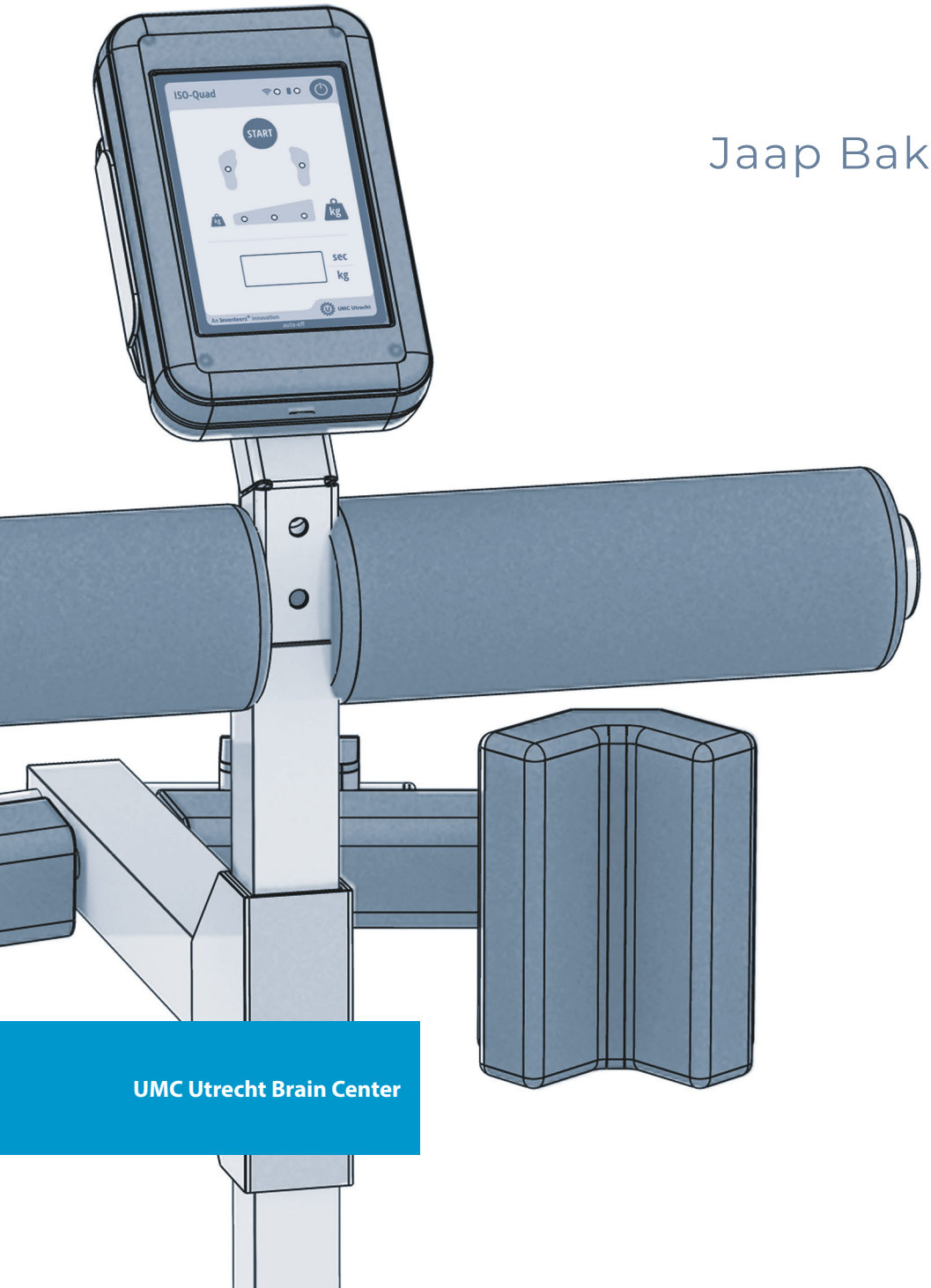
TOWARDS REMOTE CLINICAL OUTCOME MEASUREMENTS IN MOTOR NEURON DISORDERS

Jaap Bakers



UMC Utrecht Brain Center

Towards remote clinical outcome measurements in motor neuron disorders

Jaap Bakers

Towards remote clinical outcome measurements in motor neuron disorders

Jaap Bakers

ISBN: 978-94-6483-058-3

Provided by thesis specialist Ridderprint, ridderprint.nl
Printing: Ridderprint
Layout and design: Harma Makken, persoonlijkproefschrift.nl

The printing of this thesis was financially supported by the Scientific College Physical Therapy (WCF) of the Royal Dutch Society for Physical Therapy (KNGF), Inventeers Research and Development, and ProCare BV.

Towards remote clinical outcome measurements in motor neuron disorders

Het thuismeten van klinische uitkomstmaten bij motorische zenuwziekten
(met een samenvatting in het Nederlands)

Proefschrift

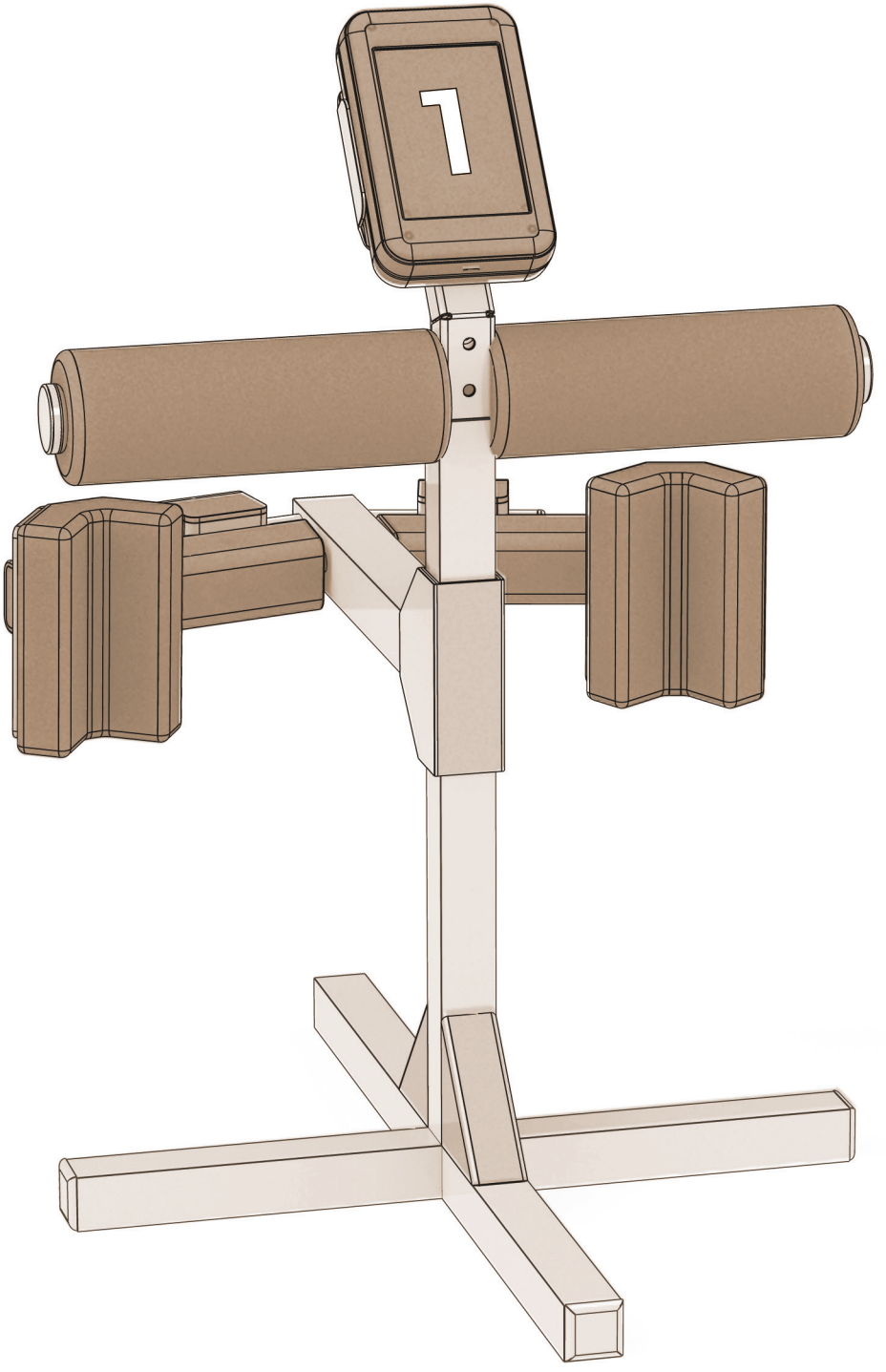
ter verkrijging van de graad doctor aan de
Universiteit Utrecht op gezag van de rector magnificus,
prof. Dr. H.R.B.M. Kummeling, ingevolge het besluit van het
college voor promoties in het openbaar te verdedigen
op dinsdag 23 mei 2023 des middags te 4.15 uur

door

Japie Nelle Engel Bakers
Geboren op 15 juni 1982 te Breda

Promotoren	Prof. dr. L.H. van den Berg
	Prof. dr. J.M.A. Visser-Meily
Copromotoren	Dr. J.A.J Beelen
	Dr. R.P.A. van Eijk
Beoordelingscommissie	Prof. dr. F. Nollet
	Prof. dr. W.L. van de Pol
	Prof. dr. S.C.C.M. Teunissen (voorzitter)
	Prof. dr. C. Veenhof
	Prof. dr. P.J. Wijksta

Contents


1

Chapter 1

GENERAL INTRODUCTION

Motor neuron disorders: ALS and MMN

Motor neuron disorders are a collection of syndromes in which central and/or peripheral motor neurons are affected. These syndromes show extensive variability in their clinical presentation and disease course. This thesis focusses on two motor neuron disorders that reflect this variability in disease pattern and course: amyotrophic lateral sclerosis (ALS) and multifocal motor neuropathy (MMN). The university medical center Utrecht provides national expertise for both ALS and MMN.

ALS is characterised by fast progressive muscle weakness with respiratory muscle involvement and spasticity. The median survival time is three to five years after the first symptom onset, but varies greatly from person to person.[1, 2] ALS develops, on average, between the ages of 45 an 65 years.[3] A rapid disease course is typical for these patients, who, at disease onset, usually very actively participate in society.[4] In addition to different survival times, the pattern of functional loss is also highly variable.[5, 6] Therefore, ALS has many different manifestations, making personalised care as well as clinical research complex.

MMN has a less aggressive, non-fatal course, without involvement of respiratory muscles. It is characterised by a relatively slowly progressive, predominantly distal limb muscle weakness. Unlike ALS, there is a treatment option available to MMN patients that can significantly slow down disease progression: the use of intravenous immunoglobulins (IVIg).[7]

Despite the differences between ALS and MMN, their peripheral motor symptoms show similarities and so the use of clinical outcome measurements in care and clinical research is partially alike for the two disorders. In the application of clinical outcome measurements to determine the optimal dosing and treatment interval of an effective treatment, MMN can be seen as a precursor to ALS.[8]

Clinical outcome measurements in the context of care and experimental research

Progressive loss of muscle strength is a hallmark of motor neuron disorders, causing severe impairment of functional ability. Therefore, the assessment of motor function in motor neuron disorders, using clinical outcome measurements, is relevant in care as well as in clinical research. However, selecting generic clinical outcome

measurements is difficult due to heterogeneity in disease symptoms and course in motor neuron disorders.

In the context of care, clinical outcome measurements are critical for detecting differences in signs and symptoms, disease course, and prognosis, in patients who are otherwise deceptively similar because of the same diagnosis or nominal classifications of neurophysiological criteria (e.g. possible, probable, definite).[9, 10] The recent availability of patient-friendly technology offers the opportunity to collect a selection of clinical outcome measurements remotely, causing a shift from in-clinic to local and home-based assessments of measurements. Remote measurements have the advantage of allowing frequent and real-time monitoring of disease symptoms, thus improving continuity of care by timely detection of a change in disease symptoms. Understanding which disease symptoms are present, and their development over time, is a prerequisite for providing personalised care at the place of preference.[11, 12]

The general premise for therapeutic interventions in motor neuron diseases is to improve motor neuron function. Selection of appropriate outcome measures is critical to the success of clinical trials. In MMN, intravenous immunoglobulin (IVIg) is the first-line treatment and its efficacy in improving muscle strength has been confirmed repeatedly.[7] The conventional outcome measurements that were applied in these trials are, therefore, still widely supported and applied for optimisation of IVIg treatment interval and dosage.[8] However, as currently available pharmacological treatment options only provide a limited prolongation of the life expectancy in ALS, the search for an effective treatment is still in full swing.[13] In the absence of definitive diagnostic and prognostic biomarkers,[14, 15] alternative efficacy endpoints are used to evaluate treatment effects in clinical trials. With the exception of survival time, these endpoints are mostly related to motor function due to progressive muscle strength loss, including respiratory dysfunction.[16] Especially in ALS, optimised evaluation of disease-related symptoms is of great importance due to the heterogeneous population with a low prevalence.[17] In order to optimise research, international collaboration is therefore necessary, and is warranted by several ALS trial networks worldwide. The Trial Research Initiative to Cure ALS (TRICALS) is an example of an ALS-specific trial network, consisting of 40 large specialised ALS centers in 14 European countries that have agreed on a standard core set of clinical outcome measurements.[18] Such a core set offers several advantages, such as the opportunity to compare the results across studies, and to perform subgroup

analyses. The TRICALS core set contains six outcome measures, including a self-reported symptoms questionnaire (ALSFRS-r), respiratory function, and isometric muscle strength.

Despite considerable efforts and progress in experimental research, that offer great potential for new treatment options,[19] the development of effective treatments has been hampered in ALS. As progressive motor neuron degeneration is for a large part responsible for loss of motor function over time, it is essential to include clinical outcome measures in trial designs to understand the effect of new therapies in ALS. However, in general, the quality of clinical measurements depends not only on the measurement properties of the instrument, but also on human performance during assessment: for example, on the expertise of the assessor carrying out the measurement, or the degree of attention and commitment of a patient during a measurement.[20] In the search for an effective treatment it is, therefore, important to explore whether the current core set of clinical outcome measurements is optimized to quantify treatment effects in ALS, also taking into account the technological advances that facilitate remote measurements.

This thesis focusses on clinical outcome measurements that assess isometric muscle strength, respiratory function, and functional ability. Isometric muscle strength can be assessed by hand-held or fixed dynamometry; respiratory function by respiratory strength, flow, or volume testing.[21] Functional ability is traditionally assessed by standardised motor function tests and questionnaires. Technological measurement instruments, such as accelerometers, are, however, currently also applied.

Aim of the thesis

In the area of motor neuron disorders, there is a lack of knowledge about how clinical outcome measures can optimally contribute to the delivery of personalised care, and to the evaluation of effectiveness during experimental research. Therefore the aim of this thesis is to gain insight into the current use and shortcomings of in-clinic assessed outcome measurements, and subsequently to explore the value of remote, independent measurements.

Outline of this thesis

This thesis is divided into three parts. The first part focuses on the clinical relevance of clinical outcome measurements in ALS and MMN. The second part identifies important shortcomings of clinical outcome measurements that are relevant for ALS care and research. The third part provides insight into how technological advances in remote outcome assessment can enhance the quality and reduce the burden of clinical outcome assessment in ALS.

Part 1: The relevance of clinical outcome measurements
In chapter 2, recommendations are provided about which respiratory function tests are relevant for the early detection of respiratory dysfunction, and which supportive interventions are available in ALS.

Chapter 3 describes that not only distal, but also proximal muscle groups improve in muscle strength following IVIg therapy in MMN.

Chapter 4 compares intravenous and subcutaneous treatment with immunoglobulins, using the course of muscle strength as an outcome measurement.

Part 2: Current limitations of clinical outcome measurements
Chapter 5 provides insight into how symptomatic treatment effects of usual care can cause unexpected, large improvements in the ALSFRS-r during experimental research.

Chapter 6 shows that variability between spirometric reference values has an effect on clinical decision-making and selection for clinical trials.

Part 3: Innovations of clinical outcome measurements
Chapter 7 shows that, when compared to hand-held dynamometry, portable fixed dynamometry reduces examiner-induced ceiling effects and improves the reliability of the measurement of isometric quadriceps muscle strength in ALS.

Chapter 8 explores accelerometry as an objective method to quantify physical functioning and, therefore, disease progression in ALS.

Chapter 9 shows that unsupervised respiratory function testing at home is a valid and well-accepted method for monitoring respiratory function.

