222	0.005*	0.069 to 0.376		
	Neglect present		-0.128	0.025*	-0.240 to 0.016		
	Dominant side affected		0.070	0.108	-0.018 to 0.157		
	Twopd affected		0.024	0.614	-0.071 to 0.120		
	FIM score		0.000	0.736	-0.002 to 0.003		
	Male		-0.046	0.306	-0.136 to 0.043		
	Pain score		-0.005	0.495	-0.020 to 0.010		
	Constant		0.380		0.086 to 0.530		
Model 3 without MEP biomarker (n= 80)						0.458	0.203
	FM score		0.007	0.000*	0.004 to 0.009		
	Neglect present		-0.115	0.059	-0.234 to 0.004		
	Dominant side affected		0.014	0.030*	0.010 to 0.200		
	Twopd affected		-0.042	0.409	-0.144 to 0.059		
	FIM score		0.000	0.635	-0.002 to 0.003		
	Male		-0.062	0.216	-0.161 to 0.037		
	Pain score		-0.000	0.983	-0.016 to 0.016		
	Constant		0.488		0.268 to 0.708		
Univariate regressions of secondary variables							
N=81	MEP +		0.405	0.000*	0.246 to 0.565	0.235	
N=85	Neglect present		-0.222	0.001*	-0.346 to 0.098	0.123	
N=87	Dominant side affected		0.130	0.025*	0.017 to 0.242	0.047	
N=84	Twopd affected		-0.140	0.018*	-0.255 to 0.025	0.055	
N=83	FIM score		0.005	0.000*	0.003 to 0.008	0.180	
N=87	Male		-0.123	0.036*	-0.238 to -0.008	0.039	
N=87	Pain score		-0.007	0.465	-0.027 to 0.013	0.006	

CI: Confidence interval. SD: Standard deviation. *The β- coefficient was statistically significant. In 34 of 40 possible patients with a SAFE < 5, MEP status was established, and in six patients it was not. In 57 patients with a SAFE ≥ 5, MEP+ was assumed. FM: Fugl-Meyer Assessment Upper Extremity. MEP: Motor-evoked potentials. Twopd: Two-point discrimination. FIM: Functional Independence Measure.

Figure 1. Flow Chart of Patients Included

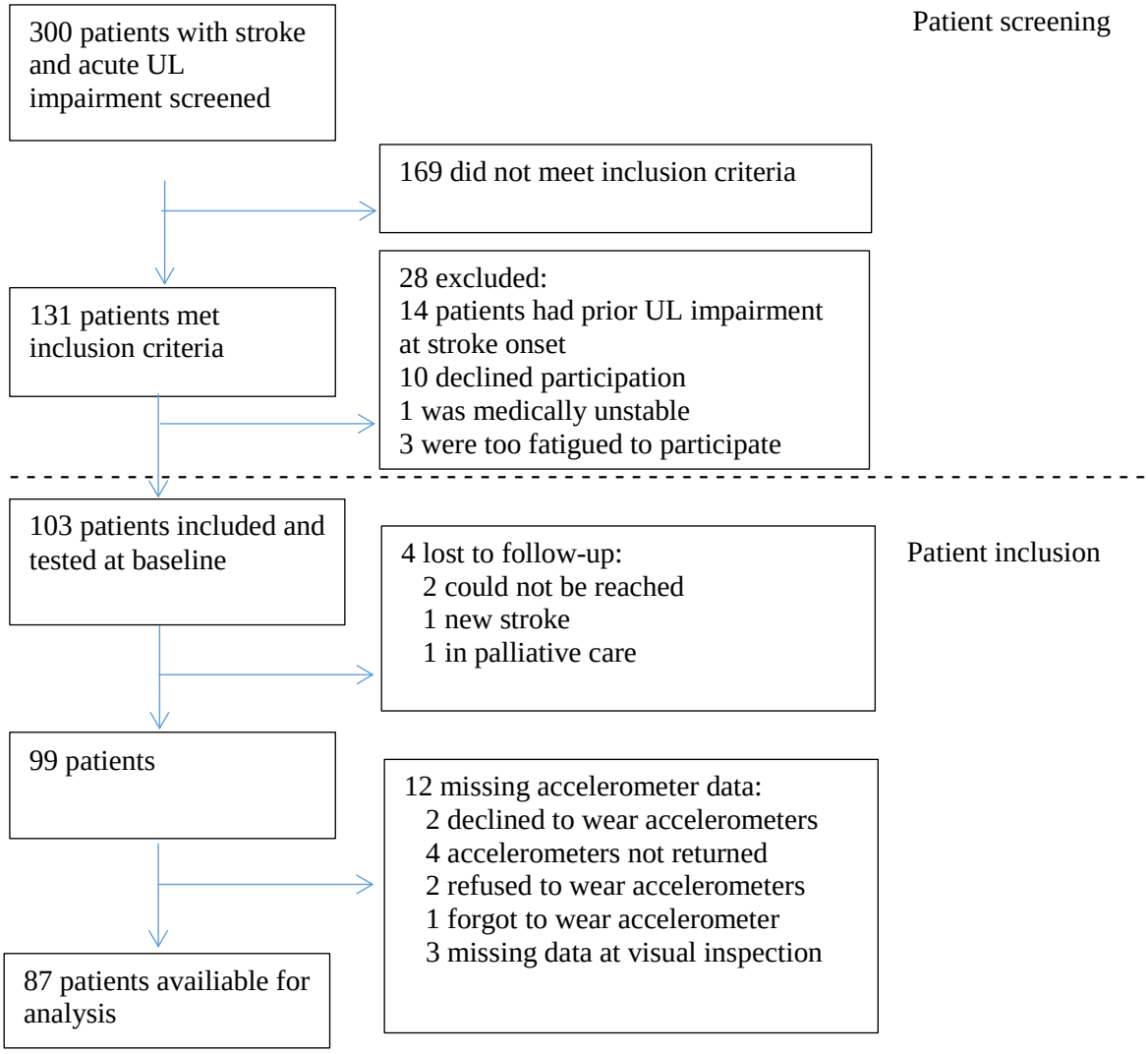


Figure 2. Association between FM at Baseline and Use Ratio at Three Months After Stroke

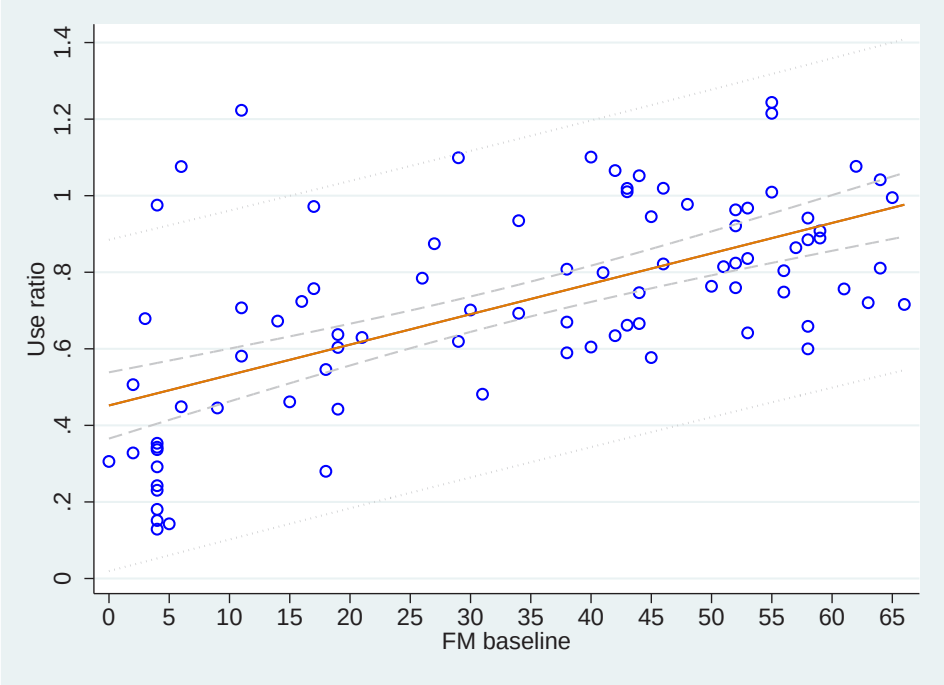


Figure 3a. Association between MEP at Baseline and Use Ratio at Three Months After Stroke

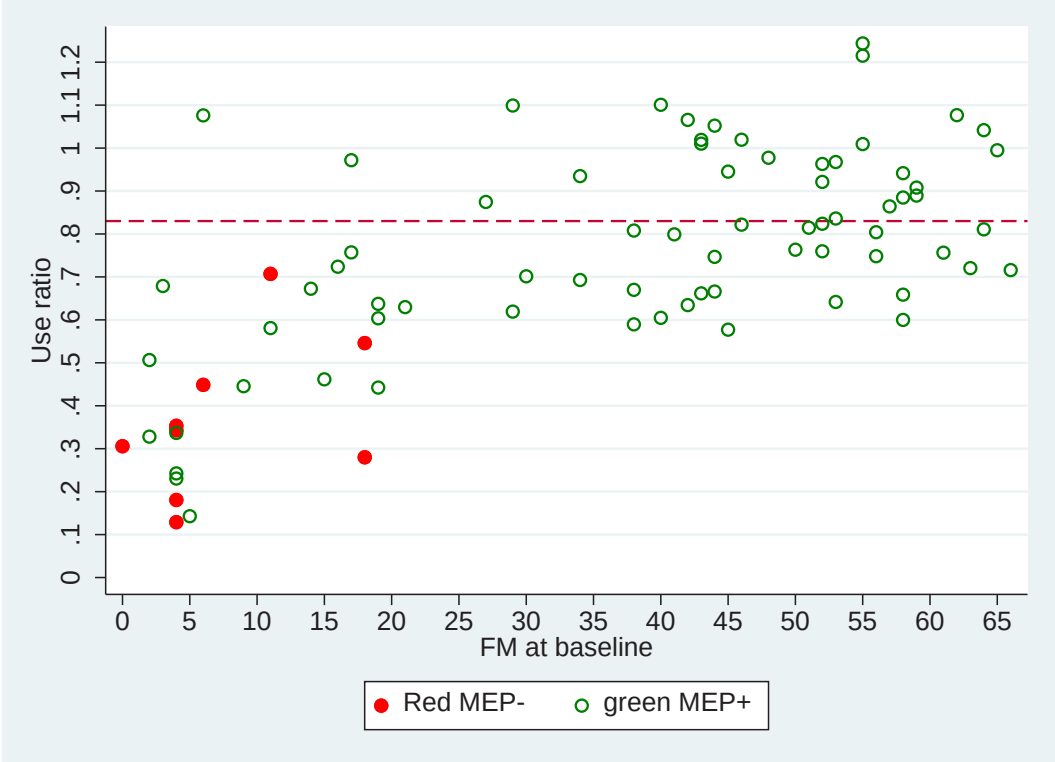


Figure 3b. Association between Neglect at Baseline and Use Ratio at Three Months After Stroke