References

1. Testa D, Lovati R, Ferrarini M, Salmoiraghi F, Filippini G. Survival of 793 patients with amyotrophic lateral sclerosis diagnosed over a 28-year period. Amyotroph Lateral Scler Other Motor Neuron Disord. 2004;5:208-212.
2. Czaplinski A, Yen AA, Appel SH. Amyotrophic lateral sclerosis: early predictors of prolonged survival. J Neurol. 2006;253:1428-1436.
3. Marin B, Logroscino G, Boumediene F, Labrunie A, Couratier P, Babron MC, et al. Clinical and demographic factors and outcome of amyotrophic lateral sclerosis in relation to population ancestral origin. Eur J Epidemiol. 2016;31:229-245.
4. Kruitwagen-Van Reenen ET, Scholten EWM, van Groenestijn A, Van Den Berg LH, Post MWM, Visser-Meily J. Participation and autonomy in the first 10 months after diagnosis of ALS: a longitudinal study. Amyotroph Lateral Scler Frontotemporal Degener. 2021;22:459-466.
5. Westeneng HJ, Debray TPA, Visser AE, van Eijk RPA, Rooney JPK, Calvo A, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. Lancet Neurol. 2018;17:423-433.
6. van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, et al. Amyotrophic lateral sclerosis. Lancet. 2017;390:2084-2098.
7. van Schaik IN, van den Berg LH, de Haan R, Vermeulen M. Intravenous immunoglobulin for multifocal motor neuropathy. Cochrane Database Syst Rev. 2005:CD004429.
8. Berger M, Allen JA. Optimizing IgG therapy in chronic autoimmune neuropathies: a hypothesis driven approach. Muscle Nerve. 2015;51:315-326.
9. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. J Neurol Sci. 1994;124 Suppl:96-107.
10. Olney RK, Lewis RA, Putnam TD, Campellone JV, Jr., American Association of Electrodiagnostic M. Consensus criteria for the diagnosis of multifocal motor neuropathy. Muscle Nerve. 2003;27:117-121.
11. Haulman A, Geronimo A, Chahwala A, Simmons Z. The Use of Telehealth to Enhance Care in ALS and other Neuromuscular Disorders. Muscle Nerve. 2020;61:682-691.
12. Dontje ML, Kruitwagen-van Reenen E, Visser-Meily JMA, Beelen A, Study Group ALSH-m, Coaching. Implementation and evaluation of an e-health innovation for personalized care for patients with amyotrophic lateral sclerosis (ALS): protocol for a participatory action research study. Implement Sci Commun. 2021;2:25.
13. Kiernan MC, Vucic S, Talbot K, McDermott CJ, Hardiman O, Shefner JM, et al. Improving clinical trial outcomes in amyotrophic lateral sclerosis. Nat Rev Neurol. 2021;17:104-118.
14. Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, et al. Amyotrophic lateral sclerosis. Nat Rev Dis Primers. 2017;3:17085.
15. Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, et al. Amyotrophic lateral sclerosis. Lancet. 2011;377:942-955.
16. Goyal NA, Berry JD, Windebank A, Staff NP, Maragakis NJ, van den Berg LH, et al. Addressing heterogeneity in amyotrophic lateral sclerosis CLINICAL TRIALS. Muscle Nerve. 2020;62:156-166.
17. van Eijk RPA, Nikolakopoulos S, Roes KCB, Kendall L, Han SS, Lavrov A, et al. Innovating Clinical Trials for Amyotrophic Lateral Sclerosis: Challenging the Established Order. Neurology. 2021;97:528-536.

18. van Eijk RPA, Kliest T, McDermott CJ, Roes KCB, Van Damme P, Chio A, et al. TRICALS: creating a highway toward a cure. Amyotroph Lateral Scler Frontotemporal Degener. 2020;21:496-501.

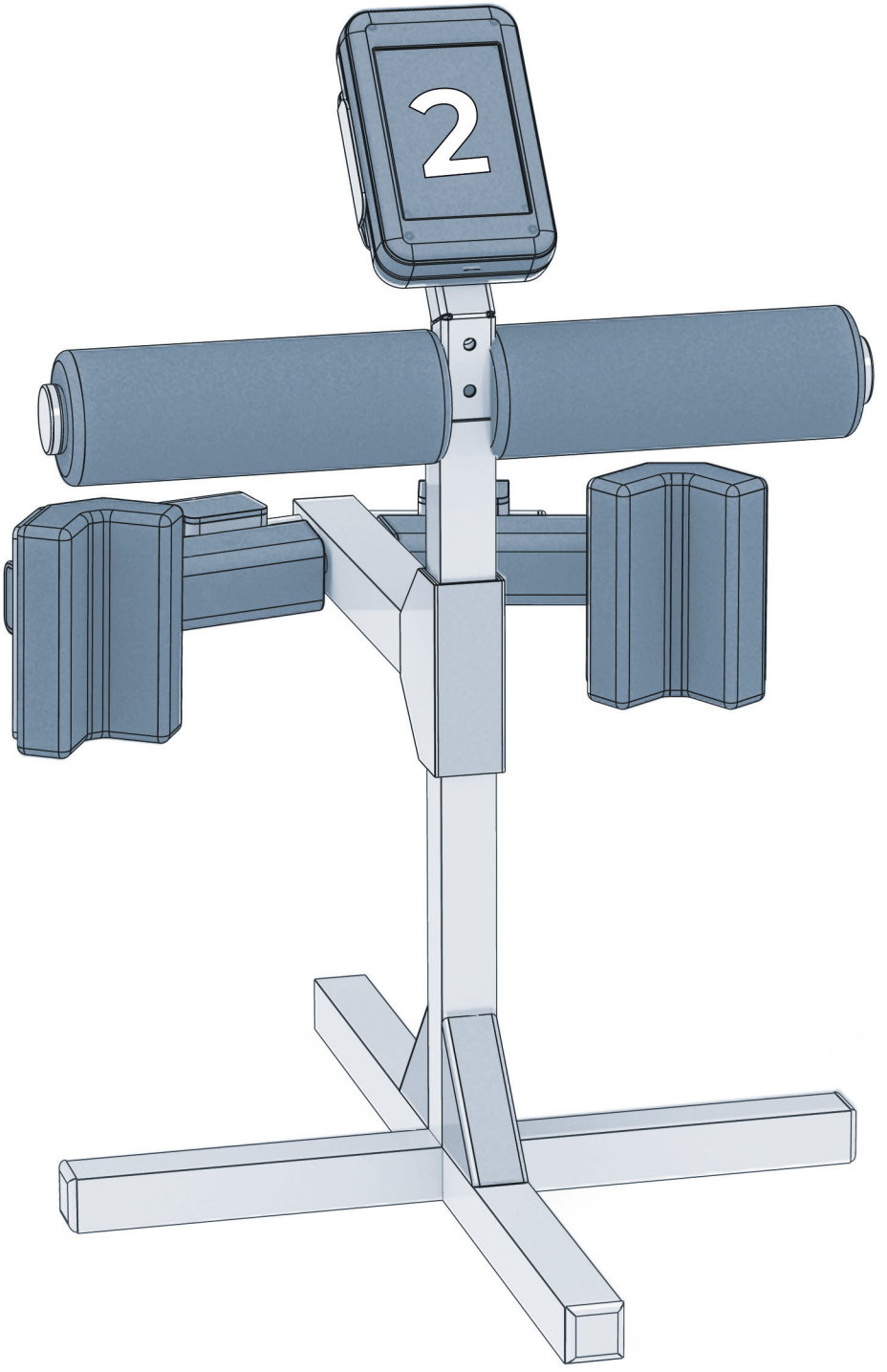
19. Sahdeo S, Goldstein DB. Translating amyotrophic lateral sclerosis genes into drug development leads. Nat Genet. 2021;53:1624-1626.

20. de Vet HC, Terwee CB, Bouter LM. Current challenges in clinimetrics. J Clin Epidemiol. 2003;56:1137-1141.

21. Lechtzin N, Cudkowicz ME, de Carvalho M, Genge A, Hardiman O, Mitsumoto H, et al. Respiratory measures in amyotrophic lateral sclerosis. Amyotroph Lateral Scler Frontotemporal Degener. 2018;19:321-330.

Part 1

THE RELEVANCE OF CLINICAL OUTCOME MEASUREMENTS IN MOTOR NEURON DISORDERS


2

Chapter 2

CLINICAL PRACTICE RECOMMENDATIONS FOR EARLY RESPIRATORY MANAGEMENT IN PATIENTS WITH ALS, PRIOR TO INITIATION OF VENTILATORY SUPPORT

Jaap Bakers
Esther Kruitwagen van Reenen
Peter Wijkstra
Leonard van den Berg
Anne Visser-Meily
Anita Beelen

Submitted

Abstract

Background: Respiratory testing at regular clinic visits can be burdensome for patients with ALS. There is a need for methods to monitor respiratory dysfunction, and interventions for symptomatic management that are easy to perform.

Objective: To provide evidence-based recommendations for (1) the assessment of respiratory function to early detect respiratory dysfunction and inefficient cough; (2) non-pharmacological interventions to improve inefficient cough and respiratory muscle weakness.

Methods: Systematic reviews were undertaken to summarize the evidence. Recommendations incorporated the evidence, values and preferences of patients and carers, clinical expertise and recommendations made in relevant guidelines.

Recommendations: 1. Assessment of respiratory function should comprise the evaluation of signs and symptoms of respiratory weakness, combined with sitting Vital Capacity (VCsit). However, VCsit appears not sufficiently sensitive and should be supplemented with supine Vital Capacity or -in case of orthopnoea- with Maximal Inspiratory Pressure, or Sniff Nasal Inspiratory Pressure. Cough strength should be assessed by Peak Cough Flow (PCF).

2. In case of inefficient cough (PCF <270L/m), or earlier with signs and symptoms of reduced cough ability, airway clearance techniques, especially airstacking, are indicated. Inspiratory and expiratory muscle training are not recommended for routine use.

Conclusions: Evaluating signs and symptoms of respiratory weakness in combination with relevant respiratory function tests are recommended for the early detection of respiratory dysfunction and inefficient cough in ALS and have the potential for remote monitoring. In the case of inefficient cough, airway clearance techniques, that are easy to perform are available before ventilator support is needed.

Introduction

In patients with Amyotrophic Lateral Sclerosis (ALS), timely assessment and management of respiratory dysfunction is challenging due to heterogeneity in time of onset, rate of decline in respiratory function, and symptoms of respiratory weakness (1). Early recognition of respiratory dysfunction is important to initiate prompt interventions that can alleviate complaints (2).

Current clinical guidelines for respiratory management in ALS direct to quantification of respiratory function in a clinical setting (3, 4). Measurement of respiratory function in ALS can be complicated due to several factors, for example facial muscle weakness, cognitive impairments, and difficulties in coordination of respiratory movements (5). Respiratory function tests are therefore typically assessed by expert pulmonologists in a clinical setting. However, hospital visits may be burdensome for patients, making unnecessary hospital visits for monitoring and symptom management are undesirable. Nevertheless, during the build-up to respiratory insufficiency, complaints due to early respiratory dysfunction may already arise (6). Accelerated through the COVID-19 pandemic, remote monitoring and -management are proposed as effective tools to reduce burden of hospital visits, and facilitate self-management of patients as well as caregivers (7). Together with an increasing availability of patient friendly digital technology, this development is expected to cause a future shift to more local and home-centred care (8). Therefore, there is a need for easy to perform assessments to monitor respiratory function, including cough effectiveness, and interventions that can be carried out at home with minimal instruction and supervision. As part of the development of a clinical practice guideline for physiotherapy in ALS, we therefore aimed to provide evidence-based recommendations for respiratory management before ventilatory support starts.

In this paper, we describe the systematic review of evidence, and the development of key recommendations for 1) respiratory function tests that are indicative for early respiratory dysfunction and inefficient cough, and 2) non-pharmacological interventions to improve inefficient cough and respiratory muscle weakness.

Methods

Identification of key clinical questions

Surveys were conducted to identify topics that provide quality improvement opportunities and were sent to professionals of 38 Dutch multidisciplinary ALS teams.

A separate survey eliciting the patients' perspectives on quality improvement of ALS Care was developed for and sent to ALS patients and their carers. Based on these surveys, a national Practice Recommendations Development Group (PRDG), and the steering committee, prioritized the topics to determine the focus of the guidelines and established the research questions. The PRDG consisted of three expert physiatrists, and two expert neurologists, one expert physiotherapist and one methodologist.

Strategy for searching and selecting literature

A systematic literature search for existing ALS guidelines and other relevant scientific publications was performed in the electronic data bases Medline and EMbase. The review protocols (including critical outcomes), and search strings used are described in **Supplement I**. Furthermore, cross-references and expert recommended references were evaluated, and a search was carried out for studies using the literature reference lists of the selected articles. To be selected, publications had to address humans and be published in English or Dutch. Publications were selected if at least one of the critical outcome measures was reported, and sufficient data were reported.

Quality assessment of individual studies

The methodological quality of individual studies was systematically assessed on the basis of internationally held methodological quality criteria, in order to allow estimation of bias.

Extracting practice recommendations

On the basis of the systematic literature search, practice recommendations were deduced according to international standards for guideline development (9). This development process was guarded by the PRDG.

Clinical expertise and patient preferences

To guarantee implementability of the clinical practice recommendations, expertise from patient representatives, an expert panel, and the PRDG was incorporated in the evidence synthesis. Two patient representatives of the Dutch ALS patient association (ALS Patients Connected), one scientific member of the Dutch patient organisation for neuromuscular diseases (Spierziekten Nederland), and one scientific member of the Royal Dutch Society for Physiotherapy (KNGF) reviewed a draft of the clinical recommendations. In addition an independent review panel of 8 Dutch professionals with specific expertise on respiratory dysfunction in ALS (pulmonologist, respiratory physiologist, neurologists, physiatrists and physiotherapists) reviewed a draft of the

practice recommendations. Their comments were collected and discussed with the PRDG and the draft guideline was amended. The final guideline was submitted to the Royal Dutch Society for Physiotherapy and both patient organisations for authorisation, and was published online in December 2019.

Results and clinical practise recommendations

Respiratory function tests that are indicative for early respiratory dysfunction and inefficient cough

Key evidence

Nine studies were included in the review. The included papers were graded as low quality evidence as the studies were either observational in nature (4) or had a high risk of bias (5). All outcomes/respiratory function parameters reported were graded as either Low or Very Low quality for detecting respiratory dysfunction.

Evidence from these studies are summarised in **table 1**. Six studies showed that vital capacity (VC) in sitting position is not sensitive enough to indicate an early stage of respiratory dysfunction (10-15) There are indications from the literature for superiority of maximal inspiratory pressure (MIP) (10, 11) sniff nasal inspiratory pressure (SNIP) (10, 12, 14) and supine VC (15, 16) over sitting VC in detecting respiratory dysfunction. However there is insufficient evidence available to demonstrate the superiority of one specific parameter of respiratory function tests.

One study showed that in case of severe weakness in the bulbar region, respiratory function tests have insufficient predictive power for respiratory insufficiency (12). There is evidence that the PCF is an overall parameter of strength of the respiratory muscles involved in the cough reflex (17).

Table 1. Identified studies for key recommendations of respiratory function tests that are indicative for early respiratory dysfunction and inefficient cough.

Author, year	Sample (n)	Study design	Study aim	Results and authors' conclusion	**RoB score**
Bourke, 2003 (22)	22	Cohort	To determine the optimal criteria for initiation of NIV.	9 of the 11 participants with orthopnea continued using NIV after the initial trial, with a large improvement in QoL and excellent compliance. Of these 11 participants with orthopnea, 5 participants had normal PaCo2 levels. Within these 5 participants, 4 participants continued with NIV.	Low
Jackson, 2001 (11)	20	Cohort	To determine which lung function parameter correlates best with symptoms of early nocturnal hypoventilation	The identification of nocturnal oxygen desaturation < 90% for at least one cumulative minute appears to be the most sensitive indicator for detecting early respiratory insufficiency in this study. The mean FVC% at the time nocturnal oxygen desaturations were initially identified was 77%, which is much higher than the FVC% currently recommended for the initiation of NPPV. All 13 patients who were documented to have nocturnal oxygen desaturation <90% also had at least two clinically significant symptoms of nocturnal hypoventilation based upon their responses to the pulmonary symptom scale, suggesting that these desaturations were indeed clinically significant.	Low
Lechtzin, 2002 (16)	25	Cross-sectional	To determine which respiratory test (FVC sitting and in supine, MIP, MEP and paCO2) predicts diaphragm weakness (measured by transdiaphragmic pressure: Pdi) best.	Supine FVC as a percent of the predicted value was more closely correlated with Pdi-sniff than any other predictor variable, and explained 76% of the variance among the observed values of Pdi-sniff. Supine FVC can be used to predict Pdi-sniff using the following egression equation: Pdi-sniff = -10.52 + (0.91 x supine FVC percent predicted). Though MIP was not significantly associated with Pdi-sniff, it is a commonly used measure of inspiratory muscle strength and we therefore examined its contribution to a model containing supine FVC (percent predicted). This model is an excellent predictor of Pdi-sniff ($p < 0.001$, $R2 = 0.93$). The regression equation to predict Pdi-sniff from this model is as follows: Pdi-sniff = - 3.22 + (1.03 x supine FVC) + (0.23 x MIP).	High
Lyall, 2001 (12)	81	Cross-sectional	To determine the value for predicting hypercapnia (expressed in pCO2 of ≥6 kPa) and the relationship between the VC, MIP, MEP and SNIP.	Predictive value of SNIP (with cut-off value of <30 cm H_2O) for detection of hypercapnia was higher (OR 25, specificity 85%, sensitivity 81%) than that of upright VC (cut-off 50% predicted; OR 9, specificity 89%, sensitivity 53%) and % MIP (cut-off 25% predicted; OR 6, 83%, 55%). In the subgroup of patients with significant bulbar weakness (Norris Bulbar Score of 22 (±8)), none of the tests had significant predictive power for the presence of hypercapnia.	Low
Mendoza, 2007 (13)	161	Retro-spective cohort	To compare the sensitivity of the MIP (<60 cm H20) and of the FVC (<50% of the predicted value) for NIV initiation.	At NIV enrolment, more patients met the MIP-criterion (65% 71/109), than the upright FVC criterion (8%, 9/109). The MIP is a more sensitive indicator of early respiratory insufficiency in ALS than FVC. Variability was notably higher in MIP tests than upright FVC tests.	High

Table 1. *(continued)*

Author, year	Sample (n)	Study design	Study aim	Results and authors' conclusion	**RoB score**
Varrato, 2001 (15)	47	Cohort	To determine the usefulness of the difference between the FVC in sitting and supine position, for respiratory evaluation of ALS patients.	The average decrease in %FVC from erect to supine (e-s%FVC) was 15.6% in patients with ALS (range 0 to 45%). These results reached significance. Of 46 patients, 43 had an upright FVC of greater than 50%. Seven of 43 (16.2%) fell below 50% when supine. The largest values for e-s%FVC occurred in patients with dyspnea, orthopnea, and fatigue, in which an average value of 19.6%, 21.4%, and 26.3%, respectively, was noted (*p* 0.074, 0.018, and 0.001, respectively). There was no significant relationship between the eFVC and orthopnea.	Low
Fregonezi, 2013 (10)	31	Cross-sectional	To determine the relationship between the Forced Vital Capacity (FVC) and respiratory muscle strength (MIP / MEP / SNIP) in healthy people and people with ALS.	In ALS, monitoring respiratory muscle strength assists in early diagnosis of respiratory dysfunction as opposed to the isolated use of FVC. Sensitivity and specificity for MIP, MEP and SNIP in detecting weakness in the respiratory muscles was 75/58%, 81/67% and 75/67%, respectively. ROC curve indicated that the MIP, MEP and SNIP can identify differences in respiratory muscle strength between ALS and healthy individuals at 0.89, 0.9 and 0.82, respectively.	High
Tilanus, 2017 (14)	110	Retro-spective cohort	To determine which respiratory function tests (FVC, PCF, MIP, MEP and SNIP) are most suitable for predicting NIV initiation within 3 months.	Out of 5 respiratory function tests (FVC, PCF, MIP, MEP and SNIP,) the PCF (cut-off value 386 L/m, sensitivity 88%, specificity 36%) and SNIP (cut-off value 45 cm H_2O, sensitivity 87%, specificity 40%) had the best predictive value for NIV initiation in ALS patients.	Low
Tzani, 2014 (17)	49	Cross sectional	To determine the value of the respiratory muscle strength (MIP / MEP), stomach pressure (Cough Pgas), and Peak Cough Flow (PCF) for detecting an ineffective cough	When related to the parameters of expiratory muscle strength, PCF showed a positive correlation with MEP (r=0,771, P<0,001), Cough Pgas (r=0,704, P<0,001) and MIP (r=0,739, P<0,001), which supports the validity of the PCF as an overall parameter for cough.	High

Table 2 Overview of signs and symptoms for early detection of respiratory dysfunction and reduced cough ability.

Signs (anamnestic complaints)	Symptoms (findings of physical examination)
In supine position	In supine position
Orthopnoea	Orthopnoea
Nocturnal dyspnoea	Paradoxical breathing (inwards movement of the stomach during inspiration)
Other	Other
Reduced exercise tolerance	Ronchi
Dyspnoea on exertion	Superficial breathing
General fatigue	Weak voluntary cough
Retention of sputum	Use of accessory respiratory muscles
Inefficient cough	Respiratory alternans (alternating every few breaths between chest and stomach breathing)
Recurrent infections of the airways	

Key considerations

Reduced diaphragm function plays an important role in the development of complaints (18). The diaphragm alone accounts for 60% of the total inspiratory muscle strength. A significant decrease in recruitable lung volume gives an important indication of diaphragm weakness (19), and it is the only respiratory muscle that is not suppressed during sleep (20). The diaphragm is therefore the primary inspiratory muscle in supine position, and early diaphragmatic weakness is often noted first in this position and especially during sleep (21). Therefore especially orthopnoea (22), but also other of signs and symptoms of respiratory dysfunction (11), are complementary to respiratory function testing. Beside evidence for the superiority of the VC in supine position in detecting diaphragm weakness (16, 23), another advantage of supine testing is that in this position also the presence of orthopnoea can be examined as well as paradoxical breathing, in case of severe weakness of the diaphragm. The PDRG also pointed out that in some cases the delta VC (difference between sitting and supine VC) is of added value to detect diaphragm weakness. A delta VC of ≥20% (percent decay of supine VC relative to sitting VC) indicates diaphragm weakness in patients with neuromuscular diseases (24, 25). Therefore the PRDG pointed out that the VC in supine position is the first test of choice for the early detection of respiratory dysfunction due to diaphragm weakness. However, if orthopnoea is present, or another reason that prevents the supine position from being feasible, an alternative respiratory function test in sitting position is preferred. The PRDG proposed the MIP an SNIP as an alternative, but pointed out that unwanted variation, for example due to a learning effect must be taken into account (26, 27). SNIP is suggested as suitable for respiratory function testing in patients with bulbar dysfunction as is does not require a seal around a mouthpiece (21). Nevertheless, there are also indications that SNIP underestimates

respiratory function in patients with bulbar weakness due to upper airway collapse (28). In this case the PRDG proposed blood gas analysis as an alternative (12).

Key recommendations

Combining assessment of signs and symptoms **(table 2),** with assessment of the following respiratory function tests, is essential for early detection of reduced respiratory dysfunction and cough ability (**quick reference 1** with specific cut off points for the respiratory function tests):

- To detect reduced cough ability, the PCF is the first test of choice.
- For early detection of respiratory dysfunction, first measure sitting VC and assess signs and symptoms.
- Determine if orthopnoea or paradoxical breathing is present in supine position
- If orthopnoea is not present, measure the VC *supine position* for early detection of respiratory dysfunction in the form of diaphragmatic weakness.
- If it is not feasible to measure the VC in supine position, for example due to orthopnoea, the MIP or SNIP in sitting position should be measured.
- If respiratory function tests can no longer be carried out according to the protocol due to bulbar weakness, or if instructing the patient is problematic, perform blood gas analysis if respiratory dysfunction is suspected.

Non-pharmacological interventions to improve inefficient cough and respiratory muscle weakness.

Key evidence

Seven studies were included in the review. The included papers were graded as low quality evidence due to small sample sizes and imprecision. Six studies were assessed with high risk of bias.

Evidence from these are summarised in **table 3**. Three studies show a beneficial effect of air stacking, but there is insufficient evidence to demonstrate a difference in effectiveness on cough strength (PCF), morbidity and mortality between air stacking and the coughing machine (29-31). One study reported that a mechanical cough assist device has a greater effect on increasing cough strength (PCF) than air stacking, and that air stacking has a greater effect on the cough strength than the combination of spontaneous coughing with manual support (31). There are indications that cough strength is also increased in bulbar patients by air stacking (29-31). Two studies showed that inspiratory muscle training (IMT) has no effect on

Table 3. Identified studies for key recommendations of non-pharmacological interventions to support inefficient cough and to slow down respiratory decline

Author, year	Sample (n)	Study design	Study aim	Results and authors' conclusion	ROB score
Mustfa, 2003 (29)	47	RCT	To compare the effect of MAC (manual assisted cough) and MI-E (coughing machine) on cough effectiveness in bulbar and non-bulbar ALS patients.	MI-E increased the PCF 15% more than MAC in bulbar and in non-bulbar patients. The greatest improvements were in patients with the weakest coughs.	High
Senent, 2011 (31)	16	Quasi RCT	To longitudinally examine the effect of various manual and instrumental cough supporting techniques on the PCF was in ALS patients with a PCF of <270L / m.	Manual techniques: Unassisted cough without coaching Unassisted cough with coaching Cough with abdominal thrust Instrumental techniques: Air stacking with abdominal thrust Bi-level pressure with abdominal thrust IPAP of +30 cm H2O with abdominal thrust MI-E (40/-40 cm H2O) In bulbar and in non-bulbar patients, all the instrumental techniques were statistically better than the manual technique ($p < 0.0001$) on improving the PCF. Patients rated technique 5as the most comfortable method, technique 7 as the most effective method and technique 1 as the least. It appears to be beneficial to test an array of techniques in order to tailor the optimal cough improvement techniques to the individual. The existence of bulbar signs, even marked, should not dissuade clinicians from evaluating cough assistance techniques. Standard deviations of the effects were not reported.	High

Table 3. *(continued)*

Author, year	Sample (n)	Study design	Study aim	Results and authors' conclusion	ROB score
Rafiq, 2015 (30)	40	RCT	To examine the efficacy of the cough machine over breath stacking on the number and duration of respiratory infections, the chance of hospitalization during an airway infection, quality of life and survival.	The average per month decline in SVC was 0.94% in the breath-stacking group and 0.45% in the MI-E group (p 0.47). The PCF declined on average by 5.77 l/min/month in the breath stacking group and improved by 0.9 l/min/month in the MI-E group (p 0.43) Lack of statistically significant differences due to sub-optimal power and confounders precludes a definitive conclusion with respect to the relative efficacy of MI-E over breath stacking. However, the authors recommend the breath-stacking technique as a low-cost, first-line intervention for volume recruitment and cough augmentation in patients with ALS who meet the criteria for intervention with non-invasive ventilation. Mean change scores and standard deviations for the effects of breath-stacking and MI-E, on respiratory function were not reported.	High
Cheah 2009 (32)	19 (9/10)	RCT	to determine whether a 12-week inspiratory muscle training program attenuated the decline in respiratory function and inspiratory muscle strength in patients with ALS/MND.	Consistent, but non-significant trends for improvement across all respiratory parameters (FVC, VC, TLC, MIP and SNIP) were demonstrated in the experimental group that received IMT. Improvements in inspiratory muscle strength were observed in the experimental group as well as the control group over the training period of 12 weeks, although the effect was less marked in the control group. This magnitude of effect was more pronounced on MIP testing, which although non-significant, was 6.10 ± 6.93% greater in the experimental group than the control group (p = 0.39). Inspiratory muscle strength subsequently declined following 8 weeks after withdrawal after of the training device in both treatment arms (SNIP: 6.55 ± 1.88%; 95% CI 2.55-10.54; p <0.05; MIP: 4.53 ± 2.15%; 95% CI -0.03-9.08, p=0.05).	Low

Table 3. *(continued)*

Author, year	Sample (n)	Study design	Study aim	Results and authors' conclusion	ROB score
Pinto 2012 (33)	26 (13/13)	RCT	To evaluate the potential role of respiratory exercise by implementing specific inspiratory muscle training in a selected population of early-affected ALS	IMT was unable to slow down deterioration of inspiratory force in ALS patients in the early phase of the disease MIP 70 ± 28 versus 78 ± 26, p 0.45). No adverse effects of IMT were found on respiratory muscle strength or function.	High
Plowman 2016 (35)	25	Cohort	To evaluate feasibility and impact of expiratory muscle strength training (EMT) on respiratory and bulbar function in persons with ALS	A significant main effect was observed for the primary outcome variable, MEP P <0.03. Post-hoc testing revealed a significant increase in MEPs between baseline and post EMT (P = 0.01) No significant differences were noted for voluntary cough spirometry measures (PCF). 10 of 25 patients dropped out during the 5 week EMT intervention.	High
Plowman 2019 (34)	48 (24/24)	RCT	To determine the impact of an in-home expiratory muscle strength training (EMT) program on pulmonary, swallow, and cough function in individuals with ALS	The six weeks training intervention resulted in a significant improvement of the MEP (+ 25.5, CI 14.3 to 36.7 cmH2O in the intervention group and + 6.6, CI -3.4 to 16.5 cmH2O in the control group, p = 0.009). No significant improvements were found for voluntary cough spirometry and FVC. Group differences and change scores were not reported.	High

the deterioration of inspiratory muscle strength in people with ALS (32, 33). There is very low quality of evidence from 2 studies for effectiveness of expiratory muscle training (EMT) on the expiratory muscle power (MEP), however no effect has been found on the clinical parameters of coughing (PCF) (34, 35).

Key considerations

An inefficient cough in the context of a neuromuscular disorder is generally defined as a cough strength (PCF) of <270L/m (3, 31, 36). However, the PRDG emphasized that due to a lack of reliable normative values for the PCF, and well-founded cut-of points, assessment of signs and symptoms of inefficient cough is complementary to cough strength testing. Based on clinical experience, the PRDG identified various airway clearance techniques which can support an inefficient cough: Active Cycle of Breathing Techniques (ACBT), Positive Expiratory Pressure (PEP), spontaneous cough with abdominal compression, air stacking and the mechanical cough assist device (MI-E). As well as supporting the cough, airway clearance techniques can also be used to maintain or increase the compliance of the lung tissue and the chest wall. The PRDG pointed out that in general, it is preferable to personalize the treatment by choosing the most active and effective airway clearance technique for the individual patient.

*Key recommendations**

- Apply (a combination of) airway clearance techniques structurally if the cough strength is inefficient (PCF<270L/m), or earlier if there are signs and/or symptoms of reduced cough strength (**Table 1**), recurrent airway infections or rapid respiratory deterioration.
- Consider ACBT or PEP as the first-line treatment for inefficient cough strength. For patients with bulbar dysfunction, or in cases where ACBT/PEP is not effective, consider air stacking. Only if air stacking is not effective, or if a respiratory infection is present, consider – in consultation with a specialised pulmonologist – the use of a mechanical cough assist device. Abdominal/costal compression can be applied in combination with all airway clearance techniques, but should only be applied if there is a clear added value.
- Tailor the airway clearance technique to the individual and discuss options with the patient and his/her environment. Which technique is best tolerated and produces sufficient effect is a matter of trial and error.
- Do not routinely apply IMT and EMT. Consider IMT (20-40% of the MIP) in motivated patients in whom diaphragm weakness is not evident or only mild

(VC in supine position >80%). An evaluation should be carried out after 6 weeks of IMT taking into account the preference of the patient and the effect achieved on the MIP, in order to decide whether to continue the training.

* **Quick reference 2** provides an hierarchical overview of airway clearance techniques, and a list of contra-indications for air stacking.

Conclusions and future directions

Our study provides evidence-based clinical recommendations for the symptomatic management of early respiratory complaints, during the build-up of respiratory insufficiency in ALS. We consider the combination of signs and symptoms of respiratory weakness with relevant respiratory function tests as an adequate and implementable strategy that can contribute to the early detection of respiratory dysfunction and inefficient cough in ALS. In case of respiratory dysfunction, several supportive, easy to perform interventions are available before ventilatory support is needed. As is often the case in the supportive disease management field, the availability and quality of the included scientific evidence was generally low. Therefore the expertise from the patient representatives, the expert panel, and the PRDG was of important value for the formulation of the recommendations of this practical guideline.

We suggest that future research should further clarify which respiratory function test is most valid for the (early) detection of respiratory dysfunction, by means of broad, mutual comparison. This allows further refinement of relevant and timely supportive interventions during the build-up of respiratory insufficiency. Furthermore, in respiratory function testing, reference standards can be used to compare a patient's respiratory function to a healthy population of a similar age, body height, sex, and ethnicity (37). However, currently, high quality reference standards for the VC in supine position, MIP, SNIP and PCF are lacking. As a result, absolute cut-off values are often proposed, in which no distinction is made for morphological differences, therefore limiting tailor made clinical decision making. Another research topic is the optimal frequency and intensity of airway clearance techniques, to support inefficient cough.

Implementation of practise recommendations

The ultimate goal of this clinical guideline is to facilitate shared decision making by timely detection of respiratory dysfunction in ALS patients, therefore improving tailored care. To promote implementation of the guideline, online training modules

with were developed (free access at www.als-centrum.nl). In addition to the online training modules, physical courses with an emphasis on practical skills training are regularly offered.