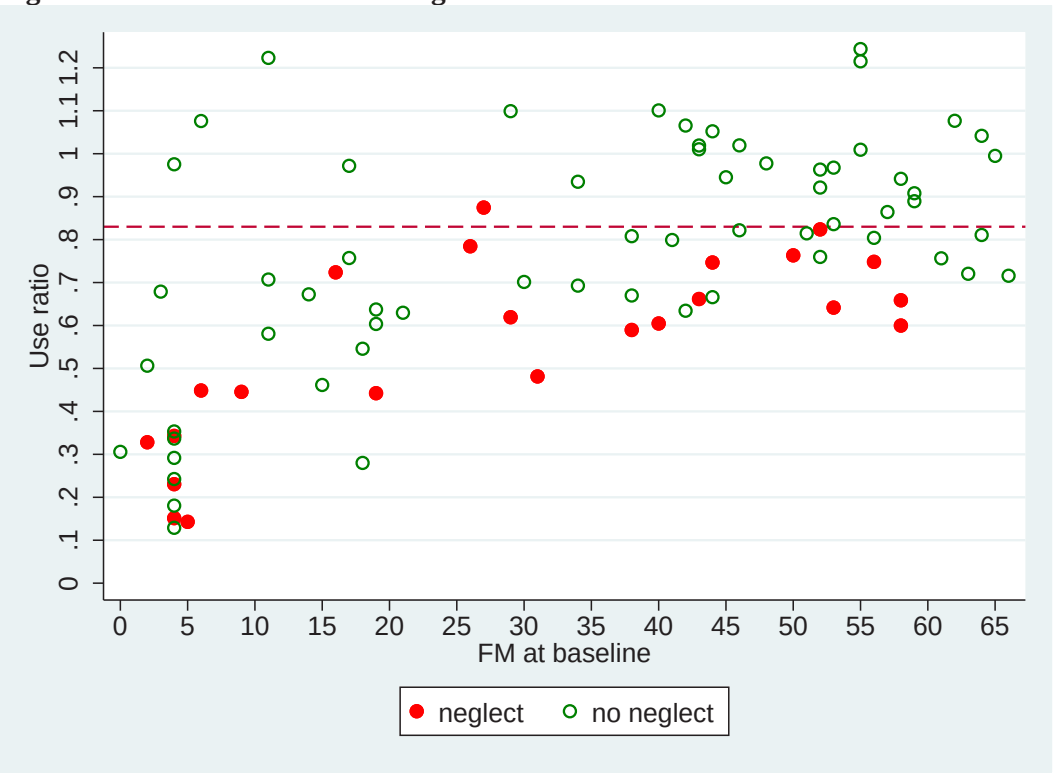
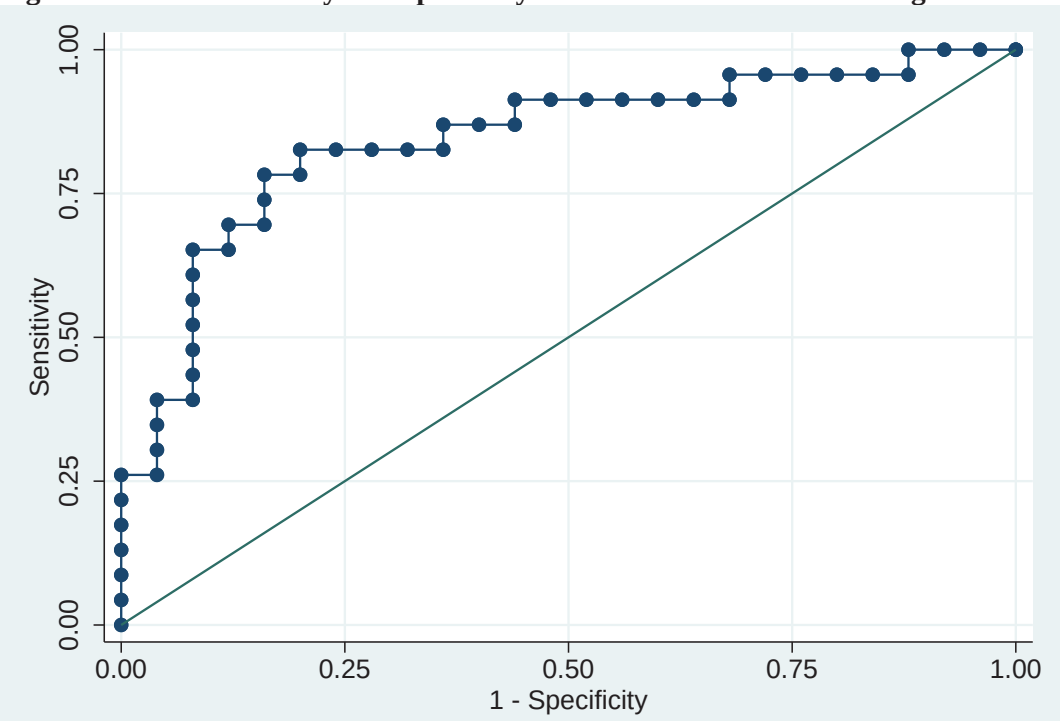


Figure 4. Roc of Sensitivity and Specificity for Prediction of not Achieving a Normal Use Ratio



List of Figure captions:

(Figure 1. no caption)

Figure 2. The solid red line is the best-fitted prediction line of the association between FMA at baseline and use ratio at three months. The 95% confidence interval is displayed with dashed lines and the wider 95% prediction interval is displayed with the dotted lines. With 95% accuracy, the true mean use ratio for a given FMA score will fall within the 95% CI. The PI is an estimate of the interval in which a future observation of UL use ratio for an individual patient will fall, with 95% probability, given what has already been observed.

Figure 3a. Horizontal red line: Threshold for normal use ratio. MEP status for a total of 81 patients. None of the nine patients who were MEP- achieved a normal use ratio. Of the remaining 72 patients, 44 patients did not and 28 patients did achieve a normal use ratio.

Figure 3b. Horizontal red line: Threshold for normal use ratio. Neglect was examined in a total of 85 patients and found present in 23 patients. Almost all, 22 of 23 patients with neglect did not achieve normal use ratio. Among the 62 patients without neglect, 28 did and 34 did not achieve a normal use ratio.

Figure 4. Prediction of achieving non-normal use ratio for patients who had MEP+ and were with¬out neglect. The ROC was based on a multivariate logistic regression with the variables FM, dominant side, twopd and FIM. The AUC was 0.84 (95% CI 0.73-0.96%). If a cut point of 0.55 was chosen, the odds of achieving a non-normal use ratio could be predicted with a sensitivity of 0.80 (95% CI 0.61- 0.91) and a specificity of 0.83 (0.63-0.93).

Paper 3

BMJ Open

Exploring physiotherapists' and occupational therapists' perceptions of the upper limb prediction algorithm PREP2 after stroke in a rehabilitation setting. A qualitative study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-038880.R2
Article Type:	Original research
Date Submitted by the Author:	n/a
Complete List of Authors:	Lundquist , camilla; Regional Hospital Hammel Neurocenter, Research Department Pallesen, Hanne; Regional Hospital Hammel Neurocenter, Research Department ; Aarhus University Tjørnhøj-Thomsen, Tine; University of Southern Denmark Brunner , Iris ; Regional Hospital Hammel Neurocenter, Research Department; Aarhus University, Hammel Neurocenter University Hospital
Primary Subject Heading:	Qualitative research
Secondary Subject Heading:	Rehabilitation medicine, Neurology
Keywords:	Stroke < NEUROLOGY, REHABILITATION MEDICINE, QUALITATIVE RESEARCH

BMJ Open

Title page

Exploring physiotherapists' and occupational therapists' perceptions of the upper limb prediction algorithm PREP2 after stroke in a rehabilitation setting. A qualitative study.

Camilla B. Lundquist[a*]; Hanne Pallesen[a]; Tine Tjørnhøj-Thomsen[b], Iris Brunner[a]

[a]Hammel Neurorehabilitation Centre and University Research Clinic, Hammel, Denmark.

[b]University of Southern Denmark, Copenhagen, Denmark

[*]Address for correspondence: Camilla Biering Lundquist, Research Department, Hammel Neurorehabilitation Centre and University Research Clinic, Voldbyvej 15, 8450 Hammel, Denmark. Mobile phone: +45 30 52 75 16; E-mail: camilla.lundquist@midt.rm.dk

Word count: 7419

Keywords: Stroke, Prognosis, Algorithm, Rehabilitation, Implementation

ABSTRACT

Objective To explore how physiotherapists (PTs) and occupational therapists (OTs) perceive upper limb (UL) prediction algorithms in a stroke rehabilitation setting and identify potential barriers to and facilitators of their implementation.

Methods and analysis This was a qualitative study taking place at a neurorehabilitation center. We conducted four focus group interviews with 3-6 physiotherapists and occupational therapists in order to explore therapists' perceptions of UL prediction algorithms, in particular the Predict Recovery Potential algorithm (PREP2). The Consolidated Framework for advancing Implementation Research (CFIR) was used to develop the interview guide. Data was analyzed using a thematic content analysis. Meaning units were identified and subthemes formed. Information gained from all interviews was synthesized, and four main themes emerged.

Results The four main themes were current practice; perceived benefits; barriers; and preconditions for implementation. The participants knew of UL prediction algorithms. However, only few had a profound knowledge and few were using the Shoulder Abduction Finger Extension (SAFE) test, a core component of the PREP2 algorithm, in their current practice. PREP2 was considered a potentially helpful tool when planning treatment and setting goals. A main barrier was concern about the accuracy of the algorithm. Furthermore, participants dreaded potential dilemmas arising from having to confront the patients with their prognosis. Preconditions for implementation included tailoring the implementation to a specific unit, sufficient time for acquiring new skills, and an organization supporting implementation.

Conclusion In the present study, experienced neurological therapists were skeptical towards prediction algorithms due to the lack of precision of the algorithms and concerns about ethical dilemmas. However, the PREP2 algorithm was regarded as potentially useful.

Strengths and limitations of this study

- To ensure successful implementation health care providers have to regard an intervention useful. A strength of this study is the focus on the therapists' perceptions.
- A strength of this study is the use of focus group interviews as these were an appropriate method to stimulate discussion between participants and illuminate their diverse perceptions.
- Generalizability of results may be compromised, as perceived barriers and facilitators for implementation will differ between sites.

FUNDING STATAMENT
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

BACKGROUND
Stroke is a leading cause of long-term disability in the western world.(1, 2) Upper limb (UL) impairments are common, resulting in functional limitations affecting daily life activities.(3, 4) Accurate prediction of recovery of UL function after stroke is desirable since it can lead to targeted rehabilitation in times of limited resources in health care.(5-7) Some researchers claim that accurate prediction can provide patients and therapists with realistic expectations for UL function and help to set goals for rehabilitation.(5)

From a clinical point of view, a prediction algorithm may be needed most in patients with severe UL impairment. These patients represent a particular challenge for therapists, as it is difficult, based on clinical measures alone, to distinguish patients who regain UL function from those who remain paralyzed.(6, 8, 9) In patients with severe UL impairment, the use of a biomarker may improve prediction accuracy for motor recovery.(7, 10, 11) A biomarker widely used to assess corticospinal excitability is motor-evoked potentials (MEPs), assessed with transcranial magnetic stimulation (TMS).(7, 10, 11)
An UL prediction algorithm that combines clinical assessment with the use of a biomarker is the Predict Recovery Potential algorithm (PREP2), displayed in Figure 1. PREP2 involves several

steps, depending on the severity of paresis. The first step encompasses a clinical assessment of UL function, using the Shoulder Abduction and Finger Extension (SAFE) test. In addition, for patients with low levels of UL function, the motor pathways between the stroke-affected side of the brain and the affected UL are examined using transcranial magnetic stimulation (TMS). Information on age and severity of stroke further contributes to predicting UL recovery.(7) The PREP2 algorithm has an overall accuracy of 75% and its prediction accuracy for patients with severe UL impairment exceeds the accuracy of other prediction algorithms.(5-7) Hence, research indicates that PREP2 is a promising tool for clinical application and in the setting where PREP2 was developed it was found to increase therapist confidence and rehabilitation efficacy.(12) To facilitate implementation of PREP2 in other settings the researchers behind the algorithm are hosting homepages that explains the rationale behind the algorithm and provide relevant instructions to therapists and patients.(13, 14) In addition, a recent paper by Connell et all discuss barriers to implementation and how implementation of PREP2 can be facilitated.(15) However, before we commenced the present study we were not able to identify reports on clinical implementation of PREP2 outside the setting where it was developed.

One of the barriers for implementation described by Connell et al. was the use of TMS, which requires special equipment and trained staff.(15) Another barrier for the use of PREP2 in a rehabilitation setting may be the first step of PREP2, the early administration of the SAFE test. To perform a SAFE test within the first 72 hours may not be possible at a rehabilitation setting where patients are admitted at a later point in time. The present project takes place at Hammel Neurorehabilitation Centre and University Research Clinic (RHN). From June 2018 to October 2019 approximately 2/3 of patients with stroke were admitted to RHN within 14 days of stroke, with a median of 10 days (IQR 6-27). Thus, the majority of patients were admitted too late to obtain a PREP2 prediction while in rehabilitation and the various different acute hospitals do currently not perform UL predictions. However, prediction may be still relevant to rehabilitation focus and goals. Connell et al. suggests that future research may determine if the time windows for obtainment of SAFE and TMS can be expanded.(15) It would ease implementation if the PREP2 could be applied 2 weeks post stroke with satisfactory accuracy. Other factors of importance for implementation may exist. To ensure successful implementation in a clinical setting, a crucial first step is identifying and describing potential barriers and facilitating factors.(16-18) To ensure successful implementation, healthcare providers have to regard an intervention as meaningful and useful for themselves and their patients.(18, 19) Physiotherapists (PTs) and occupational therapists (OTs) responsible for UL treatment are the clinicians most likely to obtain and use the PREP2 predictions.

The aim of this study was, therefore, to explore how therapists perceive UL prediction with the help of the PREP2 algorithm in a stroke rehabilitation setting and to identify potential barriers to and facilitators of implementation.

METHODS

Study Design

This was a qualitative study. We used an implementation framework to develop the interview guide, performed focus group interviews and applied a thematic content analysis. Focus groups are appropriate to illuminate both shared experiences and different perspectives of the group.(20, 21) Group interaction was expected to stimulate discussion of thoughts, beliefs and attitudes towards UL prediction.(20, 21) The interviews were explorative and focused on feasibility, acceptability and perceived usefulness of UL prediction algorithms.(17)

The Consolidated Framework for advancing Implementation Research (CFIR) was applied as a guiding framework to develop a semi-structured interview guide and structure data collection.(15-17) The CFIR is composed of five major domains: intervention characteristics, outer setting, inner setting, characteristics of the individuals involved, and the process by which implementation is accomplished.(16-18) The domains from the CFIR most thorough explored in this study are intervention characteristics, inner setting and characteristics of the individuals involved. The participants' views and attitudes within these three domains are expected to be important to a future implementation. On the contrary, the structure and organization of the fourth domain, outer setting, will not be influenced by the views and attitudes of the participants and the final domain, implementation process, is still in a preliminary phase.

The focus group interviews were centered around UL prediction algorithms, in particular the PREP2 algorithm. PREP2 is a three-step process, see Figure 1. The first step is a calculation of the SAFE score by scoring shoulder abduction and finger extension strength separately between a minimum of 0 and a maximum of 5. SAFE is based on the medical research council grades for limb power and two sub scores are added to form a SAFE score of a maximum of 10. The second step of PREP2 depends on the SAFE score. For patients with an initial high degree of UL function reflected in a SAFE score of 5 or above, information on age is used. For patients with a SAFE score below 5, the function of motor pathways between the stroke-affected side of the brain and the affected arm is examined using TMS to elicit MEPs. For patients in whom MEPs cannot be elicited a measure of stroke severity, the patient's National Institute of Health Stroke Scale (NIHSS) is used. If this scale is not available, the equivalent Scandinavian Stroke Scale (SSS) score may be used.(21) PREP2 predicts UL function at 3 months in one of four categories, from "poor" to "excellent". Patients who are predicted to fall in the category "poor" are unlikely to regain useful movement in their hand and arm within 3 months, while patients in the category "excellent" have the potential to make a complete, or near complete, recovery of hand and arm.

Insert Figure 1 The Predict Recovery Potential (PREP2) algorithm around here

Study Setting

The interviews were performed at RHN, Denmark. The RHN is distributed across three physically distinct rehabilitation units. Patients are admitted to one of the three units. Unit 1 has approximately 70 beds, and units 2 and 3 have 30 and 15 beds, respectively. While adult patients with stroke attend all three units some of the beds at unit 1 are allocated patients with severe (traumatic) acquired brain injury. A research department is placed in connection to Unit 1. Clinical staff at all three locations work in teams, and in total 67 physiotherapists and 67 occupational therapists, involved in treatment of patients, are employed. Some of the therapists have key positions, e.g. specialist physiotherapists or specialist occupational therapists, and are responsible for professional development.

In recent years, therapists at the RHN have developed an interest in using UL prediction algorithms in clinical practice at patient level. The evidence, local relevance and potential implementation have been examined and discussed by a group consisting of 2 OTs and 4 PTs assigned positions within research or professional development. Based on these discussions, the most relevant algorithm for clinical use on an individual level appeared to be the PREP2 algorithm. The main reason was that the predictive value of PREP2 for patients with severe paresis exceeded the accuracy of other prediction algorithms.(5, 6, 12)

Several organizational obstacles prevented an implementation of PREP2 at a local level in the rehabilitation unit. The first part of PREP2, the SAFE test, should be performed within the first 72

hours, while patients are frequently admitted to RHN at a later point in time. Another barrier was the use of TMS, which requires special equipment and specially trained staff. Despite a desire to attain systematic prediction as a clinical routine, implementation of PREP2 in its current form was not possible.

However, two steps to ease a future implementation were taken: First, a prospective longitudinal cohort study was commenced to examine the accuracy of PREP2 when the SAFE score and TMS examination were obtained at a later point in time than originally proposed. Second, the current study was conducted to explore facilitators and barriers for a future implementation.

Participants and procedure

Before the actual data collection, the interview guide was tested for comprehensibility in an interview with an OT and a PT who were both involved in research and implementation. After this, a pilot focus group interview was performed with three PTs invited by the first author, CBL, a physiotherapist and a PhD student. The test interview and pilot focus group interview resulted in minor corrections: the number of questions was reduced, some questions were merged, and information about UL prediction algorithms was simplified. For the complete interview guide see Table 1. Information posters displaying relevant illustrations about the topic, e.g. the PREP2 algorithm, were produced in order to explain and facilitate discussion in the subsequent interviews.