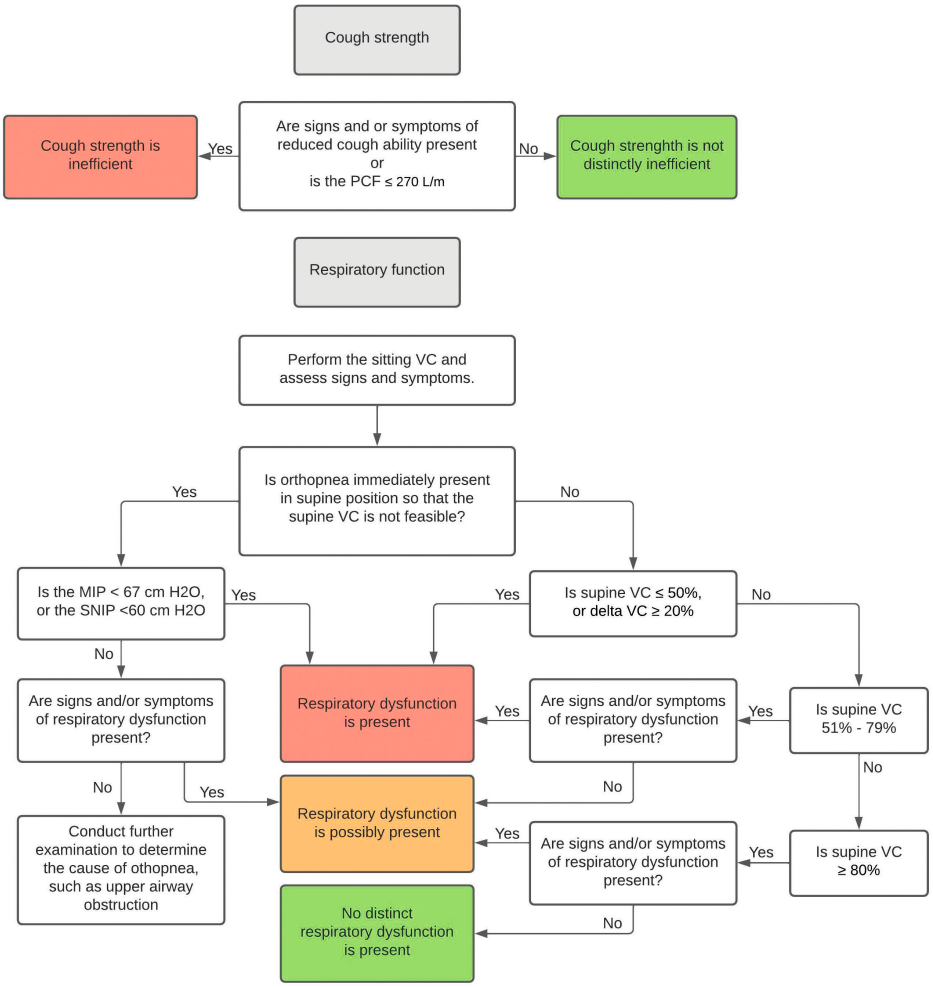
Acknowledgments:
We thank members of the PRDG, Professor M. de Visser MD, PhD (Department of Neurology, Amsterdam UMC), H. Grupstra, MD (Department of Rehabilitation, Amsterdam UMC) A. van Groenestijn, MD, PhD (Department of Rehabilitation, Amsterdam UMC) H. Cremers, (Department of Rehabilitation, Amsterdam UMC), S. Offeringa, MSc (Department of Rehabilitation, Amsterdam UMC). Members of the expert review panel J. Raaphorst, MD, PhD (Department of Neurology, Amsterdam UMC), J. Groothuis (Department of Rehabilitation, Radboud UMC Nijmegen), J. ten Broek-Pastoor (Department of Rehabilitation, Radboud UMC Nijmegen), P. Zanen, PhD (Division of heart and lungs, UMC Utrecht), M. Biever, PT (Basalt rehabilitation centre, The Haque), F. de Ruiter-ten Doeschate, PT ('t Roessingh rehabilitation centre, Enschede), V. van Dongen, PT (Department of Rehabilitation, Radboud UMC Nijmegen) F. Driehuis, PhD (Royal Dutch Society for Physiotherapy, (KNGF)). Patient representatives R. Kolijn (Dutch ALS patient association (ALS patients connected)), C. van der Meijden (Dutch ALS patient association (ALS patients connected)), A Horemans, PhD (Dutch patient organisation for neuromuscular diseases (Spierziekten Nederland)). None of these persons had a role in the preparation of this manuscript or the decision to submit this manuscript for publication.

References

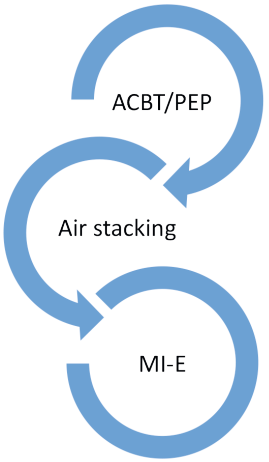
1. van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, et al. Amyotrophic lateral sclerosis. Lancet. 2017;390(10107):2084-98.
2. Bourke SC, Steer J. Practical respiratory management in amyotrophic lateral sclerosis: evidence, controversies and recent advances. Neurodegener Dis Manag. 2016;6(2):147-60.
3. Motor Neurone Disease: Assessment and Management. National Institute for Health and Care Excellence: Clinical Guidelines. 2016.
4. Miller RG, Jackson CE, Kasarskis EJ, England JD, Forshew D, Johnston W, et al. Practice parameter update: the care of the patient with amyotrophic lateral sclerosis: drug, nutritional, and respiratory therapies (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2009;73(15):1218-26.
5. Lechtzin N, Cudkowicz ME, de Carvalho M, Genge A, Hardiman O, Mitsumoto H, et al. Respiratory measures in amyotrophic lateral sclerosis. Amyotroph Lateral Scler Frontotemporal Degener. 2018;19(5-6):321-30.
6. Benditt JO, Boitano L. Respiratory treatment of amyotrophic lateral sclerosis. Phys Med Rehabil Clin N Am. 2008;19(3):559-72, x.
7. Hatcher-Martin JM, Busis NA, Cohen BH, Wolf RA, Jones EC, Anderson ER, et al. American Academy of Neurology Telehealth Position Statement. Neurology. 2021;97(7):334-9.
8. Norris SP, Likanje MN, Andrews JA. Amyotrophic lateral sclerosis: update on clinical management. Curr Opin Neurol. 2020;33(5):641-8.
9. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. J Clin Epidemiol. 2010;63(12):1308-11.
10. Fregonezi G, Araujo PR, Macedo TL, Dourado Junior ME, Resqueti VR, Andrade Ade F. Monitoring respiratory muscle strength assists in early diagnosis of respiratory dysfunction as opposed to the isolated use of pulmonary function evaluation in amyotrophic lateral sclerosis. Arq Neuropsiquiatr. 2013;71(3):146-52.
11. Jackson CE, Rosenfeld J, Moore DH, Bryan WW, Barohn RJ, Wrench M, et al. A preliminary evaluation of a prospective study of pulmonary function studies and symptoms of hypoventilation in ALS/MND patients. Journal of the neurological sciences. 2001;191(1-2):75-8.
12. Lyall RA, Donaldson N, Polkey MI, Leigh PN, Moxham J. Respiratory muscle strength and ventilatory failure in amyotrophic lateral sclerosis. Brain : a journal of neurology. 2001;124(Pt 10):2000-13.
13. Mendoza M, Gelinas DF, Moore DH, Miller RG. A comparison of maximal inspiratory pressure and forced vital capacity as potential criteria for initiating non-invasive ventilation in amyotrophic lateral sclerosis. Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases. 2007;8(2):106-11.
14. Tilanus TBM, Groothuis JT, TenBroek-Pastoor JMC, Feuth TB, Heijdra YF, Slenders JPL, et al. The predictive value of respiratory function tests for non-invasive ventilation in amyotrophic lateral sclerosis. Respir Res. 2017;18(1):144.
15. Varrato J, Siderowf A, Damiano P, Gregory S, Feinberg D, McCluskey L. Postural change of forced vital capacity predicts some respiratory symptoms in ALS. Neurology. 2001;57(2):357-9.
16. Lechtzin N, Wiener CM, Shade DM, Clawson L, Diette GB. Spirometry in the supine position improves the detection of diaphragmatic weakness in patients with amyotrophic lateral sclerosis. Chest. 2002;121(2):436-42.
17. Tzani P, Chiesa S, Aiello M, Scarascia A, Catellani C, Elia D, et al. The value of cough peak flow in the assessment of cough efficacy in neuromuscular patients. A cross sectional study. Eur J Phys Rehabil Med. 2014;50(4):427-32.

18. Pinto S, Carvalho M. Breathing new life into treatment advances for respiratory failure in amyotrophic lateral sclerosis patients. Neurodegener Dis Manag. 2014;4(1):83-102.
19. Mier-Jedrzejowicz A, Brophy C, Moxham J, Green M. Assessment of diaphragm weakness. Am Rev Respir Dis. 1988;137(4):877-83.
20. Braun AT, Caballero-Eraso C, Lechtzin N. Amyotrophic Lateral Sclerosis and the Respiratory System. Clin Chest Med. 2018;39(2):391-400.
21. Hardiman O. Management of respiratory symptoms in ALS. J Neurol. 2011;258(3):359-65.
22. Bourke SC, Bullock RE, Williams TL, Shaw PJ, Gibson GJ. Noninvasive ventilation in ALS: indications and effect on quality of life. Neurology. 2003;61(2):171-7.
23. Schmidt EP, Drachman DB, Wiener CM, Clawson L, Kimball R, Lechtzin N. Pulmonary predictors of survival in amyotrophic lateral sclerosis: use in clinical trial design. Muscle & nerve. 2006;33(1):127-32.
24. Fromageot C, Lofaso F, Annane D, Falaize L, Lejaille M, Clair B, et al. Supine fall in lung volumes in the assessment of diaphragmatic weakness in neuromuscular disorders. Arch Phys Med Rehabil. 2001;82(1):123-8.
25. Minami T, Manzoor K, McCool FD. Assessing Diaphragm Function in Chest Wall and Neuromuscular Diseases. Clin Chest Med. 2018;39(2):335-44.
26. Bauer M, Czell D, Hartmann S, Goldman B, Muller D, Weber M. Limitations of sniff nasal pressure as an outcome measurement in amyotrophic lateral sclerosis patients in a clinical trial. Respiration. 2012;84(4):306-11.
27. Hart N, Polkey MI, Sharshar T, Falaize L, Fauroux B, Raphael JC, et al. Limitations of sniff nasal pressure in patients with severe neuromuscular weakness. J Neurol Neurosurg Psychiatry. 2003;74(12):1685-7.
28. Chaudri MB, Liu C, Watson L, Jefferson D, Kinnear WJ. Sniff nasal inspiratory pressure as a marker of respiratory function in motor neuron disease. Eur Respir J. 2000;15(3):539-42.
29. Mustfa N, Aiello M, Lyall RA, Nikoletou D, Olivieri D, Leigh PN, et al. Cough augmentation in amyotrophic lateral sclerosis. Neurology. 2003;61(9):1285-7.
30. Rafiq MK, Bradburn M, Proctor AR, Billings CG, Bianchi S, McDermott CJ, et al. A preliminary randomized trial of the mechanical insufflator-exsufflator versus breath-stacking technique in patients with amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration. 2015;16(7-8):448-55.
31. Senent C, Golmard JL, Salachas F, Chiner E, Morelot-Panzini C, Meninger V, et al. A comparison of assisted cough techniques in stable patients with severe respiratory insufficiency due to amyotrophic lateral sclerosis. Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases. 2011;12(1):26-32.
32. Cheah BC, Boland RA, Brodaty NE, Zoing MC, Jeffery SE, McKenzie DK, et al. INSPIRATIonAL--INSPIRAtory muscle training in amyotrophic lateral sclerosis. Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases. 2009;10(5-6):384-92.
33. Pinto S, Swash M, De Carvalho M. Respiratory exercise in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis. 2012;13(1):33-43.
34. Plowman EK, Tabor-Gray L, Rosado KM, Vasilopoulos T, Robison R, Chapin JL, et al. Impact of expiratory strength training in amyotrophic lateral sclerosis: Results of a randomized, sham-controlled trial. Muscle & nerve. 2019;59(1):40-6.
35. Plowman EK, Watts SA, Tabor L, Robison R, Gaziano J, Domer AS, et al. Impact of expiratory strength training in amyotrophic lateral sclerosis. Muscle & nerve. 2016;54(1):48-53.
36. Bach JR, Ishikawa Y, Kim H. Prevention of pulmonary morbidity for patients with Duchenne muscular dystrophy. Chest. 1997;112(4):1024-8.
37. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012;40(6):1324-43.

Quick reference 1. Early detection of reduced cough ability and respiratory dysfunction.



Quick reference 2. Non-pharmacological interventions for the support of inefficient cough.



Phases of Active Cycle of Breathing Techniques. Alternate phases, and adjust the order to the individual.

- Breathing control: relaxed breathing on tidal volume. Use breathing control to recover between phases.
- Deep Breathing Exercises: Breathe in slowly as deep as possible. Hold the breath for 3 seconds and exhale slowly. Repeat 3-5 times.
- Forced expiration technique (huffing). The goal is to move sputum from the upper airways to the throat. Huffing is a forceful and -if possible- prolonged exhalation, with open mouth and open throat.
- If the forced expiration technique is not productive, try coughing in combination with abdominal/costal compression.
- Breathing through a PEP mask creates resistance during expiration, and can be useful as an alternative/addition to ACBT

Air stacking has different options for application. Try out en tailor to the individual.

- Mouthpiece, with or without nose clip, or full face mask.
- Operate the ambu bag with the hands, just below the armpit, between the knees, or by a carer (after instruction).
- In case of inability to close the glottis, consider a one-way-valve, which blocks exhalation.
- In case of sputum retention, try coughing after acquiring maximal inspiratory capacity, possibly in combination with abdominal/costal compression

The mechanical cough assist device (MI-E) can be used for people that are no longer able to actively cough up sputum. It's use requires specialist guidance.

! Air stacking can cause complications, and requires skill. Patients should give explicit consent before application.

! Contra indication for air stacking are

- COPD GOLD 3 or 4, and other diseases that affect the elasticity of lung tissue
- Uncooperative patient (reduced executive functioning, no disease insight, low consciousness)
- Acute respiratory insufficiency
- Bilateral vocal cord paralysis
- Cuffed tracheal cannula
- Recent lobectomy
- Recent pneumothorax
- Increased intracranial pressure, also ventricular drains

Supplement I Review protocols and search strings

Review question 1	Which respiratory function tests are indicative for early respiratory dysfunction and reduced cough ability.
Objectives	To identify respiratory function parameters that are sensitive to assess respiratory dysfunction (respiratory insufficiency, nocturnal hypoventilation and cough ineffectiveness) To identify the optimum frequency of assessment required to monitor pulmonary function in people with ALS
Review population	Adults (aged 18 and over) with ALS
Outcomes	**Critical:** respiratory insufficiency, nocturnal hypoventilation cough effectiveness **Non-critical:** FVC, SVC, FEV1, PCF, PEF, MIP, MEP, SNIP, pCO2, pO2, SpO2, pH, bloodgas
Search strategy	**Databases:** Medline **Date limit:** No date limit applied **Date of search:** August 2nd 2018 **Language:** English and Dutch **Study designs:** systematic reviews, meta-analysis, cross-sectional study, prospective cohort study, Retrospective cohort study Minimum sample size: 20
Exclusion criteria	Letter, editorial, news, experimental historical article, anecdotes as topic, comment, case report, randomized controlled trial, experimental animal studies

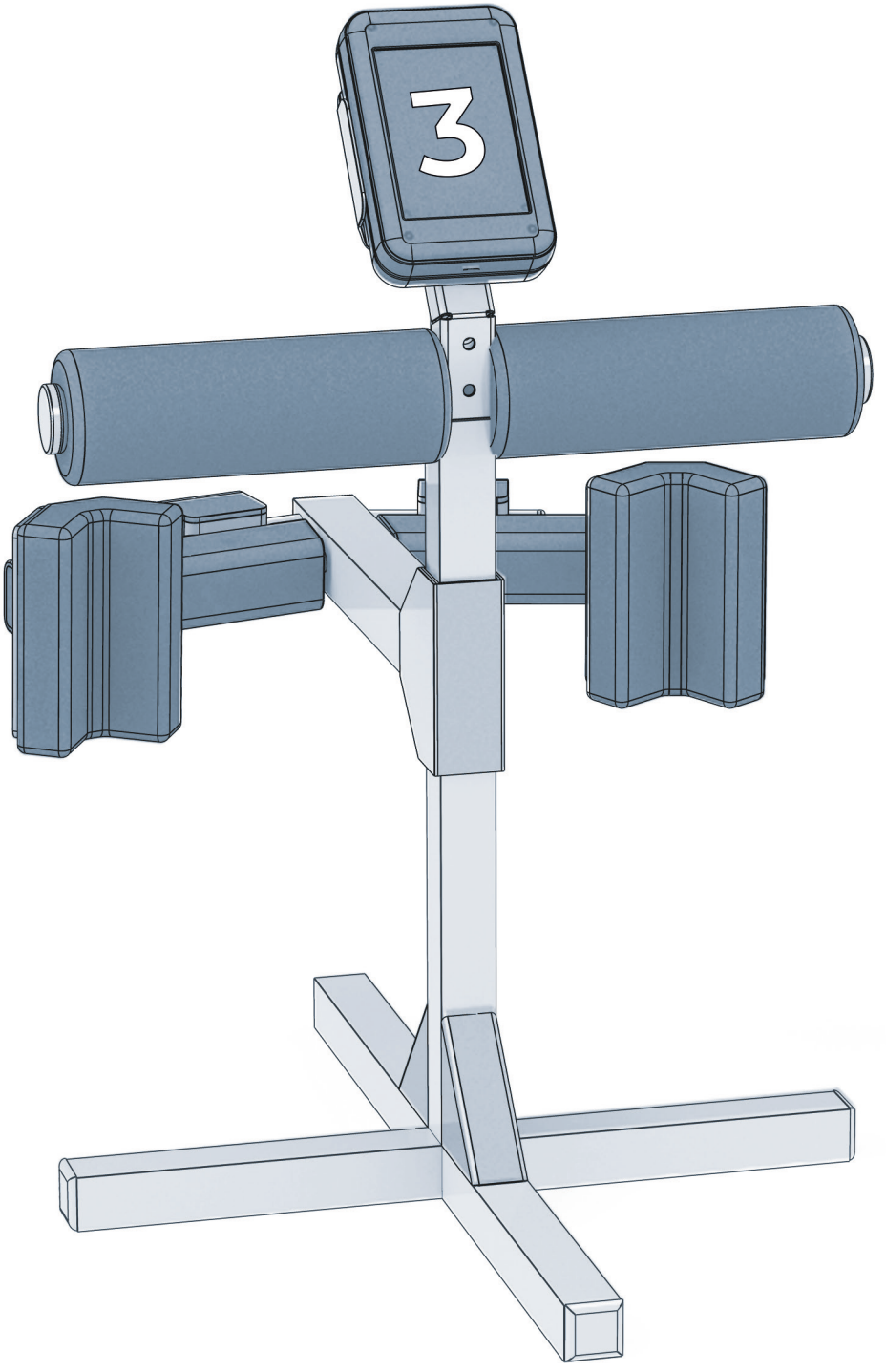
Review question 2	Which non-pharmacological interventions can improve inefficient cough and respiratory muscle weakness
Objective	To evaluate the effectiveness of cough augmentation techniques and respiratory muscle training for patients with ALS
Criteria	**Population:** Patients with ALS **Interventions:** Basic cough augmentation techniques: Active cycle of breathing techniques (ACBT) for example TEE, breathing control, huffing Postural drainage and manual techniques (shaking/percussion/vibration), GAP, positioning Manual cough assisted coughing technique (quad coughing, assisted coughing) Maximal insufflation capacity techniques (MIC) e.g. breath stacking(unassisted)/thoracic range of movement exercises, GPB Respiratory muscle training (IMT, EMT) Devices (maximal insufflation capacity techniques/lung inflation capacity techniques): Mechanical cough assist device (mechanical insufflation-exsufflation) Intrapulmonary percussive ventilation Lung volume recruitment techniques (for example LVR bags, NIV device to increase the inspiratory phase of cough to increase cough capacity) Suction pump **Comparisons:** Compared with each other, or with nothing **Outcomes:** Critical: Survival, Health-related quality of life, Patient/carer reported outcomes (ability to cough, ability to clear secretions, concordance, breathlessness, fatigue) Important: Change in peak cough flow, Reduction of chest infection (community- or hospital-acquired pneumonia and aspiration), Hospital admissions (and unplanned admissions)
Search strategy	**Databases:** Embase **Date limit:** No date limit applied **Date of search:** August 2nd 2018 **Language:** English and Dutch language **Study designs:** RCTs or systematic reviews of RCTs; if no RCTs are retrieved, we will search for cohort studies with a sample size >20.
Inclusion criteria	**Aggregated studies:** systematic reviews, meta-analysis **Controlled trials:** RCT, CCT

Search string Medline

#	Searches Medline	Results
1	motor neuron disease/	4105
2	amyotrophic lateral sclerosis/	16344
3	bulbar palsy, progressive/	797
4	exp *motor neuron/	23085
5	(motor neuron* or moto neuron* or motoneuron* or motorneuron* or motor-neuron* or motor-neuron*).ti,ab,kf.	41681
6	((primary or amyotrophic) adj lateral scleros*).ti,ab,kf.	19817
7	(progressive adj (muscular atroph* or bulbar pals*)).ti,ab,kf.	572
8	(pseudopolyneur* or pseudo-polyneur* or psuedo polyneur*).ti,ab,kf.	38
9	((pseudobulbar or pseudo-bulbar or pseudo bulbar) adj pals*).ti,ab,kf.	372
10	((bulbar or respirat* or limb) adj onset*).ti,ab,kf.	451
11	lou gehrig*.ti,ab,kf.	202
12	((anterior or ventral) adj (horn or column) adj3 (disease* or disorder*)).ti,ab,kf.	172
13	(flail* adj (arm* or leg*) adj (syndrome* or disorder*)).ti,ab,kf.	32
14	(guam adj (disease* or disorder* or syndrome*)).ti,ab,kf.	8
15	monomelic amyotroph*.ti,ab,kf.	97
16	frontotemporal dementia/	1907
17	((frontotemporal or fronto temporal or fronto-temporal) adj dement*).ti,ab,kf.	5589
18	or/1-17 [MND]	72067
19	transcutaneous electric nerve stimulation/	4112
20	((function* or neuromuscul* or peripheral* or transcutan* or electric*) adj4 stimulat*).ti,ab,kf.	88223
21	TENS.ti,ab,kf.	12800
22	ultrasonography/	169683
23	(sonograph* or ultrasound* or ultrason*).ti,ab,kf.	356442
24	exp physical therapy modalities/	135294
25	exp exercise/	162613
26	(physiotherap* or exercis* or stretch* or resist* or position*).ti,ab,kf.	1718347
27	(physical adj2 therap*).ti,ab,kf.	22193
28	muscle cramp/ or muscle rigidity/ or muscle spasticity/ or muscle weakness/	19696
29	((muscle* or muscular) adj2 (cramp* or rigid* or spast* or weak* or tight* or stiff* or twitch* or spasm*)).ti,ab,kf.	27206
30	(fasciculat* or contract*).ti,ab,kf.	264560
31	transcranial magnetic stimulation/	9176
32	((transcran* or intramusc*) adj4 stimulat*).ti,ab,kf.	15747
33	or/19-32 [Muscle problems]	2602907
34	18 and 33	13941
35	meta-analysis/	84725

Search string Medline

#	Searches Medline	Results
36	meta-analysis as topic/	16313
37	(meta analy* or metanaly* or metaanaly*).ti,ab.	121577
38	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.	154755
39	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.	36367
40	(search strategy or search criteria or systematic search or study selection or data extraction).ab.	41396
41	(search* adj4 literature).ab.	47893
42	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.	159891
43	cochrane.jw.	13808
44	or/35-43 [SR's]	345865
45	randomized controlled trial.pt.	475559
46	controlled clinical trial.pt.	96070
47	randomi#ed.ab.	496292
48	placebo.ab.	194458
49	randomly.ab.	288207
50	clinical trials as topic.sh.	189098
51	trial.ti.	186891
52	or/45-51 [RCT]	1195381
53	epidemiologic studies/	7771
54	exp case control studies/	900957
55	exp cohort studies/	1734121
56	cross-sectional studies/	256049
57	case control.ti,ab,kf.	107637
58	(cohort adj (study or studies or analys*)).ti,ab.	151876
59	((follow up or observational or uncontrolled or non randomi#ed or nonrandomi#ed or epidemiologic*) adj (study or studies)).ti,ab.	203420
60	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort*)).ti,ab.	1143511
61	or/53-60 [observational studies]	2648876
62	or/44,52,61	3759548
63	34 and 62	1450


3

Chapter 3

PATTERN OF MUSCLE STRENGTH IMPROVEMENT AFTER INTRAVENOUS IMMUNOGLOBULIN THERAPY IN MULTIFOCAL MOTOR NEUROPATHY

Jaap Bakers
Ruben van Eijk
Leonard van den Berg
Johanna Visser-Meily
Anita Beelen

Muscle & Nerve 2021

Abstract

Introduction: In multifocal motor neuropathy (MMN), knowledge about the pattern of treatment response in a wide spectrum of muscle groups, distal as well as proximal, after intravenous immunoglobulin (IVIg) initiation is lacking.

Methods: Hand-held dynamometry data of 11 upper and lower limb muscles, from 47 patients with MMN was reviewed. Linear mixed models were used to determine the treatment response after IVIg initiation and its relationship with initial muscle weakness.

Results: All muscle groups showed a positive treatment response after IVIg initiation. Changes in standard deviation (SD) scores ranged from +0.1 to +0.95. A strong association between weakness at baseline and the magnitude of the treatment response was found.

Discussion: Improved muscle strength in response to IVIg appears not only in distal, but to a similar degree also in proximal muscle groups in MMN, with the largest response in muscle groups that show the greatest initial weakness.

Introduction

Multifocal motor neuropathy (MMN) is a rare motor neuropathy, characterized by progressive, asymmetric and predominantly distal arm and leg weakness.[1] Intravenous immunoglobulin (IVIg) is the first-line treatment in MMN and its efficacy in improving muscle strength has been confirmed repeatedly.[2-6]

Muscle involvement and treatment response are often based on muscle strength assessments. In MMN, the method most commonly used to quantify muscle strength is hand-held dynamometry (HHD).[2-4, 6] However, trials that quantified muscle strength using HHD evaluated preselected distal muscle groups with preserved contraction, or calculated sum scores. Therefore, knowledge about the pattern of treatment response in individual muscle groups, distal as well as proximal, after IVIg initiation is not supported by observational data. [7, 8]

Insight into the course of muscle strength in both distal and proximal muscle groups after IVIg initiation could contribute to optimization of therapeutic strategies.[9] This study aims to determine the pattern of treatment response at a group level in a wide spectrum of individual muscle groups after IVIg initiation in treatment-naive MMN patients, and to explore the relationship between initial muscle weakness and treatment benefit.

Methods

Study design

In this cohort study, we reviewed data from the electronic medical records of all consecutive treatment-naïve patients, diagnosed with probable or definite MMN according the diagnostic criteria of the EFNS and PNS,[8] who received their first treatment with IVIg (Gammagard; Hyland Baxter, Glendale, CA), and who underwent muscle strength assessments in the University Medical Center Utrecht (The Netherlands) between 2012 and 2018. Subjects were excluded if they had been treated with other immunosuppressive drugs in the 6 months before initial measurements or had comorbidities that had a direct effect on muscle strength.

The Medical Ethics Committee of the UMCU (protocol number 17-832) approved the study protocol, and concluded that no informed consent was required, provided that the data was anonymized before analysis.

Procedures

Testing procedure

During each objective muscle strength evaluation using HHD, eleven muscle groups were tested bilaterally. Hand-grip strength was tested with a hydraulic hand dynamometer (Jamar, Sammons & Preston, Bolingbrook, IL, USA). Pinch- and key-grip strength were tested with a hydraulic pinch gauge (Baseline, Fabrication Enterprises, White Plains NY, USA). The other eight muscle groups (wrist extension, elbow flexion and extension, shoulder abduction, ankle dorsiflexion, knee extension, hip abduction and flexion) were manually tested with a hand-held dynamometer (MicroFET2, Hogan Health Industries, Salt Lake City UT, USA) using the break method. With this technique, the examiner pushes the dynamometer against the subject's limb until the subject's maximal effort is overcome, and the joint gives way. To reduce the risk of measurement errors, objective muscle strength evaluations were performed according to standard procedures described elsewhere, including test positions, placement of the HHD, duration of contraction, and verbal encouragements. [10, 11] All muscle strength evaluations with HHD were conducted by the same physical therapist (J.N.E.B.).

Each measurement was repeated twice. If the value of the second measurement deviated by more than 10% from the first, the measurement was repeated until the difference between two individual scores was less than 10%. In this case, generally, a third or fourth measurement was sufficient. Of these two scores, the higher value was noted. The entire testing protocol took approximately 40 minutes per patient.

Muscle strength assessments of the second treatment were not available for all patients. Between 2012 and 2015, according to the local protocol, strength evaluation using HHD was done only after the first IVIg treatment. After 2015, muscle strength using HDD was evaluated both after first and second treatment. Pre-treatment measurements with HHD were carried out within one week prior to or during the first 3 days of hospitalization for IVIg treatment. According to the local protocol, the follow-up (post-treatment) measurements were obtained between 21 and 35 days after the first IVIg treatment course, and (if performed) between 21 and 35 days after the second IVIg treatment course . Datasheets with muscle strength values were added to the electronic patient file.

Data collection

All data were collected prospectively from the time of diagnosis. Demographic data were retrieved, and objective muscle strength evaluations were collected at the first and, if available, second follow-up. After data collection (by J.N.E.B), the datasheet

was anonymized by the data manager of the research institute of the UMCU). This was then provided to the researchers for statistical analysis.

Analysis

Baseline data were summarized by calculating the mean and the standard deviation (SD) for normally distributed data, and the median for non-normally distributed data. Muscle strength data were standardized by sex- and age.[12,13] The standardized scores, or *z*-scores, indicate by how much the patient's values deviate (measured in number of standard deviations) from the average muscle strength of a control population of similar age and sex. To illustrate, a standardized score of -2 indicates that the patient is 2 standard deviations below the expected muscle strength for his or her age.

Due to the repeated measurements per patient, we used a linear mixed model to estimate the mean change in standardized muscle strength from baseline at the first and second follow-up visits. The measurement prior to IVIg initiation was taken as baseline and incorporated in the model as a fixed covariate, in addition to follow-up visit as factor (i.e. visit 1 or 2). Each model was fitted per muscle group and incorporated a random intercept for subject number. In order to estimate the average, pooled effect, all muscle groups were combined into a single model, accounting for clustering within muscle groups and patients by specifying two random intercepts for each clustering variable. To compare the average treatment response between distal and proximal muscle groups, muscle groups were clustered accordingly and an additional indicator variable was added to the model as a fixed factor.

Results

Patient characteristics

Data were available for 47 MMN patients who fulfilled the inclusion criteria, and comprised a total of 2,574 isometric muscle strength measurements. Their patient characteristics are summarized in **Table 1**.

Treatment response of individual muscle groups

Figure 1 shows the average treatment response at a group level of individual muscle groups over time, when compared with baseline measurements (muscle strength prior to IVIg initiation). The change from baseline (CFB), expressed as number of standard deviations, indicated a statistically significant treatment response in most muscle groups at the first follow-up, and in all muscle groups at the second follow-up (see **supplemental Table 1**).

Table 1 Patient characteristics

Total number of included patients	47	
Number of patients with second follow-up	20	
Age, years	52	(11)
Male/female	40/7	
Disease duration, months	25	(2-161)
Treatment delay, months	1	(0-52)

Abbreviations: Age, age at diagnosis (SD).
Disease duration = time from first weakness to start of treatment (median and range).
Treatment delay = time from diagnosis to start of treatment with IVIg (median and range).

The average, pooled effect over all muscle groups was 0.56 (95% CI 0.37 – 0.75, $p < 0.001$). Changes in SD-scores ranged from 0.42 to 0.89 in the distal muscle groups (i.e. pinch, thumb, hand, wrist and ankle), and 0.1 to 0.95 in the proximal muscle groups (i.e. elbow, shoulder, knee and hip). Pairwise comparison between distal and proximal muscle strength groups showed no difference in muscle strength gain ($p = 0.77$). Compared to baseline, this treatment response was significant for both the first and the second follow-up, except for the first follow-up of elbow extension, knee extension, and hip abduction. In the upper extremity, elbow flexion and wrist extension showed the largest change from baseline scores. Hip flexion and ankle dorsiflexion showed the largest change in the lower extremity.

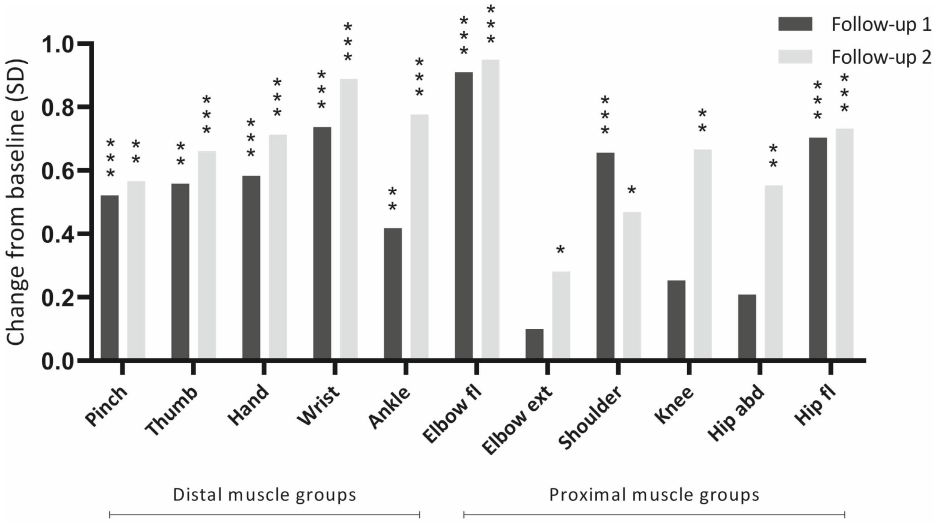
The treatment response increased during the second follow-up in almost all individual muscle groups, except for shoulder abduction. Pairwise comparison showed a significant additional increase in muscle strength between first and second follow-up for ankle flexion, knee extension and hip abduction (all $p < 0.05$).