Table 1. Interview Guide

Main categories	**Questions**
General questions	In patients with paresis of arm and hand: Which factors do you consider relevant for future arm and hand function? (important elements)
	What is relevant for your own approach to treatment of the arm and hand? (write down three-four issues/ things)
Thoughts on prediction	What are your thoughts about prediction of arm and hand function at an early point in time? What are the likely consequences?
	Which patients/ groups of patients would benefit from knowledge of prognosis (e.g. paralytic UL)?
	UL prediction models: to whom will it not make sense?
	Does age matter for prognosis (in general and for UL in particular)?
	Severity of stroke from onset is relevant for UL prognosis. Where do you seek this information (e.g. ward round, patient record, looking for particular scores as NIHSS or SSS)?
	Do your expectations of future UL function influence your approach and choice of UL treatment?
SAFE score	Before participation, you were asked to perform a SAFE score on at least three patients. How was it?
	What are your thoughts on using specific UL tests for (all) patients with reduced strength in arm and hand (e.g. SAFE score, Fugl-Meyer score)
	Are you aware of other hospitals focusing on UL prediction? E.g. if they use SAFE?
Knowledge of evidence	How do you get knowledge updates on UL treatment?
	Do you have the time and opportunity to get updated on new knowledge?
	Exercise: I explain the PREP2 model and show pictures of the elements:
	What are the pros and cons of a UL prediction model similar to the PREP2?
	What should it take for you to use a UL prediction model?
	Do you see patients for whom a prediction model would make no sense?
	Would use of a UL prediction model change your approach to a patient?
	PREP2 can predict future UL function with approximately 75% accuracy. What is your opinion on that?
	Transcranial magnetic stimulation (TMS) - can it be use in this setting?
Summarising	What we have talked about.
	Do you have anything you would like to add?

The ward managers at each of the three hospital units were asked to invite participants based on the following criteria: a mix of PTs and OTs, involved in the treatment of patients, at least one year of experience in neurorehabilitation, and from different wards. Experience with UL prediction algorithms was not a requirement. The intention was to achieve maximal variation in profession, clinical experience, and degree of specialization.**(22)**
After being appointed for the interview an information letter was sent to the participants, in which the purpose of the interviews and the background for UL prediction algorithms for patients with stroke were presented. The participants were specifically informed about the PREP2 algorithm and were instructed to perform step 1 of the algorithm, the SAFE test, on at least three patients before participation. For this purpose, they were given a written scoring instruction. Performance of the SAFE test should ensure practical experience with the test and qualify the discussions during the interviews.

The interviews took place at the participants' work site. The interviews started with a few broad questions about what the participants considered important factors for UL prognosis. The purpose of these broad questions was to make the participants relax and feel comfortable and get their spontaneous opinions. Afterwards, the questions were more specifically about prediction algorithms, the use of tests, and attitudes towards evidence-based practice. Finally, the PREP2 algorithm was introduced and discussed.

The focus group interview was moderated by CBL, who was aware of ensuring a confident atmosphere that welcomed a diversity of opinions. It was emphasized that there were no right or wrong answers. A senior researcher, HP, functioned as an observer, providing feedback to the moderator and observing interactions in the focus group. HP asked clarifying and supplementary questions during the interviews. Directly after the interviews, the overall impression of the interview and any spontaneous reflections and considerations were noted. The interviews were audio-recorded and transcribed verbatim by CBL.

Analysis
The interview transcripts were imported to the qualitative research software program NVivo12 to facilitate coding and make data analysis more manageable. The pilot focus group interview was considered to add interesting dimensions to the topic and data from this interview was included and analyzed along with data from the succeeding three focus group interviews. A thematic content analysis of the interviews was performed by CBL.(22, 23) The qualitative data analysis was both a deductive and an inductive process.(22, 23) Deductive as we used the CFIR framework and sought to answer the specified research question regarding barriers and facilitators for implementation (theory-based coding). Inductive as we let the material talk (data-based coding) because attitudes towards UL prediction algorithms have not previously been explored, and knowledge of how to implement algorithms into the clinic setting is scarce. First, the four interviews were individually open-coded in NVivo and meaning units were identified. Second, the interviews were compared for similarities and differences and based on the meaning units, subthemes were formed. Finally, information gained from all four interviews was synthesized, and four main themes, considered of great importance to the participants and relevant for implementing prediction algorithms emerged (See Figure 2).

Insert Figure 2. Diagram showing example of theme formation around here

The coding and interpretation of results were continuously discussed with co-authors. According to Malterud this triangulation between authors with different positions and perspectives will increase

the understanding of complex phenomena.(24) The four interviews revealed a broad array of relevant considerations, many of them appearing repeatedly, indicating data saturation.

In the results section below quotations are used to display from what kind of original data categories are formulated, thereby documenting and substantiating findings and increasing the trustworthiness of the study.(23-25) Where cited, the context is quoted in parentheses with anonymized participant initials and focus group origin, in accordance with Table 2 e.g. participant E from focus group 2 would be quoted as *(participant E, F2).* To ensure credibility, a participant from each focus group interview has reviewed the interview transcripts and the interpretation of the findings.(22) The participants agreed to the transcripts, recognized themselves in the descriptions and provided further nuance to the findings.

Ethical Considerations

The study was conducted in accordance with the Helsinki Declaration 2008. Participation was voluntary. All participants signed an informed consent form and were informed that they could withdraw at any time. Anonymity was preserved by changing names and identifiable places or situations. In accordance with Danish legislation on research ethics, and due to the nature of the study, approval by the Research Ethics Committee was not required.

Patients and Public Involvement Statement

Patients or the public were not involved in the design or conduct of this study.

RESULTS

A pilot focus group interview followed by three ordinary focus group interviews were performed from January to April 2019 and had a duration of 68 to 90 minutes. In the pilot focus group interview, three PTs participated. All had clinical experience in either neurorehabilitation or acute neurology and were engaged in either a Master's degree or a PhD. In the succeeding three interviews, all participants were employed at neurorehabilitation wards. The number of participants in the focus group interviews corresponded to the size of the rehabilitation unit: six participants from unit 1, four from unit 2 and three from unit 3. Both PTs and OTs participated; three were specialists in the field, one was a student advisor, all were female, had graduated 5-23 years previously, and had 3-20 years of neurological experience (see Table 2 for characteristic of participants).

Table 2. Characteristics of focus group participants

Group	Pilot focus group (F1)	Focus group 1(F2)	Focus group 2 (F3)	Focus group 3 (F4)
Number of participants	3	6	4	3
Profession	3 PT	3 PT; 3 OT	2 PT; 2 OT	1 PT; 2 OT
Assigned position	1 specialist	2 specialists, 1 student advisor		
Educational level	2 Master; 1PhD	5 Bachelor; 1 Master	4 Bachelor	3 Bachelor
Gender	2 F; 1 M	6 F	4 F	3 F
Average years since graduation (range)	15 (12-18)	12 (5-17)	20 (13-23)	17 (9-23)
Average years of experience in neurorehabilitation (range)	11 (10-18)	10 (3-17)	17 (13-20)	12 (2-18)

Current unit of employment	Unit 1 and Acute Neurology	Unit 1	Unit 2	Unit 3
Anonymized initial of participant when quoted	A; B; C	D; E; F; G; H; I	J; K; L; M	N; O; P

PT= Physiotherapist, OT= Occupational Therapist; F= Female, M= male

Findings and quotations in relation to the four main themes: current practice, perceived benefits, barriers, and preconditions for successful implementation are presented below and in Figure 3.

Insert Figure 3 The four main themes and their subthemes around here

1. Current practice

Knowledge of current practice is a precondition for understanding the participants' perceptions of barriers and perceived benefits. This first main theme concerned the participants' considerations on current practice and encompassed three subthemes*: limited use of UL assessments, considerations on UL prognosis and treatment*, and *professional identity*.

1.1 Limited use of UL assessments

As prediction algorithms comprise the performance of standardized assessments, information about the use of UL assessments was relevant. Participants in all four interviews agreed that UL tests were used, but on a limited scale. There was consensus that the UL test had to be clinically relevant for the specific patient and not a routine test for everyone. According to several participants UL tests were primarily meaningful and used for patients with moderate to good UL function:

> *Yes.. MAS I believe I use a lot. With arms that can…do a bit more (participant F, F2)*

In addition, the test had to be quick to perform and easy to administer:

> *One has to prioritize the time to do it. So it has to make sense to do it. (participant B, F1)*

1.2 UL prognosis and treatment

Many factors were considered important for UL recovery. Some but not all aligned with factors highlighted in the literature. Pain was highlighted in all four interviews and had to be prevented and treated for UL function to occur:

> *Well, I believe pain has a big say. Because…if they have pain, they don't move their arm. They just try to protect it… (participant F, F2)*

Initial UL function and time since stroke were also mentioned in all interviews, but not stressed by the participants as important predictors:

> *I think that having some function is important. We have a lot…I believe where the SAFE score is zero…because they are paralyzed… you cannot palpate any muscle activity. That has a huge importance for… whether they regain any function at all… (participant G, F2)*

Other factors mentioned in the interviews as important for recovery were sensory motor deficits, time since stroke, location of stroke, type of stroke, and initial medical treatment. According to all of the participants, cognition was vital, especially neglect and awareness of own disabilities:

> *Yes. And cognition, all things considered…yes that matters. The ability to understand instructions. And maybe even to be able to perform self-training…that they understand the importance of focusing on arm and hand training. (participant K, F3)*

Many participants mentioned the importance of past experiences, self-efficacy, motivation, and inner drive:

> *And the patients that have a good inner drive…they have a good prognosis [the group agrees]. (participant F3)*

The PREP2 algorithm includes information on age and initial score on stroke severity. However, age was not considered particularly important for UL prognosis, and only a few participants were aware of initial scores performed in the acute units.

When planning UL treatment and choosing interventions, the participants took many of the same elements into account as when considering UL prognosis. Importantly, they found that the patients' individual goal should guide whether or not UL treatment was a main priority:

> *The patient's priority counts. If the most important thing is to get that arm and hand going. Right now, I have a patient where eating was the most important issue and what I prioritized. (participant G, F2)*

1.3 Professional identity

Professional identity concerns how the participants perceive themselves in relation to their profession and membership of their profession.*)*

In all of the interviews, the participants agreed that use of UL assessment and algorithms such as the PREP2 aligned more with the PT profession than with the OT profession. Even though both professions use similar UL interventions and approaches, the PTs traditionally treat patients on an impairment level, while the OTs focus on activities and activity limitations:

> *I believe our examinations differ. PTs have this…what can I say…very body-level examination, while we involve them during activities and in the bathroom or kitchen [the group agrees]. (participant K, F2)*

> *Well, if I have a patient I look for … because I am an OT… for activity limitations in relation to the use of arms and hands…because I am an OT. (participant H, F2)*

Most of the participants considered themselves experienced neuro-therapists; and according to many participants, prediction algorithms may make most sense for recently qualified therapists. Recently qualified therapists will need a simple tool, while the more experienced can draw on years of experience:

> *I believe this PREP2 is for more recently qualified therapists…a lot easier to*

access…because then you can draw on the cold facts: this is what we have to guide us. And they are more schooled in that that the rest of us. [the group agrees] (participant M, F3)

2. Perceived benefits

This theme centers around how the participants thought an algorithm could aid and ensure UL treatment and rehabilitation. Subthemes were *the SAFE score is easy; a helpful tool; a positive algorithm can motivate;* and *positive towards new technology.*

2.1 The SAFE test is easy

In the pilot focus group, participants had a general knowledge of prediction algorithms, but across the other interviews, knowledge of algorithms was less profound. All participants had heard of UL prediction algorithms and in one unit, some of the participants used the SAFE test. SAFE is step one in PREP2. All participants had been asked to perform a SAFE test on at least three patients before the interviews. Especially the physiotherapists found the SAFE test easy to administer:

The SAFE test is easy and quick and you can allow yourself to do it no matter what. (participant B, F1)

Some participants found the SAFE score insensitive, as the difference between score 2 (=limited range of motion without gravity) and score 3 (=full range of motion against gravity, but not resistance) was rather large. Despite this, the same participants considered the SAFE score to be appealing, because it was quick and could be performed everywhere and without equipment.

But that big gap…we actually discussed it…. Actually, for some patients we would like to score 2½ [the group agrees]. (participant M, F3)

But apart from that, it is an easy score as…it doesn't need you to bring anything with you. And you can do it everywhere. (participant M, F3)

2.2 A helpful tool

PREP2 was considered a potentially helpful tool considering the prognostic potential of UL and for planning treatment. The algorithm might not be able stand alone, but in combination with information from other sources, it could be used as a tool or an indicator to decide what way to go, e.g. whether to intensify UL training or instead start the use of compensatory strategies.

I believe an indicator is a good word. An indicator. Because it is not an answer to functions they will not achieve…or that it will be amazingly good. But it gives an indication. For this reason, we choose to go this way. But it does not mean that when the patient is discharged from RHN, we will write: The patient will never achieve any function. It is just a good tool. (participant F, F2)

Yes as in a toolbox. Just like many other things. (participant H, F2)

It is always nice to know more about prognosis. (participant K, F3)

Across interviews, there were different views on whether UL prediction algorithms would be a prognostic aid for all patients, mainly those with no or little function or those with moderate function. The predominant opinion was that it would be particularly relevant for patients with little or no UL function, reflected in a SAFE score below 5, as it was difficult for the therapists to predict UL function for these patients.

> *The paralyzed patients. Or those nearly paralyzed. I believe those patients would benefit (participant C, F1)*
>
> *Well… if so… it is only those with a SAFE score below 5. (participant N, F4)*

2.3 A positive algorithm can motivate

All participants envisaged that a prediction algorithm could be used to motivate patients and therapists, given that the prediction was optimistic.

> *Some indication… would be nice. It could be used to motivate when progression is slow and you think nothing is happening in an arm. If I could say: I KNOW if we do this exercise for the next four weeks every day, then it will come; that would motivate the patient. And me as a therapist. (participant P, F4)*

Even though the participants preferred prediction algorithms to be as accurate as possible, many believed that an algorithm could be an aid without being 100% accurate. Several participants said that an accuracy of 75% would give an indication of whether your treatment plan was on the right track and what you could expect. It could still be used by the team or individual therapist along with other indications and tools of prognosis.

> *For me, it will be a tool to use in a team. I often believe that…with FIM and other functional measures…it is so interesting when…it does not fit. Then we get some beneficial discussions. (participant E, F2)*

2.4 Positive towards new technology

In all four interviews, the attitude towards TMS and MEP was positive. The participants found it appealing that use of TMS and obtainment of a MEP could add information to UL prediction that could not be obtained by a clinical test. They imagined this information would motivate both patient and therapist:

> *But what I find really interesting is that you can have this.. MEP…? If there is a connection in the corticospinal tract. So you can have a SAFE below five and still expect a good function. (participant L, F3)*
>
> *There might be some people where you think they should have got some more…because if we had that examination, TMS… (participant G, F2)*

3. Barriers

This theme encompassed the participants' perceptions of the limitations of prediction algorithms and potential barriers to their implementation in clinical practice. Three subthemes emerged within this main theme: *an algorithm must be accurate, ethical dilemmas*, and *fear of consequences*.

3.1 An algorithm must be accurate

All participants agreed that an algorithm should be as accurate as possible:

> *Definitely, definitely [the group agrees]. (participant L, F3)*
>
> *It must, of course, be very precise for us to use it. (participant O, F4)*

However, disagreement existed on whether the 75% accuracy of the PREP2 algorithm was precise enough. For some, a precision of 75% would be a barrier, and one participant stated that even if the algorithm was 100% accurate, she still might not follow it.

3.2 Ethical dilemmas

Whether or not to present and discuss the UL prediction with patients emerged as a dilemma for many participants. If a patient was predicted to have little or no function, this might depress the patient and would conflict with the participants' desire to motivate the patient:

> *Yes. And what day do we tell the patient? Is it when they arrive and have been here in…? Well. I really don't know. On top of everything else? (Participant N, F4)*

Even if some participants were skeptical, they were still open for discussion and dialogue when other participants responded that informing the patient could make it easier to focus on other aspects of the rehabilitation where improvement seemed more realistic:

> *I find it difficult to shatter someone's dream. You need to dream and believe this one will gain function. For some time. Of course, not for several years. (participant G, F2)*
>
> *Well…Well it is a balance isn´t it. We have patients who come and tell us they are sorry that they weren't told…so the most important thing is to dare tell them, to be honest…well why should we treat an arm that we are nearly 100% will never function again? (participant H, F2)*
>
> *No…No.. but… (participant G, F2)*
>
> *I don't believe we necessarily shatter someone's dream…necessarily… by letting the patient know how much this arm can improve. Instead, we consolidate and focus rehabilitation. (participant H, F2)*

3.3 Fear of consequences

The participants believed that UL treatment affected future UL function and should be offered regardless of initial function. The general view across interviews was that all patients deserved that therapists did their best to restore UL function.

In focus groups 3 and 4, concern was expressed that use of a prediction algorithm would dictate which patients should receive treatment and which should not. If so, patients with a negative prediction would receive little or no UL treatment, and the algorithm would serve as a self-fulfilling prophecy. As such, the participants feared that introduction of a prediction algorithm would alter their approach to the patients:

> *And then I might prioritize other issues instead. I am afraid so. And I hope I wouldn't. Because I believe that they need all the treatment they can get.…. Because, truly, there is a chance in reality. (participant O, F4)*

> *If I had a diagram that could tell…your arm will never be good.…. then I believe the patient should get the opportunity to prove this wrong. (participant J, F3)*

In three of the interviews, it was mentioned that an algorithm could be used to stratify and prioritize which patients should receive treatment. In one interview, the participants regarded this a positive consequence because it could be used to optimize treatment in times of limited resources:

> *Because it is such a difficult matter already - and we do not have that many rehabilitation beds. So that would be an enormous help, I believe. (participant A, F1)*

On the other hand, participants from unit 3 and especially unit 2 looked at algorithms in light of pressure from budget cutbacks; they feared that an algorithm would be used to accelerate and shorten rehabilitation periods, regardless of the patients' rehabilitation potential. In addition, some participants feared that an algorithm would introduce a too simplified view on humans.

> *It depends how – if you can say so - our managers wish to use this tool…because we are under pressure. And will this be a tool to evaluate…which patients should be here? (participant K, F3)*

> *And I think it is like a tendency in society. That we need something that can be measured and recorded, and hard facts [the group agrees] (participant M, F3)*

4. Preconditions for implementation

Preconditions for future successful implementation were grouped in two subthemes: *tailored implementation* and *organizational structure and resources*.

4.1 Tailored implementation

The focus group interviews were performed at three different units. The overall impression was that despite being part of the same rehabilitation hospital (RHN), different cultures existed at the three units. Especially at unit 1, the participants (see Table 2) were open to new ideas and implementing new knowledge seemed an integrated part of their culture. The participants at the other two rehabilitation units seemed open to new ideas, too, but were at the same time more skeptical. In all interviews, the participants discussed the importance of tailoring implementation to the specific unit, ward, and patient. If something new had to be implemented, a persistent focus on the topic was needed:

I believe that you must realize that implementation is just a lot more time-consuming and difficult than you imagine. A single day - when you present, discuss and maybe do something practical - is just not enough. (participant D, F2)

I can say…in my ward…if something must be implemented, you have to take the specific patient and the patient's team to make it work. We cannot say something general about you having to…..in all upper limbs...to do so and so. It has to be specific so they can relate to that. (participant I, F2)

We aren't different from the patients. We, too, need a lot of repetition to implement something new and learn it. [the group agrees] (participant M, F3)

4.2 Organizational structure and resources

Time to get acquainted with new evidence and practice new skills was considered insufficient. All interviews showed that there was a sense among participants of being well-informed, while time to incorporate and practice new skills and routines was lacking:

No. we don´t even have the time to plan our daily treatments. So no, not at all. That is a real challenge [the group agrees] (participant E, F2)

The level of information is actually okay…it is more the time afterwards…to incorporate it. (participant J, F3)

Yes exactly. (participant K, F3)

True …to make it a routine. (participant J, F3)

Several participants' mentioned that prioritization and support from the ward manager were important for success. In all interviews, participants mentioned that weekly or monthly meetings were planned ahead and could be dedicated to specific issues. These meetings were valued and considered important by the participants when new knowledge had to be practiced, modified, and implemented. Flexibility from colleagues was acknowledged, and several participants considered this a prerequisite for attending a course. Generally, the participants felt that there was a culture of sharing the acquired knowledge with colleagues.