Relationship between initial muscle strength and treatment response

Table 2 shows the average CFB for each individual muscle group compared to its baseline muscle strength. Despite variability between individual patients, most analysed muscle groups showed a strong association between initial weakness at baseline and the magnitude of the treatment response, with the largest CFB in the weakest muscle groups. Regression coefficients ranged from -0.06 (SE 0.04) to -0.42 (SE 0.08), and were significant in all muscle groups, except for hand grip. The coefficient indicates that for each standard deviation loss at baseline, patients gained up to an additional 0.42 standard deviation after IVIg initiation. Especially in the more proximal muscle groups, i.e. the shoulder (-0.33), hip (-0.40) and knee (-0.42), the magnitude of the treatment response depended more strongly on muscle strength

at baseline as compared to the more distal muscle groups like hand grip (-0.06) or ankle dorsiflexion (-0.17). **Supplemental Figure 1** provides additional supporting data.

Figure 1 Treatment response during follow-up visits 1 and 2



Mean improvement from baseline in number of standard deviations. Muscle strength measurements per muscle group were standardized according to sex and age. Abbreviations: SD = standard deviation, fl = flexion, ext = extension, abd = abduction. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2 Relationship between initial weakness at baseline and treatment response after IVIg initiation

	Slope (SE)	p value
Pinch	-0.22 (0.03)	<0.001
Thumb	-0.17 (0.03)	<0.001
Hand	-0.06 (0.04)	0.158
Wrist	-0.14 (0.04)	<0.001
Elbow fl	-0.24 (0.05)	<0.001
Elbow ext	-0.20 (0.05)	<0.001
Shoulder	-0.33 (0.08)	<0.001
Ankle	-0.17 (0.05)	0.001
Knee ext	-0.42 (0.08)	<0.001
Hip abd	-0.40 (0.08)	<0.001
Hip fl	-0.28 (0.07)	<0.001

Abbreviations: fl = flexion, ext = extension, abd = abduction. See also supplemental Figure 1 for a graphical representation of the relationship initial muscle strength and treatment response.

Discussion

The current study showed: (1) an overall positive treatment response in both distal and proximal muscle groups after IVIg initiation, and (2), an increase in treatment response as initial muscle strength (prior to IVIg initiation) decreased.

Despite variation on a group level, proximal muscle groups responded at a similar level as distal muscle groups during IVIg initiation. The similarity in treatment response between proximal and distal muscle groups was also noted in by previous research.[14]

When the first and second follow-up with IVIg were compared, the second follow-up caused an overall, additional treatment response in all muscle groups. However, within the wide range of tested muscle groups, this additional effect was only significant ($p < 0.05$) in muscles around the ankle, knee and hip; muscle groups of the lower extremity. Since symptoms in the lower extremity initially occur in only 34% of MMN patients, [1] this finding is surprising, and might suggest that muscle groups in the lower extremity responded more slowly to IVIg initiation than muscle groups in the upper extremity. This delay in treatment response may have led to an underestimation of the effectiveness of IVIg in the lower extremity.

The positive treatment response in more proximal muscle groups of arms and legs, may not have been detected previously due the use of the Medical Research Council (MRC) scale both in the clinical and experimental setting. The use of the MRC scale for the quantification of muscle strength is controversial because of its subjectivity.[15] Moreover, its insensitivity for detecting muscle weakness results in a risk of classifying muscle groups that are actually weakened as MRC 5.[16, 17] This insensitivity may cause structural underestimation of muscle weakness, in particular in large muscle groups that initially exhibited mild muscle weakness. Although distal arm and/or leg weakness is distinct in MMN, it is possible that MMN also causes a mild loss of strength in the proximal muscles relative to the level before disease onset. This possibly more widespread loss of muscle function could explain the fact that MMN patients report that, in addition to primary loss of manual dexterity, they also have fatigue and loss of walking ability.[18]

Finally, the current study found that muscle groups which were very weak (<-4 SD) prior to IVIg initiation still benefitted from IVIg. In fact, the weaker the initial muscle strength, the greater the treatment response in SD. This finding suggests that severe weakness in the early phase of MMN may be related to reversible damage, such as

focal demyelination, or inflammatory processes involving of the nodes of Ranvier, rather than irreversible axonal damage.[19] However, it is important to note that this finding was the result of a group analysis, and that on an individual level there were also patients who experienced prolonged weakness or paralysis in some muscle groups (also see **supplemental Figure 1**).

Although objective muscle strength measurements using HHD can provide a more reliable assessment of muscle strength than MRC scores, they can also be prone to measurement errors. The reliability of HHD depends on the technique and strength of the examiner. [20, 21] A ceiling effect can occur when it is difficult for an examiner to provide sufficient resistance at high forces (commonly around 200 to 300 Newton) in strong muscle groups such as the quadriceps. [22] In this study, the impact of these errors was reduced as much as possible by using a standardized protocol applied by the same experienced evaluator. Since the treatment response increased consistently during consecutive measurements, a learning effect may also have had an impact on the results. To minimise this potential learning effect in the current study, patients were familiarised with the test. In addition, the results of this study show that the treatment response continued to increase during the second follow-up, making a significant influence of a learning effect less likely. The difference in the magnitude of the treatment response between elbow flexion and elbow extension may have been caused by the normative values that were used, although a true difference in treatment responsiveness cannot be excluded. Furthermore it cannot be ruled out that part of the initial muscle weakness in the proximal muscle groups in arms and legs can be explained by deconditioning. Also the delayed response in proximal lower extremity muscles could be explained by better physical fitness in patients who are more active when mobility is improved due to stronger distal muscles. Placebo effect could also have played a role, and though rare, some patients probably have had a conduction block in proximal muscles. Due to the relatively short follow-up period of our study, it is not possible to accurately discriminate between true or placebo effects. To gain more insight in the mechanisms that underlie the treatment response, a longer follow-up study with repeated electrodiagnostic evaluation is required.

The results of this study revealed an overall positive treatment response after IVIg initiation, demonstrated in proximal as well as distal muscle groups. In addition, weaker muscle groups benefitted most from IVIg treatment, even if there initially was complete paralysis at baseline.

References

1. Cats EA, van der Pol WL, Piepers S, Franssen H, Jacobs BC, van den Berg-Vos RM, et al. Correlates of outcome and response to IVIg in 88 patients with multifocal motor neuropathy. Neurology. 2010;75:818-825.
2. Azulay JP, Blin O, Pouget J, Boucraut J, Bille-Turc F, Carles G, et al. Intravenous immunoglobulin treatment in patients with motor neuron syndromes associated with anti-GM1 antibodies: a double-blind, placebo-controlled study. Neurology. 1994;44:429-432.
3. Federico P, Zochodne DW, Hahn AF, Brown WF, Feasby TE. Multifocal motor neuropathy improved by IVIg: randomized, double-blind, placebo-controlled study. Neurology. 2000;55:1256-1262.
4. Hahn AF, Beydoun SR, Lawson V, Team IiMS, Oh M, Empson VG, et al. A controlled trial of intravenous immunoglobulin in multifocal motor neuropathy. J Peripher Nerv Syst. 2013;18:321-330.
5. Leger JM, Chassande B, Musset L, Meininger V, Bouche P, Baumann N. Intravenous immunoglobulin therapy in multifocal motor neuropathy: a double-blind, placebo-controlled study. Brain : a journal of neurology. 2001;124:145-153.
6. Van den Berg LH, Kerkhoff H, Oey PL, Franssen H, Mollee I, Vermeulen M, et al. Treatment of multifocal motor neuropathy with high dose intravenous immunoglobulins: a double blind, placebo controlled study. Journal of neurology, neurosurgery, and psychiatry. 1995;59:248-252.
7. Leger JM, Guimaraes-Costa R, Iancu Ferfoglia R. The pathogenesis of multifocal motor neuropathy and an update on current management options. Ther Adv Neurol Disord. 2015;8:109-122.
8. van Schaik IN, Leger JM, Nobile-Orazio E, Cornblath DR, Hadden RDM, Koski CL, et al. European Federation of Neurological Societies/Peripheral Nerve Society Guideline on management of multifocal motor neuropathy. Report of a Joint Task Force of the European Federation of Neurological Societies and the Peripheral Nerve Society - first revision. Journal of the Peripheral Nervous System. 2010;15:295-301.
9. Kumar A, Patwa HS, Nowak RJ. Immunoglobulin therapy in the treatment of multifocal motor neuropathy. J Neurol Sci. 2017;375:190-197.
10. Andrews AW, Thomas MW, Bohannon RW. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. Phys Ther. 1996;76:248-259.
11. Bohannon RW. Comparability of force measurements obtained with different strain gauge hand-held dynamometers. J Orthop Sports Phys Ther. 1993;18:564-567.
12. Bohannon RW. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. Arch Phys Med Rehabil. 1997;78:26-32.
13. Mathiowetz V, Kashman N, Volland G, Weber K, Dowe M, Rogers S. Grip and pinch strength: normative data for adults. Arch Phys Med Rehabil. 1985;66:69-74.
14. Harbo T, Andersen H, Hess A, Hansen K, Sindrup SH, Jakobsen J. Subcutaneous versus intravenous immunoglobulin in multifocal motor neuropathy: a randomized, single-blinded cross-over trial. Eur J Neurol. 2009;16:631-638.
15. Knepler C, Bohannon RW. Subjectivity of forces associated with manual-muscle test grades of 3+, 4-, and 4. Percept Mot Skills. 1998;87:1123-1128.
16. Beasley WC. Influence of method on estimates of normal knee extensor force among normal and postpolio children. Phys Ther Rev. 1956;36:21-41.
17. Bohannon RW. Manual muscle test scores and dynamometer test scores of knee extension strength. Arch Phys Med Rehabil. 1986;67:390-392.
18. Erdmann PG, Lindeman E, Cats EA, van den Berg LH. Functioning of patients with multifocal motor neuropathy. J Peripher Nerv Syst. 2010;15:113-119.

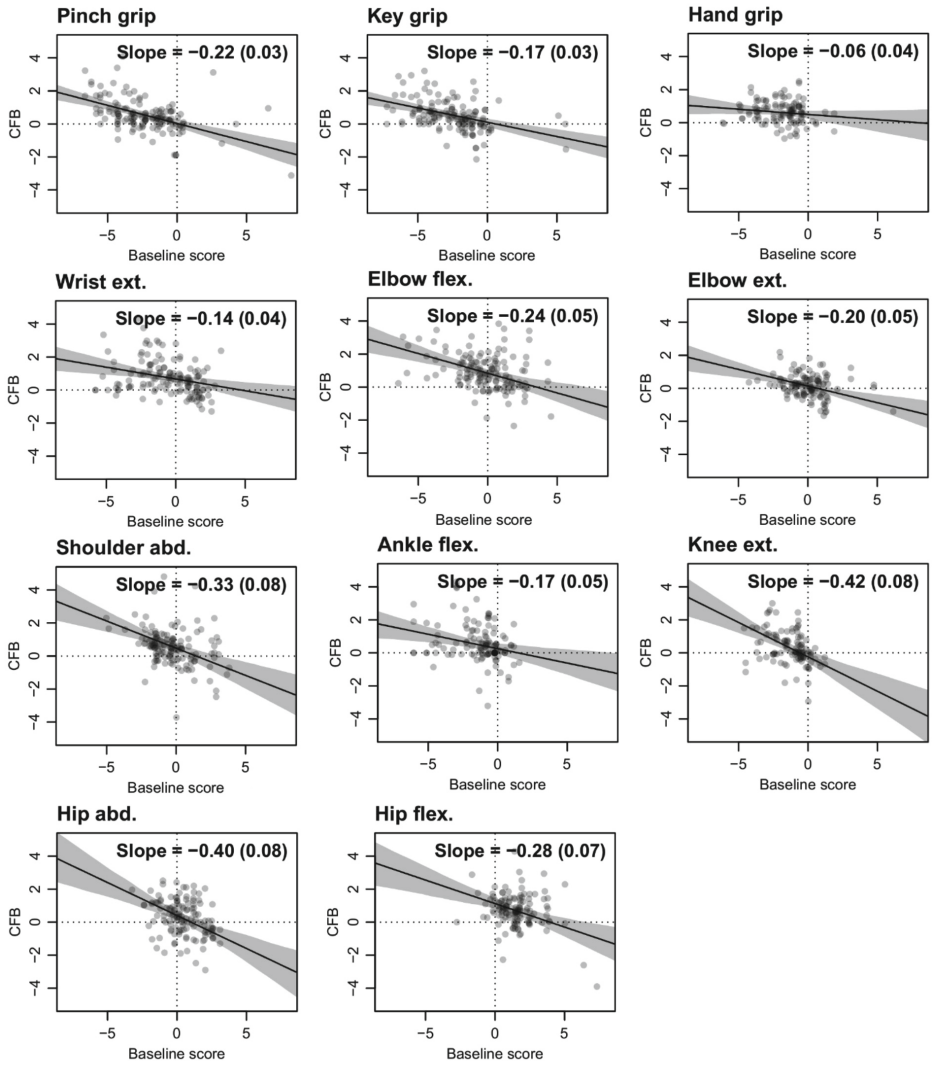
19. Van Asseldonk JT, Van den Berg LH, Kalmijn S, Van den Berg-Vos RM, Polman CH, Wokke JH, et al. Axon loss is an important determinant of weakness in multifocal motor neuropathy. Journal of neurology, neurosurgery, and psychiatry. 2006;77:743-747.

20. Kolber MJ, Cleland JA. Strength testing using hand-held dynamometry. Physical Therapy Reviews. 2005;10:99-112.

21. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held Dynamometry Correlation With the Gold Standard Isokinetic Dynamometry: A Systematic Review. Pm&R. 2011;3:472-479.

22. Visser J, Mans E, de Visser M, van den Berg-Vos RM, Franssen H, de Jong JM, et al. Comparison of maximal voluntary isometric contraction and hand-held dynamometry in measuring muscle strength of patients with progressive lower motor neuron syndrome. Neuromuscul Disord. 2003;13:744-750.

Supplemental Figure 1

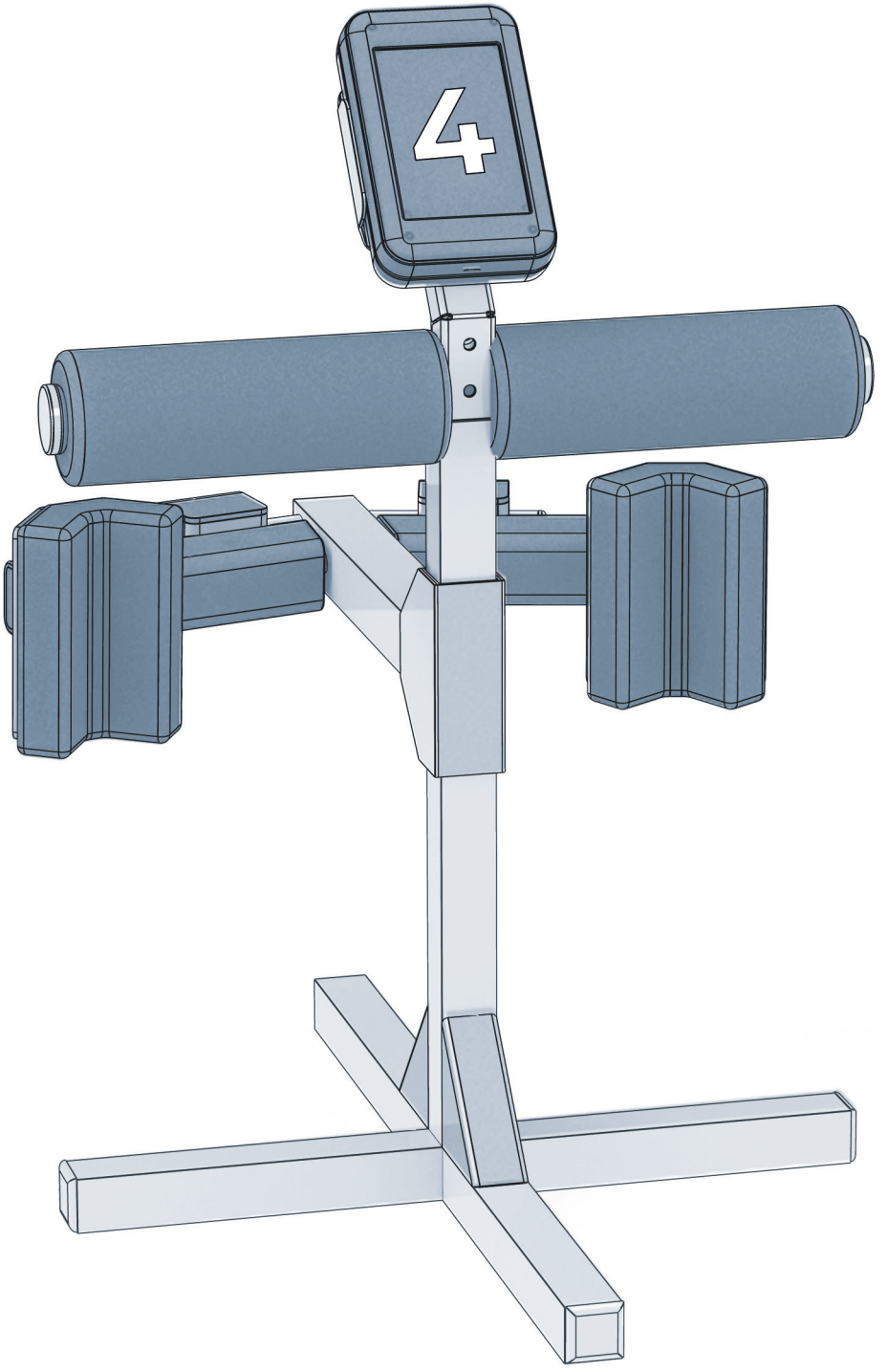


Graphical representation of the relationship between initial weakness at baseline and treatment response (CFB) after IVIg initiation. The raw CFB data for each muscle group is plotted on the y-axis, against the baseline score on the x-axis. For each muscle group, 4 quadrants can be distinguished by means of the two dotted lines that run through the zero point of the x- and y-axis. This makes it possible to, for example, identify cases of prolonged weakness in the bottom left quadrant, with weakness at baseline (x-axis), in combination with progressive weakness in that particular muscle group despite IVIg treatment (y-axis). Using SD units, slope scores for each muscle group were expressed by their regression coefficients and standard error. The coefficient indicates that for each standard deviation loss at baseline, patients gained up to an additional 0.42 standard deviation after IVIg initiation. Abbreviations: CFB = change from baseline, flex = flexion, ext = extension, abd = abduction

Supplemental Table 1 Linear Mixed Models results: Treatment response during follow-up 1 and 2

	Baseline	**Follow-up 1**				**Follow-up 2**			
	Standardised mean (SD)	Change from baseline (SE)	95% CI	SD-change	p value	Change from baseline (SE)	95% CI	SD-change	p value
Pinch	-2.12 (2.99)	0.52 (0.11)	0.30 to 0.75	1.01	<0.001	0.57 (0.17)	0.24 to 0.90	0.99	0.001
Thumb	-2.34 (2.68)	0.56 (0.14)	0.29 to 0.83	0.90	0.001	0.66 (0.19)	0.28 to 1.04	0.94	<0.001
Hand	-1.97 (1.56)	0.58 (0.09)	0.40 to 0.77	0.65	<0.001	0.71 (0.12)	0.47 to 0.96	0.76	<0.001
Wrist	-0.77 (2.17)	0.74 (0.12)	0.49 to 0.98	1.04	<0.001	0.89 (0.17)	0.55 to 1.23	1.00	<0.001
Elbow fl	-0.39 (2.20)	0.91 (0.14)	0.64 to 1.18	1.10	<0.001	0.95 (0.18)	0.59 to 1.31	1.05	<0.001
Elbow ext	-0.14 (1.57)	0.10 (0.09)	-0.07 to 0.27	0.63	0.25	0.28 (0.12)	0.04 to 0.52	0.85	0.024
Shoulder	-0.45 (1.69)	0.66 (0.15)	0.36 to 0.96	1.18	<0.001	0.47 (0.18)	0.10 to 0.84	0.86	0.013
Ankle	-1.38 (1.72)	0.42 (0.15)	0.11 to 0.73	1.12	0.006	0.78 (0.19)	0.41 to 1.15	1.23	<0.001
Knee ext	-0.40 (1.34)	0.25 (0.15)	-0.04 to 0.55	0.96	0.084	0.67 (0.19)	0.28 to 1.05	1.11	0.001
Hip abd	0.30 (1.31)	0.21 (0.14)	-0.07 to 0.49	1.05	0.121	0.55 (0.19)	0.18 to 0.92	1.15	0.004
Hip fl	1.45 (1.49)	0.70 (0.15)	0.41 to 1.00	1.06	<0.001	0.73 (0.18)	0.37 to 1.10	0.93	<0.001

Abbreviations: SD= standard deviation of the population mean; SE = standard error of the estimated mean using a linear mixed effects model; CI = confidence interval; SD Change = crude standard deviation of the change from baseline and reflects the differences between individuals in treatment response; fl = flexion, ext = extension, abd = abduction


4

Chapter 4

HUMAN IMMUNE GLOBULINE 10% WITH RECOMBINANT HUMAN HYALURONIDASE IN MULTIFOCAL MOTOR NEUROPATHY

Ingrid Herraets
Jaap Bakers
Ruben van Eijk
Stephan Goedee
Ludo van der Pol
Leonard van den Berg

Journal of Neurology 2019

Abstract

Objective: The primary aim was to determine the safety of treatment with human immune globulin 10% with recombinant human hyaluronidase (fSCIg) compared to intravenous immunoglobulin (IVIg) in a prospective open-label study in patients with multifocal motor neuropathy (MMN).

Methods: Our study consisted of two phases: the *IVIg phase* (visits 1-3; 12 weeks), in which patients remained on IVIg treatment, and the *fSCIg phase* (visits 4-7; 36 weeks), in which patients received fSCIg treatment. After visit 3, IVIg was switched to an equivalent dose and frequency of fSCIg. Outcome measures were safety, muscle strength, disability and treatment satisfaction.

Results: Eighteen patients were enrolled in this study. Switching to fSCIg reduced the number of systemic adverse events (AE) (IVIg 11.6 *vs.* fSCIg 5.0 AEs/per person-year, $p < 0.02$), and increased the number of local reactions at the injection site (IVIg 0 *vs.* fSCIg 3.3 AEs/per person-year, $p < 0.01$). Overall, no significant differences in muscle strength and disability between fSCIg and IVIg were found. Treatment with fSCIg was perceived as optimal treatment option by 8 of the 17 patients (47.1%) and they continued with fSCIg after study closure because of improved independence and flexibility to administer treatment.

Conclusion: Treatment with fSCIg can be considered a safe alternative for MMN patients on IVIg treatment. fSCIG could be a favorable option in patients who prefer self-treatment and more independency, and in patients who experience systemic adverse events on IVIg or have difficult intravenous access.

Introduction

Multifocal motor neuropathy (MMN) is an immune-mediated neuropathy characterized by asymmetric muscle weakness, predominantly of the upper limbs [1-3]. Various trials have shown a beneficial effect of intravenous immunoglobulins (IVIg) on muscle strength in MMN and a comparable effect of subcutaneous immunoglobulins (SCIg) [4-7].

Although a large number of studies have demonstrated that IVIg treatment is well tolerated, various systemic adverse events have been reported: the majority, such as headache, malaise and chills, are transient and relatively mild, but some rare adverse events, such as anaphylactic and skin reactions, are serious [4]. Moreover, repeated venous access and administration in hospital or at home, in the presence of a nurse, is a burden for the patient. SCIg treatment is considered a good alternative as it can be administered by the patient or informal caregiver and produces fewer systemic adverse reactions [5,8]. However, limitations of subcutaneous infusion volumes and reduced bioavailability require more frequent infusion and an increase in dose in approximately 50% of the patients [5].

A relatively new treatment that overcomes the disadvantages of the conventional SCIg is Human Immune Globulin 10% with recombinant human Hyaluronidase (*fSCIg*). Subcutaneous administration of hyaluronidase increases SCIg dispersion and absorption and therefore provides higher doses of SCIg with less frequent infusion and with the benefit of a higher bio-availability [9-11]. Treatment with fSCIg has been approved by the Food and Drug Administration (FDA) for primary immunodeficiency (PID), but not for inflammatory neuropathies including MMN. This study explores the safety and treatment satisfaction of fSCIg compared to IVIg in MMN patients.

Methods

Study design and patients

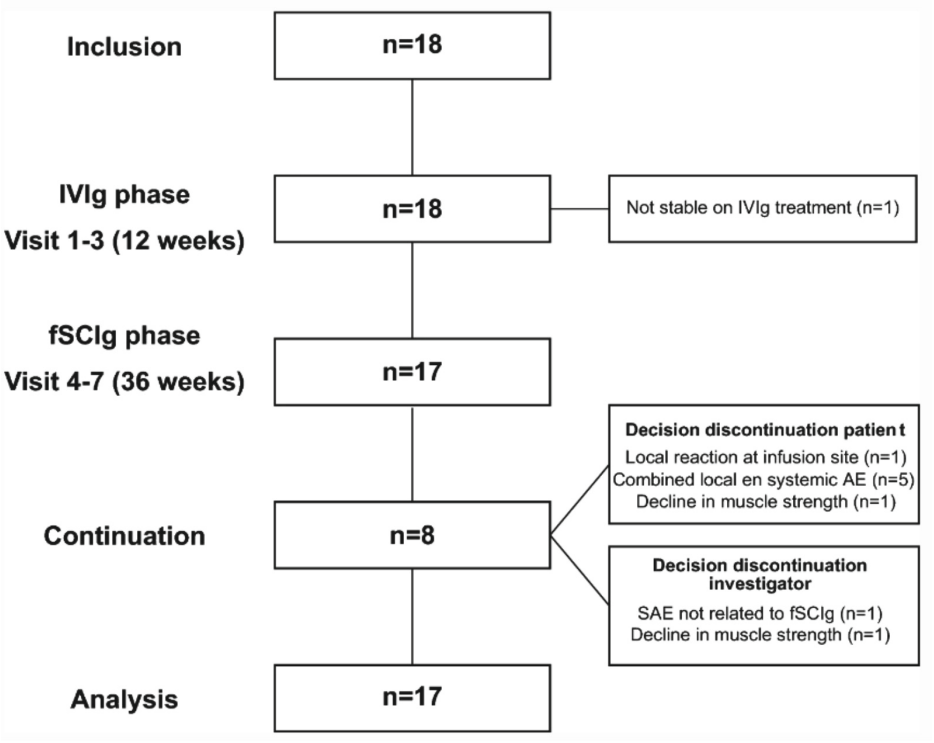
This prospective, open-label study was performed between November 2016 and February 2018 in the UMC Utrecht, a tertiary referral center for neuromuscular disorders. Patients with the diagnosis of MMN according to the EFNS/PNS criteria, who had been stable on IVIg therapy for ≥1 year, were eligible for inclusion in this study. Exclusion criteria for this study were: 1) treatment with other immunosuppressive drugs (e.g. cyclophosphamide, azathioprine, cyclosporine) in the 6 months preceding the study, 2) age <18, and 3) female patient pregnant or breast-feeding. The study

protocol was approved by the local medical ethics committee Utrecht (METC Utrecht; file ID NL52642.041.15) and registered in the Eudra-CT (2015-000828-28) and clinicaltrial.gov (NCT02885259) databases and has, therefore, been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All patients gave written informed consent.

Outcome measures

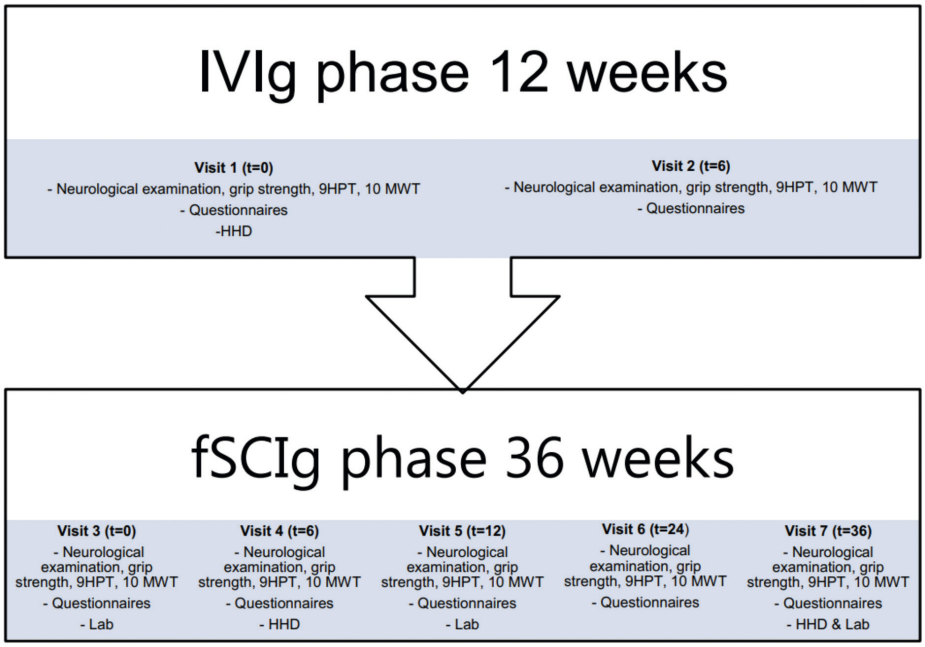
This study consisted of two successive phases: the *IVIg phase* lasting 12 weeks and the *fSCIg phase* of 36 weeks (Figure 1 and Figure 2). During the *IVIg phase*, patients visited the outpatient clinic every 6 weeks (visit 1-3). In the *fSCIg phase*, patients visited the outpatient clinic on weeks 18 (visit 4), 24 (visit 5), 36 (visit 6) and 48 (visit 7). At each visit all outcome measures were collected, except for hand-held dynamometry (HHD) (visits 1-4-7) and laboratory tests (visits 3-5-7) (Figure 2).

Figure 1 Flowchart study



fSCIg = Human Immune Globulin 10% with recombinant human Hyaluronidase, IVIg = intravenous immunoglobulins; Local reactions = local reactions at infusion site, SAE = serious adverse event, systemic AE = systemic adverse event

Figure 2 Outcome measures collected per visit



Questionnaires consisted of a standardized questionnaire for adverse events, treatment satisfaction rated on a 0-10 VAS scale, Guy's Neurological Disability Scale and Self-Evaluation Scale.
fSCIg = Human Immune Globulin 10% with recombinant human Hyaluronidase, HHD = hand-held dynamometry, IVIg = intravenous immunoglobulins, Lab = laboratory tests, t= time in weeks, 9 HPT = 9-Hole Peg Test, 10 MWT = 10-Meter Walk Test

The primary aim was to assess the safety of fSCIg treatment. During the study we documented safety using a standardized questionnaire that included a number of adverse events and laboratory tests, including hemoglobin, hematocrit, haptoglobin, reticulocytes, lactate dehydrogenase, bilirubin, and direct Coombs test to exclude hemolytic anemia due to fSCIg. In addition, blood samples were obtained to explore a possible association between rHuPH20-binding antibody positivity and adverse events. In case of a serious adverse event (SAE) related to fSCIg, the study treatment had to be discontinued. If a patient experienced an adverse event, the investigator or the patient him/herself could decide to discontinue the study treatment and resume regular IVIg treatment.

The second aim of this study was to measure muscle strength. All patients underwent a standardized neurological examination, and motor function of 18 muscle groups (abduction of the arm, flexion and extension of the forearm, wrist and fingers, spreading of the fingers, abduction, adduction and opposition of the thumb, flexion

of the hip, flexion and extension of the knee, and flexion and extension of the foot and toes) was graded bilaterally using the Medical Research Council (MRC) scale to calculate the MRC-sum score. Grip strength was determined bilaterally with the Martin-Balloon-Vigorimeter (Firma Gebrüder Martin, Tuttlingen, Germany) and measured in Kilopascals (kPa) Hand-held dynamometry (HHD) was performed bilaterally in nine muscle groups (abduction of the arm, flexion of the forearm, extension of the wrist and fingers, spreading of the fingers, abduction of the thumb, flexion of the hip, and extension of the foot and big toes) by a physiotherapist using the microFET2 (Hoggan health industries, Draper, UT, USA). Muscle strength with HHD was measured in Newton (N).

In addition, disability was determined with the Guy's Neurological Disability Scale and Self-Evaluation Scale (SES).To measure hand function and finger dexterity, the 9-Hole Peg Test (9-HPT) was performed with the most affected hand and the mean duration (in seconds) of five subsequent trials was calculated. Walking was evaluated with the 10-Meter Walk Test (10 MWT), for which the mean duration (in seconds) and number of steps of three repeats was calculated. Finally, patients were asked to rate their treatment satisfaction on a 0-10 point VAS-scale.