It depends on how your ward works, I believe. How generous your colleague or ward manager is in relation to….well there is some economy in it too…but how much energy will we put into this? And are the rest prepared to run faster while someone is attending a course….? That is the culture and what you want. (participant D, F2)

Members of staff were assigned specific positions as a specialist OT or specialist PT. Specialists were considered a resource, capable of, and responsible for presenting and implementing evidence.

And when some of the specialist therapists have been out in the wide world and return home and tell us about it, or some colleagues have been at a course… (participant M, F3)

For me, it is about responsibility. Someone has to take responsibility. Because if all are responsible, nothing happens. I, as a specialist, can be the one responsible and say: Your patient, has he got an UL problem? (participant F, F2)

DISCUSSION

Summary of main findings

In the current qualitative study four main themes were identified when exploring therapists' perceptions of prediction algorithms: current practice, perceived benefits, barriers, and preconditions for implementation. Most participants knew of UL prediction algorithms. However, in practice, only some elements were applied and by a few therapists. Most participants considered themselves experienced neurotherapists and regarded UL prediction algorithms as particularly useful for more recently qualified therapists. The PREP2 algorithm was considered a potentially helpful tool when planning treatment and setting goals. The perceived benefits centered on the SAFE test, a core component of the PREP2 algorithm. In addition, participants appreciated the use of TMS if it could add information to UL prediction. The main barriers were concern about accuracy of the algorithm and dilemmas arising from having to confront patients with their prognosis. Preconditions for implementation encompassed tailoring the implementation to a specific unit, having sufficient time, and being part of an organization supporting implementation.

Comparison with previous findings

Current practice was characterized by limited knowledge and use of UL measurements and UL prediction algorithms. This result is corroborated by a recent Danish survey study by Kier et al. that revealed, that prediction models for UL function after stroke are not yet a part of daily practice in Danish stroke rehabilitation.(26) This is not surprising as the Danish clinical stroke guidelines do not recommend the use of any particular UL measurement or UL algorithm.(27, 28) Moreover, international recommendations of standard use of UL measurements and have not yet been implemented in clinical practice.(29, 30)

According to previous research, the initial UL function is the main predictor for UL recovery.(6-9, 31-33) The participants in our study acknowledged the predictive value of initial motor function. At the same time, they considered several other aspects important, such as the patients' goals, their motivation, and their self-efficacy. Studies have shown that other factors as individual goals, motivation, self-efficacy, aphasia and depression influence rehabilitation outcomes and should be considered.(34-37) The experienced therapists in our study drew on their clinical knowledge and expertise,. This is in line with their views on the PREP2 algorithm as a useful but supplementary tool that cannot stand alone. Therefore, they suggested that a UL algorithm would be particularly helpful for recently trained therapists. However, previous research indicates that prognoses based on clinical expertise are not superior to those of algorithms, even among experienced therapists.(38)

In the present study, participants were positive towards the SAFE test and found it easy to use. The PREP2 algorithm has three steps and requires information from few sources.(7) For approximately 2/3 of the patients only the SAFE test is needed and according to Connell et al. simple prediction algorithms are more likely to be implemented.(15)

Prediction of UL function in severely impaired patients was considered a particular challenge. The participants therefore welcomed new technologies such as TMS, which could help to distinguish patients who regained function from those who remained paralyzed. However, their views on

prediction algorithms differed depending on the prediction outcome. A favorable prediction was considered motivating for both therapists and patients. By contrast, most participants found a negative prediction demotivating. This is in line with the findings of Connell et al., who points to that individualized prediction is a new field for therapists, and that negative predictions may be particularly challenging.(15) According to Connell et al. therapists may need assistance in delivering negative predictions (15) and the PRESTO homepage as well as the PREP2 training homepage by Stinear et al. provides suggestions on how to phrase and deliver negative predictions.(13, 14)

Some participants considered the 75% accuracy of the PREP2 algorithm to be insufficient to improve clinical decision-making. To enable more precise predictions, participants in the present study proposed combining PREP2 with other sources of prognostic information. However, many factors the participants claimed to be relevant for UL prognosis have, in fact, already been examined.(6, 7). During the development of the PREP2 algorithm Stinear et al. showed, that the most important predictors to incorporate in PREP2 were MEP status, SAFE score and NIHSS.(7) The predictors irrelevant to include were sex, hemisphere affected, hand affected, stroke classification, thrombolysis, and previous stroke. UL outcome was not predicted by these factors; nor was it modified by UL therapy dose.(7) In the present study, participants also mentioned the importance of personality traits that are not easily quantifiable, e.g. inner drive and the approach to life. Such personality traits have so far received little attention in research and may need further investigation.

Although the three units were part of the same rehabilitation hospital, different cultures existed, e.g. expressed in participants from unit 2 and 3 being more sceptic to implementation of an UL algorithm. A reason for this might be greater focus on UL prediction algorithms prior to the interviews at unit 1 resulting in more knowledge and experience with UL prediction algorithms than in the other two units. Furthermore, some of the observed differences in culture may be attributed to the characteristics of the participants and the site. Participants who had graduated more recently, were employed at the largest site and in proximity to the research unit were more prone to a positive attitude towards prediction algorithms. Differences in culture stress the importance of tailoring a future implementation to the particular setting in which it is intended to be used. This is in accordance with similar studies that focus on the importance of translating and implementing research into practice.(16-18)

The present study was the first step in an implementation process. By identifying potential barriers and facilitators, these can be addressed. CFIR was used as an overarching framework, ensuring that aspects relevant for future implementation were systematically captured. Emphasis was on the participants' perspectives and the CFIR constructs to which they could easily relate: intervention characteristics, inner setting, and characteristics of the individuals involved. This approach was chosen as the beliefs of healthcare staff about an intervention are often more influential for implementation than other factors such as the strength of evidence supporting the intervention.(16, 17, 19)

Limitation and strengths

The scientific trustworthiness of the present study was evaluated using the concepts credibility, confirmability, and transferability.(22-24) Also, a checklist for focus groups interviews was used to assure that important aspects considering the research team, study methods, context of the study, analysis and interpretations were addressed.(25)

Credibility was ensured by involving several researchers with different positions and perspectives who could supplement and challenge each other in the analysis. According to Malterud, multiple researchers can in this way strengthen a study.(22, 24) To further ensure credibility, a participant from each focus group interview reviewed transcripts and findings. To assure confirmability, we aimed not to let our pre-understandings influence the interpretation of the findings.(22) Being aware of our own preconceptions is essential as this enabled us to analyze the interview transcripts with open minds.(23, 24)

Transferability or generalizability concerns the application of the study findings beyond the context in which the study was undertaken.(24) perceived barriers and facilitators for implementation will differ between sites, depending on the characteristics of the clinical setting and the people involved.(15, 17) As a consequence, findings from the current study will not necessarily be generalizable to other settings. Nevertheless, the systematic use of CFIR as framework even before the start of implementation can be transferred to other contexts.

A potential bias of the present study is that the ward managers either appointed participants or asked participants to volunteer for the present study. Therapists with an interest in UL algorithms or implementation may be more eager to participate. If so, the expressed perceptions towards UL prediction algorithms may well be more positive than what is the case amongst therapists in general.

Another limitation of the present study is that the participants did not try to perform the complete PREP2 before attending the interviews. PREP2 is comprised of the SAFE test, information on age, NIHSS score and MEP status. While the NIHSS score or the equivalent SSS score are always performed at the acute units and can be found in the medical records, knowledge of MEP status are on the contrary not easily obtainable. Performance of TMS to obtain MEP status requires longer period of training, which was not feasible for the present study. For this reason, the participants' thoughts on TMS are merely theoretical. However, TMS is expensive to purchase and requires ongoing training of staff to operate and therefore constitutes an obvious barrier to a future implementation of PREP2. Still, the participants practiced the most essential part of the PREP2, the SAFE test, before attending the interviews.

Future directions

The present study reveals that the perceptions of the participants only partly align with current scientific evidence, reflecting a lack of translation from evidence to applied knowledge. Connell et al. states, that the beliefs of health care staff about interventions are often more influential than other factors such as the strength of evidence for the intervention.(19) As a consequence, the evidence behind UL prediction algorithms should be presented and discussed in more details with the therapists prior to implementation. In this context it is important to consider, that prediction algorithms should not be implemented in clinical practice until both development and validation studies have been conducted.(6) Before implementation in our local setting we need to establish if the high accuracy of PREP2 can be preserved if the time windows for obtainment of the SAFE score and the MEP status are expanded. Also, knowledge from the present study should be considered in at future implementation.

CONCLUSION

In the present study, we found that experienced neurological therapists knew about UL prediction algorithms. However, only few had a profound knowledge and few were using the SAFE test. The

participants regarded algorithms as potentially useful tools and particularly relevant for recently qualified therapists and for patients with little or no UL function. They were positive about using the two main components in the PREP2 algorithm, the SAFE score and TMS.
Performance of the SAFE score aligned more with the physiotherapy profession than the occupational therapy profession. If PREP2 is to be implemented, PTs may be the ones performing the algorithm. A future implementation strategy should address how to support therapists in handling and delivering predictions, especially if they are negative.

Declaration of interest
The authors have no interests to declare.

Data availability statement
The Danish Data Protection Agency's terms and conditions were complied with, and the data were deleted at the end of the study.

Acknowledgement
We thank the therapists at RHN for participation.

Author statement
Camilla Biering Lundquist: Drafting the study. Developed and revised the interview protocol. The focus group interview was moderated and transcribed by the first author. Analyzing data, writing the article. Final approval of the version to be published.

Hanne Pallesen: Gave advice on qualitative methods. Helped develop and revise the interview protocol. Was an observer, providing feedback to the moderator and observing interactions in the focus groups. Reviewing and discussing the findings. Final approval of the version to be published.

Tine Tjørnhøj-Thomsen: Contributed to the conception of the study. Gave advice on qualitative methods and on the use of Implementation framework. Reviewing and discussing the findings. Final approval of the version to be published.

Iris Brunner: Drafting the study. Helped develop and revise the interview protocol. Reviewing and discussing the findings. Feedback on the article. Final approval of the version to be published.

Funding
We have not received funding for this study.

Figure 1 Predict Recovery Potential (PREP2) algorithm

Footnote for Figure 1:
SAFE: Shoulder Abduction and Finger Extension. < 80 y: Below 80 years old. MEP+: motor evoked potentials present. NIHSS: National Institute of Health Stroke Scale. Excellent: Potential to make a complete, or near complete, recovery of hand and arm function within 3 months. Good: Potential to be using their affected hand and arm for most activities of daily living within 3 months. Limited: Potential to regain some movement in their hand and arm within 3 months. Poor: Unlikely to regain useful movement in their hand and arm within 3 months. Figure copied from the PRESTO homepage.(13)

Figure 2. Diagram showing example of theme formation

Figure 3. The four main themes and their subthemes

References

1. Feigin VL, Krishnamurthi RV, Parmar P, Norrving B, Mensah GA, Bennett DA, et al. Update on the Global Burden of Ischemic and Hemorrhagic Stroke in 1990-2013: The GBD 2013 Study. *Neuroepidemiology.* 2015;45(3):161-76.
2. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. *Heart Disease and Stroke Statistics*-2017 Update: A Report From the American Heart Association. Circulation. 2017;135(10):e146-e603.
3. Faria-Fortini I, Michaelsen SM, Cassiano JG, Teixeira-Salmela LF. Upper extremity function in stroke subjects: relationships between the international classification of functioning, disability, and health domains. *J Hand Ther*: 2011;24(3):257-64.
4. Patel AT, Duncan PW, Lai SM, Studenski S. The relation between impairments and functional outcomes poststroke. *Arch Phys Med Rehab.* 2000;81(10):1357-63.
5. Stinear CM, Byblow WD. Predicting and accelerating motor recovery after stroke. *Curr Opin Neurol.* 2014;27(6):624-30.
6. Kwah LK, Herbert RD. Prediction of Walking and Arm Recovery after Stroke: A Critical Review. *Brain Sci.* 2016;6(4):E53.
7. Stinear CM, Byblow WD, Ackerley SJ, Smith MC, Borges VM, Barber PA. PREP2: A biomarker-based algorithm for predicting upper limb function after stroke. *Ann Clin Transl Neurol.* 2017;4(11):811-20.
8. Winters C, van Wegen EE, Daffertshofer A, Kwakkel G. Generalizability of the Proportional Recovery Model for the Upper Extremity After an Ischemic Stroke. *Neurorehabil Neural Repair.* 2015;29(7):614-22.
9. Byblow WD, Stinear CM, Barber PA, Petoe MA, Ackerley SJ. Proportional recovery after stroke depends on corticomotor integrity. *Ann Neurol.* 2015;78(6):848-59.
10. Hayward KS, Schmidt J, Lohse KR, Peters S, Bernhardt J, Lannin NA, et al. Are we armed with the right data? Pooled individual data review of biomarkers in people with severe upper limb impairment after stroke. *Neuro Image Clin.* 2016;13:310-9.
11. Kim B, Winstein C. Can Neurological Biomarkers of Brain Impairment Be Used to Predict Poststroke Motor Recovery? A Systematic Review. *Neurorehabil Neural Repair.* 2017;31(1):3-24.
12. Stinear CM, Byblow WD, Ackerley SJ, Barber PA, Smith MC. Predicting Recovery Potential for Individual Stroke Patients Increases Rehabilitation Efficiency. *Stroke.* 2017;48(4):1011-9.
13. PRESTO PREP2 [cited 2020 09 09]. Available from: https://presto.auckland.ac.nz/prep2-overview/.

14. PREP2 training [cited 2021 01 07]. Available from: https://preptraining.auckland.ac.nz/.
15. Connell LA, Smith MC, Byblow WD, Stinear CM. Implementing biomarkers to predict motor recovery after stroke. *NeuroRehabilitation.* 2018;43(1):41-50.
16. May C. Towards a general theory of implementation. *Implement Sci*: 2013;8:18-5908-8-18.
17. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci*: 2009;4:50-5908-4-50.
18. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the Consolidated Framework for Implementation Research to Identify Barriers and Facilitators for the Implementation of an Internet-Based Patient-Provider Communication Service in Five Settings: A Qualitative Study. *J Med Internet Res.* 2015;17(11):e262.
19. Connell LA, McMahon NE, Harris JE, Watkins CL, Eng JJ. A formative evaluation of the implementation of an upper limb stroke rehabilitation intervention in clinical practice: a qualitative interview study. *Implement Sci*: 2014;9:90-014-0090-3.
20. Krueger R, Casey M. Planning the Focus Group Study. Focus groups: A Practical Guide for Applied Research 5rd ed ed: SAGE Publications; 2015. 19-37.
21. Wilkinson S. Focus groups in health research: exploring the meanings of health and illness. *J Health Psychol.* 1998;3(3):329-48.
22. Ohman A. Qualitative methodology for rehabilitation research. *J Rehabil Med.* 2005;37(5):273-80.
23. Elo S, Kyngas H. The qualitative content analysis process. *J Adv Nurs.* 2008;62(1):107-15.
24. Malterud K. Qualitative research: standards, challenges, and guidelines. *Lancet* (London, England). 2001;358(9280):483-8.
25. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Health Care Qual.* 2007;19(6):349-57.
26. Kiaer C, Lundquist CB, Brunner I. Knowledge and application of upper limb prediction models and attitude toward prognosis among physiotherapists and occupational therapists in the clinical stroke setting. *Top Stroke Rehabil.* 2020:1-7.
27. Sundhedsstyrelsen. Anbefalinger til nationale redskaber til vurdering af funktionsevne - hos voksne med erhvervet hjerneskade. 2020.
28. Sundhedsstyrelsen. Anbefalinger for tværsektorielle forløb for voksne med erhvervet hjerneskade. 2020
29. Pohl J, Held JPO, Verheyden G, Alt Murphy M, Engelter S, Floel A, et al. Consensus-Based Core Set of Outcome Measures for Clinical Motor Rehabilitation After Stroke-A Delphi Study. *Front Neurol.* 2020;11:875.
30. Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Int J Stroke.* 2017;12(5):451-61.
31. Gebruers N, Truijen S, Engelborghs S, De Deyn PP. Prediction of upper limb recovery, general disability, and rehabilitation status by activity measurements assessed by accelerometers or the Fugl-Meyer score in acute stroke. *Am J Phys Med Rehabil.* 2014;93(3):245-52.

32. Prabhakaran S, Zarahn E, Riley C, Speizer A, Chong JY, Lazar RM, et al. Inter-individual variability in the capacity for motor recovery after ischemic stroke. *Neurorehabil Neural Repair.* 2008;22(1):64-71.
33. Buch ER, Rizk S, Nicolo P, Cohen LG, Schnider A, Guggisberg AG. Predicting motor improvement after stroke with clinical assessment and diffusion tensor imaging. *Neurology.* 2016;86(20):1924-5.
34. Winstein C, Varghese R. Been there, done that, so what's next for arm and hand rehabilitation in stroke? *NeuroRehabilitation.* 2018;43(1):3-18.
35. Jones F, Postges H, Brimicombe L. Building Bridges between healthcare professionals, patients and families: A coproduced and integrated approach to self-management support in stroke. *NeuroRehabilitation.* 2016;39(4):471-80.
36. Mayo NE, Fellows LK, Scott SC, Cameron J, Wood-Dauphinee S. A longitudinal view of apathy and its impact after stroke. *Stroke.* 2009;40(10):3299-307.
37. Subramanian SK, Chilingaryan G, Sveistrup H, Levin MF. Depressive symptoms influence use of feedback for motor learning and recovery in chronic stroke. *Restor Neurol Neurosci.* 2015;33(5):727-40.
38. Nijland RH, van Wegen EE, Harmeling-van der Wel BC, Kwakkel G, Early Prediction of Functional Outcome After Stroke I. Accuracy of physical therapists' early predictions of upper-limb function in hospital stroke units: the EPOS Study. *Phys Ther.* 2013;93(4):460-9.

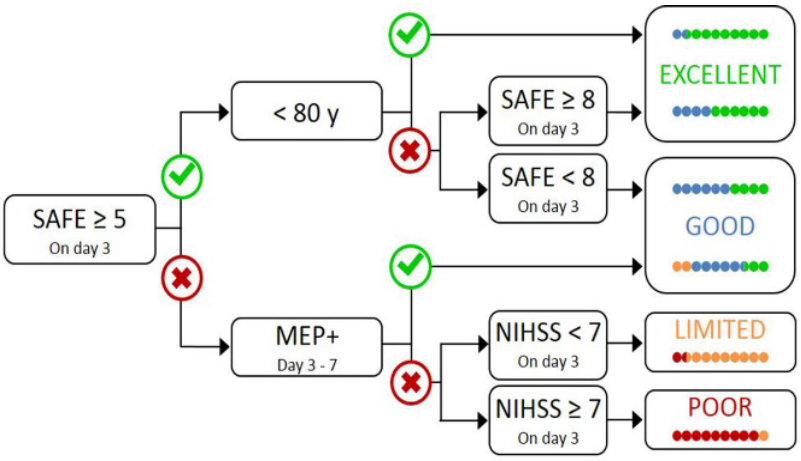
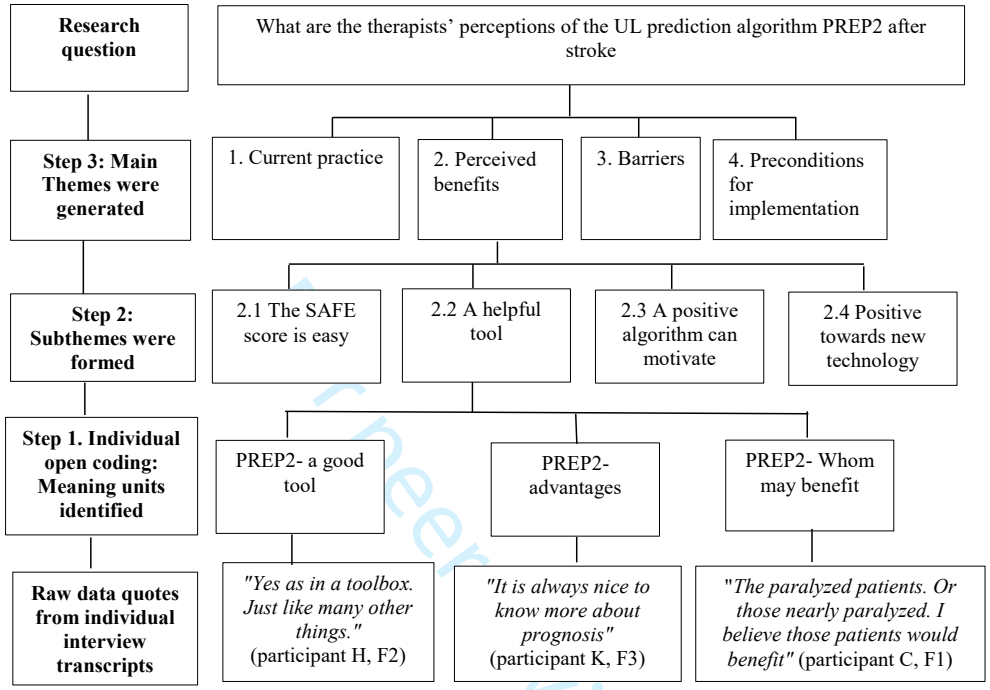
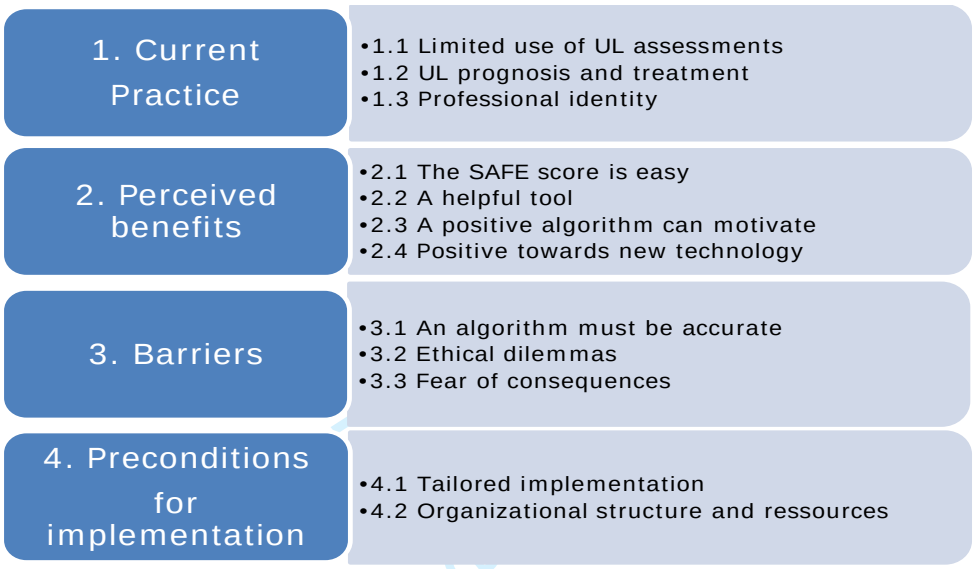


Figure 1 Predict Recovery Potential (PREP2) algorithm





Declaration of co-author ship for paper 1

AARHUS UNIVERSITY

Declaration of co-authorship concerning article for PhD dissertations

Full name of the PhD student: Camilla Biering Lundquist

This declaration concerns the following article/manuscript:

Title:	Accuracy of the Upper Limb Prediction Algorithm PREP2 applied 2 weeks poststroke: A prospective longitudinal Study.
Authors:	Lundquist CB, Nielsen JF, Arguissain FG, Brunner IC

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preperation ☐

If published, state full reference: Accuracy of the Upper Limb Prediction Algorithm PREP2 applied 2 weeks poststroke: A prospective longitudinal Study Camilla Biering Lundquist, Jørgen Feldbæk Nielsen, Federico Gabriel Arguissain, Iris Charlotte Brunner. Neurorehabil and Neural Repair 2020.1-11

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

Your contribution

Please rate (A-F) your contribution to the elements of this article/manuscript, **and** elaborate on your rating in the free text section below.

A. Has essentially done all the work (>90%)
B. Has done most of the work (67-90 %)
C. Has contributed considerably (34-66 %)
D. Has contributed (10-33 %)
E. No or little contribution (<10%)
F. N/A

Category of contribution	Extent (A-F)
The conception or design of the work:	B
Free text description of PhD student's contribution ***(mandatory)*** The conception of the design and planning of the study, data management and analysis, and interpretation have mainly been done by the student.	
The acquisition, analysis, or interpretation of data:	B
Free text description of PhD student's contribution ***(mandatory)*** Data have been acquired by independent raters according to procedures developed by the PhD student. Data management, and data analysis have been mainly done by the student. Interpretation of data has been a collaboration between co-authors with the student as the primary decision-maker.	
Drafting the manuscript:	B
Free text description of PhD student's contribution ***(mandatory)*** The student has drafted all versions of the text and then included co-author comments and corrections.	
Submission process including revisions:	B

AARHUS
UNIVERSITY

Submission process including revisions:	Submitted june 2020. 1 revision made sep 2020. Published nov 2020.
Free text description of PhD student's contribution ***(mandatory)*** The student has completed the submission process on her own. She has included the co-authors in the response to reviewers and re-submission	

Signatures of first- and last author, and main supervisor

Date	Name	Signature
16.02.2021	Camilla Biering Lundquist	16/2/21
17.02.21	Inj A. Bjunnar	Inj A. Bjunnar

Date: 16/2/21

Signature of the PhD student

Declaration of co-author ship for paper 2

AARHUS UNIVERSITY

Declaration of co-authorship concerning article for PhD dissertations

Full name of the PhD student: Camilla Biering Lundquist

This declaration concerns the following article/manuscript:

Title:	Prediction of Upper Limb use three months after stroke: A prospective longitudinal study
Authors:	Camilla Biering Lundquist, Jørgen Feldbæk Nielsen, Iris Charlotte Brunner

The article/manuscript is: Published ☐ Accepted ☐ Submitted ☒ In preperation ☐

If published, state full reference:

If accepted or submitted, state journal: Disability and Rehabilitation

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

Your contribution

Please rate (A-F) your contribution to the elements of this article/manuscript, **and** elaborate on your rating in the free text section below.

A. Has essentially done all the work (>90%)
B. Has done most of the work (67-90 %)
C. Has contributed considerably (34-66 %)
D. Has contributed (10-33 %)
E. No or little contribution (<10%)
F. N/A

Category of contribution	Extent (A-F)
The conception or design of the work:	B
Free text description of PhD student's contribution ***(mandatory)*** The conception of the design and planning of the study, data management and analysis, and interpretation have mainly been done by the student.	
The acquisition, analysis, or interpretation of data:	B
Free text description of PhD student's contribution ***(mandatory)*** Data have been acquired by independent raters according to procedures developed by the PhD student. Data management, and data analysis have been mainly done by the student. Interpretation of data has been a collaboration between the student and the co-author, with the student as the primary decision-maker.	
Drafting the manuscript:	B
Free text description of PhD student's contribution ***(mandatory)*** The student has drafted all versions of the text and then included co-authors comments and corrections.	
Submission process including revisions:	B

AARHUS
UNIVERSITY

Free text description of PhD student's contribution ***(mandatory)***
The student has completed the manuscript and it will be submitted within the newt two weeks

Signatures of first- and last author, and main supervisor

Date	Name	Signature
16.02.2021	Camilla Biering Lundquist	
17.02.21	Iris Charlotte Bjerrum	

Date: 16/2/21

Signature of the PhD student

Declaration of co-author ship for paper 3

AARHUS UNIVERSITY

Declaration of co-authorship concerning article for PhD dissertations

Full name of the PhD student: Camilla Biering Lundquist

This declaration concerns the following article/manuscript:

Title:	Exploring physiotherapists' and occupational therapists' perceptions of the upper limb prediction algorithm PREP2 after stroke in a rehabilitation setting. A qualitative study.
Authors:	Camilla Biering Lundquist, Hanne Pallesen, Tine Tjørnhøj-Thomsen, Iris Brunner

The article/manuscript is: Published ☐ Accepted ☐ Submitted ☒ In preperation ☐

If published, state full reference:

If accepted or submitted, state journal: BMJ Open

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

Your contribution

Please rate (A-F) your contribution to the elements of this article/manuscript, **and** elaborate on your rating in the free text section below.

A. Has essentially done all the work (>90%)
B. Has done most of the work (67-90 %)
C. Has contributed considerably (34-66 %)
D. Has contributed (10-33 %)
E. No or little contribution (<10%)
F. N/A

Category of contribution	Extent (A-F)
The conception or design of the work:	B
Free text description of PhD student's contribution ***(mandatory)*** The conception of the qualitative study design and planning of the focus group interviews, data management and analysis, and interpretation have mainly been done by the student.	
The acquisition, analysis, or interpretation of data:	B
Free text description of PhD student's contribution ***(mandatory)*** The student drafted the study. Developed and revised the interview protocol. The focus group interview was moderated and transcribed by the the student. The student was analyzing data, writing the article. The student approved of the final version to be published.	
Drafting the manuscript:	B
Free text description of PhD student's contribution ***(mandatory)*** The student has drafted all versions of the text and then included co-author comments and corrections.	
Submission process including revisions:	B

AARHUS
UNIVERSITY

Free text description of PhD student's contribution ***(mandatory)***
The student has completed the submission process on her own. During the revision all co-authors were contacted and engaged in discussions.

Signatures of first- and last author, and main supervisor

Date	Name	Signature
16.02.2021	Camilla Biering Lundquist	
17.02.21	Iris Charlotte Wiggers	

Date: 16/2/21

Signature of the PhD student