Treatment protocol

During the *IVIg phase*, patients remained on their regular IVIg maintenance therapy regimen to determine their current neurological functioning on therapy. After completion of the *IVIg phase*, patients switched to fSCIg treatment at a dose and frequency equivalent to the IVIg dose and frequency. Personalized titration schedules were devised to increase the dose of fSCIg slowly and thus allow patients to get used to the presence of fluid in their abdominal wall. In general, patients received a dose of 25% of fSCIg in week 1, of 50% in week 2 and their total dose of fSCIg in week 3. IVIg treatment was discontinued when the total dose of fSCIg was administered. Treatment with fSCIg was administered in the patients' home setting. Specialized nurses were present during the first 6 infusions to teach patients how to administer fSCIg and to monitor and treat potential adverse events.

If patients developed a decline in muscle strength during fSCIg treatment, investigators could increase the dose of fSCIg, provided there was no increase in adverse events. This decline in muscle strength was defined as a worsening of ≥1 of the outcome measures: Guy's Neurological Disability Scale (increase ≥1 in either the upper or lower limb score), SES (an increase of ≥1 at ≥2 motor activities) and HHD (a decrease of 50% in ≥2 clinically affected muscles groups). If patients showed

no improvement after increasing fSCIg dose, or if adverse events occurred, fSCIg maintenance treatment was discontinued and IVIg treatment resumed.

Statistical analysis

All data were summarized using the median and range for non-normal distributed variables, and number and percentage for categorical variables. Patient characteristics between patients that continued with fSCIg or discontinued were compared using the Fisher's exact test for categorical variables and the Mann-Whitney U test for non-normal distributed variables. The absolute frequency of adverse events with IVIg and fSCIg were compared using Fisher's exact test. For each patient we determined whether he or she switched back to IVIg, and, if so, the time spent on fSCIg. This time-to-event variable was visualized using Kaplan-Meier curves. Subsequently, we assessed which baseline factors affected the time spent on fSCIg using a Cox proportional hazards model. The mean difference of the HHD measurement was calculated as the difference between first evaluation under fSCIg (visit 4) and baseline (visit 1) and analysed using a paired t-test. The longitudinal outcome measures were analysed using linear mixed effect models (LMMs). The dependency in the data due to the repeated measures was accounted for by a random intercept per individual. The fixed effects part contained a term for treatment (IVIg or fSCIg) and a term for time (in months). Significance of both factors was determined using the likelihood ratio test. Due to the exploratory nature of this study, we did not adjust for multiple testing and results were considered significant when the *p*-value was lower than 0.05. All analysis were conducted in SPSS 22 (SPSS Inc., Chicago IL, USA) except for the LMMs that were fitted using the lmer function in the R package lme4 (version 1.1-12) [12]

Results

Patients

The MMN database of the UMC Utrecht was screened (n=130) and all patients fulfilling the inclusion criteria were invited for participation (n=102). Of these, 54 patients did not respond or could not be reached, and 30 patients declined participation. In total, 18 patients, all treated with IVIg in home setting, were enrolled in this study between November 2016 and May 2017. Clinical characteristics of participants (n=18) and non-participants (n=30) were not significantly different, except for disease duration (6.7 years versus 16.9 years). One patient appeared to be unstable on IVIg treatment during the *IVIg phase* and was excluded from the study. The baseline characteristics

of the remaining 17 patients are provided in Table 1. Two patients were lost to follow-up, both at visit 4 after discontinuation of fSCIg. In 1 patient, visit 4 was missing because of surgery for a hernia. According to the protocol, an increase of dose was required in 1 patient on IVIg treatment and in 3 patients on fSCIg treatment.

Reasons and determinants of discontinuation

Nine patients (52.9%) discontinued fSCIg during the treatment phase after an average number of infusions of 4.7 (SD: 4.6). Baseline characteristics of patients that continued with fSCIg (n=8) and discontinued (n=9) were not significantly different (Table 1). Six participants decided to discontinue because of adverse events (local reactions at the injection site (n=6), nausea (n=1), cramps (n=1), general malaise (n=2), headache (n=1)). One patient showed a decline in muscle strength but refused to increase the dose of fSCIg and chose to switch back to IVIg. The investigators withdrew 2 participants because of an unrelated SAE (ischemic stroke, n=1) and decline in muscle strength despite increasing the dose of the fSCIg (n=1) (Figure 1).

Table 1 Baseline characteristics

	Total cohort (n = 17)	**Continuation fSCIg (n=8)**	**Discontinuation fSCIg (n=9)**	**P-value**
Age at inclusion (years)	57.7 (36.5-69.5)	61.6 (36.6-70.0)	50.0 (46.2-68.9)	0.161
Sex (male)	14 (82.4)	7 (87.5)	7 (77.7)	1.0
Symptom duration (years)	6.9 (2.0-29.9)	6.6 (2.0-29.9)	10.2 (4.9-23.9)	0.67
Duration of IVIg therapy (years)	4.9 (1.2-23.8)	4.3 (1.2-23.8)	4.9 (1.2-13.5)	0.88
Dosage IVIg (g/kg)	0.5 (0.3-2.2)	0.4 (0.3-2.2)	0.6 (0.4-0.6)	0.37
Interval IVIg (days)	21 (7.0-35.0)	21.0 (7.0-28.0)	21.0 (7.0-35.0)	0.37
Abnormal CSF protein	5/6 (83.3)	3/4 (75.0)	2/2 (100.0)	1.0
Abnormal MRI brachial plexus	6/10 (60.0)	2/5 (40.0)	4/5 (80.0)	0.52
Presence of anti-GM1 autoantibodies	11/16 (68.8)	7/8 (87.5)	4/8 (50.0)	0.28

Data are shown for the total cohort (n=17) and for patients that continued with fSCIg (n=8) and discontinued with fSCIg (n=9). Data are in median (range) or n (%).
CSF = cerebrospinal fluid, fSCIg = Human Immune Globulin 10% with recombinant human Hyaluronidase, g/kg = grams per kilo, IVIg = intravenous immunoglobulins.

We evaluated which outcome measures were associated with treatment discontinuation (i.e. treatment satisfaction, Guy's Neurological Disability score, SES, 10 MWT and 9-HPT).

Interestingly, treatment satisfaction was the only baseline factor associated with continuation of fSCIg: a higher satisfaction during the *IVIg phase* of the trial was associated with the continuation of fSCIg (HR 0.31 95% CI 0.12 – 0.83, $p = 0.007$). To exemplify: after 6-months, 78% of the patients, whose satisfaction with IVIg treatment was initially ≥8, remained on fSCIg, *vs* 25% of patients with a satisfaction rate < 8 (Supplemental figure 1).

Safety

Frequencies of AEs, AEs per year and AEs per patient are shown for IVIg and fSCIg in Table 2. The frequency of systemic adverse events was lower in fSCIg (n = 87 on IVIg *vs.* n = 35 on fSCIg, $p = 0.04$); headache and general malaise occurred less often in fSCIg ($p < 0.01$; $p < 0.01$), while cramps and local reactions at the injection site occurred more often ($p = 0.03$ and $p < 0.01$). None of the patients developed hemolytic anemia, nor did any develop rHuPH20-binding antibodies after initiation of fSCIg treatment.

During the study, 3 SAEs (coronary artery disease, ischemic stroke and diabetes mellitus) occurred in 2 patients (**Table 2**). Thrombosis is a rare adverse event of immunoglobulin treatment. However, all SAEs were considered unrelated to fSCIg treatment. The first patient reported angina pectoris at visit 4, during fSCIg treatment, but, in retrospect, this complaint had already been present 3 months before the start of the study (during treatment with IVIg), and had not been reported at visits 1-3. After cardiological evaluation, coronary artery disease was diagnosed. The cardiovascular risk profile of this patient consisted of hypertension, hypercholesterolemia, recurrent transient ischemic attacks (TIAs) treated with carotid endarterectomy and smoking. Between visits 6 and 7 (during IVIg treatment), the same patient had been admitted to hospital because of new onset diabetes mellitus. The second patient reported headache and visual complaints; i.e. spots in the left visual field, after only one low dose of fSCIg (10 gram) combined with a regular high dose of IVIg (40 gram). MRI cerebrum showed a small occipital lobe infarction. After extensive work-up performed by a neurovascular specialist, a combination of cardiovascular risk factors (hypercholesterolemia, hypertension and smoking) was deemed to be the most likely cause. During follow-up, this patient made a full recovery. Recovery of the visual field was confirmed by a normal perimetry examination performed by an ophthalmologist.

Table 2 Safety profile of IVIg and fSCIg

	IVIg		fSCIg		*p*-value
	Frequency	*Rate*	*Frequency*	*Rate*	
Any systemic AE	81(14)	11.6	35 (11)	5.0	0.02
Skin reactions	12 (5)	1.6	6 (4)	0.9	0.79
Dizziness	4 (2)	0.5	2 (2)	0.3	1.00
Headache	26 (6)	3.5	6 (3)	0.9	<0.01
General malaise	17 (6)	2.3	2 (2)	0.3	<0.01
Fatigue	18 (5)	2.4	8 (3)	1.1	0.36
Increased hunger sensation	4 (1)	0.5	3 (1)	0.4	0.43
Cramps	1 (1)	0.1	5 (4)	0.7	0.03
Diarrhea	0 (0)	0.0	1 (1)	0.1	0.39
Dry mouth	0 (0)	0.0	1 (1)	0.1	0.39
Nausea	0 (0)	0.0	1 (1)	0.1	0.39
Lumbago	1 (1)	0.1	0 (0)	0.0	1.00
Palpitations	1 (1)	0.1	0 (0)	0.0	1.00
Hypertension	3 (2)	0.4	0 (0)	0.0	0.28
Local reactions at injection site	0 (0)	0.0	23 (11)	3.3	<0.01
Serious adverse event	3 (2)	0.1	0 (0)	0.0	0.29

AE = adverse event, frequency = absolute frequency of adverse events, in brackets are the unique patients; fSCIg = Human Immune Globulin 10% with recombinant human Hyaluronidase, IVIg= intravenous immunoglobulins; P-value = comparison of absolute frequency of AE with IVIg and fSCIg, Rate = number of adverse events/per person-year

Muscle strength and disability

Overall, there were no significant differences between fSCIg and IVIg expressed in vigorimetry, 9-HPT, MRC sum score or HHD total score (all $p > 0.18$). Interestingly, there was a strong improvement over time in the 10-meter walk test (both in time taken and number of steps, p-values < 0.001, Table 3). This observation may suggest a learning effect over time. Despite the adjustment for time, this learning effect might obscure accurate estimation of the difference between fSCIg and IVIg in the 10-meter walk test. The SES increased by 0.6 points (95% CI 0.1 – 1.2, $p = 0.021$) when switching to fSCIg. The deterioration in SES is temporary and improvable as it is most likely caused by a decline in muscle strength of one patient at visit 5, with a normalisation of the score when the dose of fSCIg was increased. Excluding this patient results in an increase in SES of 0.4 points (95% CI -0.1 – 0.8, p-value = 0.097).

Treatment satisfaction and reasons for continuation

Overall, treatment satisfaction remained unchanged. The average treatment satisfaction with regard to IVIg and fSCIg was 7.9 (95% CI 7.3 to 8.5) and 7.5 (95% CI

6.8 to 8.1), respectively. Main reasons for continuation of fSCIg were independence to administer treatment (n=8) and decrease in presence of adverse events (general malaise (n=1), skin reaction (n=1)).

Table 3 Longitudinal outcome measures

Endpoint	**Intercept**	**Treatment** *(IVIg vs. fSCIg)*			**Time** *(Months)*		
		Coefficient	**95% CI**	***p*-value**	**Coefficient**	**95% CI**	***p*-value**
Vigorimetry, *kPa*	131.2	-2.9	-12.2 – 6.4	0.54	0.5	-0.7 – 1.6	0.41
MRC sum (0-180)	163.4	-0.8	-2.2 – 0.7	0.30	0.2	0.0 – 0.3	0.051
SES	11.4	0.6	0.1 – 1.2	0.021	0.0	0.0 – 0.1	0.36
10-meter walk, *steps*	14.0	0.3	0.1 – 0.6	0.020	-0.1	-0.1 – 0.0	< 0.001
10-meter walk, *seconds*	8.4	0.3	0.0 – 0.5	0.040	-0.1	-0.1 – -0.1	< 0.001
9-hole peg, *seconds*	32.3	2.5	0.0 – 5.1	0.051	-0.2	-0.6 – 0.1	0.12
Treatment satisfaction	7.9	-0.5	-1.0 – 0.0	0.067	0.0	-0.1 – 0.1	0.82

Results are given per mixed model with a fixed effect for treatment and a random intercept per individual (n = 17), adjusted for time to account for potential disease progression during study follow-up. The treatment estimate is the mean difference between treatment arms. fSCIg = Human Immune Globulin 10% with recombinant human Hyaluronidase, IVIg = intravenous immunoglobulins, kPa = Kilopascals, MRC = Medical Research Council scale, SES = Self-Evaluation Scale

Table 4 Mean difference in Hand-held dynamometry

Endpoint	**Mean difference** *(Post – Pre)*	**95% CI**	***p*-value**
Dynamometry, *Newton*			
Total	-21.3	-75.8 – 33.3	0.42
Shoulder abduction	-11.4	-33.0 – 10.2	0.28
Biceps flexion	-1.1	-11.5 – 9.2	0.82
Wrist extension	1.5	-7.2 – 10.2	0.72
Finger extension	2.5	-4.5 – 9.5	0.46
Finger spreading	2.5	0.0 – 5.0	0.049
Thumb abduction	-1.8	-4.5 – 0.9	0.18
Hip flexion	-4.8	-16.3 – 6.7	0.38
Ankle flexion	-7.5	-24.8 – 9.8	0.37
Toe extension	-1.2	-9.9 – 7.6	0.78

Two patients were excluded due to missing data of fSCIg. The mean difference was calculated as the difference between first evaluation under HyQvia (Visit 4) and baseline (Visit 1).
95% CI = 95% confidence interval.

Discussion

In the present study, we showed that the safety, of fSCIg is comparable to IVIg. In The Netherlands, approximately 90% of the patients with MMN are treated in a home care program, in contrast to countries where IVIg treatment is only given in a hospital setting. For this reason, the satisfaction rate for IVIg treatment is high, as there is no burden of travelling to hospital. Nevertheless, fSCIg was preferred compared to IVIg treatment by almost half of the patients, and they continued with fSCIg after study closure, in particular because of independence and flexibility to administer treatment and a decrease in systemic adverse events. A significant number of patients remained on IVIg treatment, probably because the benefits of fSCIg (i.e. more independence and flexibility of administration and a decrease in systemic adverse events) did not outweigh the well-facilitated IVIg home program due to the local reactions at fSCIg injection site. Moreover, in countries which do not offer the option of IVIg treatment in home setting, fSCIg could be an even more favorable option.

Regarding safety of fSCIg, we found similar results compared to previous publications on SCIg in MMN or to fSCIg in primary immunodeficiencies, and to a recently published study that compared fSCIG with conventional SCIg in 20 MMN patients [5-7,9,11,13-16]. We reported local reactions at the injection sites in 64.7% of the patients, which is in accordance with previous studies that described local adverse reactions of fSCIg in 44-100% of the patients [5-7,16,17]. A systematic review and meta-analysis reported a significant reduction of 28% in the relative risk ratio of systemic adverse events of SCIg compared to IVIg; this is comparable to a significant reduction in systemic adverse events of fSCIg versus IVIg in our study [15]. Overall, similar to previous studies, muscle strength, disability and treatment satisfaction in our study remained stable, showing equal muscle strength and disability and unchanged or improved quality of life and treatment satisfaction for SCIg compared to IVIg in MMN patients [5-7,13,14,16]. Therefore, fSCIg could be a favorable alternative to IVIg treatment in MMN, as systemic adverse events may decrease, muscle strength, disability and treatment satisfaction remains stable, and there is the advantage of independence and flexibility of administration. Moreover, professional supervision of administration is not necessary for fSCIg treatment and could therefore reduce medical costs [17-19].

An advantage of treatment with fSCIg is the reduced number of infusions compared to SCIg. This is relevant since we have previously found that patients in The Netherlands preferred IVIg in home setting to SCIg because of the high number

of infusions of SCIg (unpublished data). This is in accordance with the results of a randomized single blinded cross-over trial and follow-up study by Harbo et al, investigating SCIg versus IVIg [6,20]. In this study, 4/9 patients preferred IVIg to SCIg especially because of the significantly lower number of infusions. Additionally, fSCIg allows self-administration of loading doses if necessary, as opposed to SCIg treatment, which requires IVIg loading doses and hence loss of independence and flexibility of administration [7,21].

No clinical outcomes were associated with an increased risk of discontinuing fSCIg. Remarkably, the only prognostic factor for continuation of fSCIg was a higher (≥8) satisfaction in the *IVIg phase* of the study. These findings may be explained by the expectation level of patients regarding treatment with fSCIg. Patients who were less satisfied with IVIg treatment may have had higher expectations of fSCIg treatment, but as muscle strength, disability and treatment satisfaction were comparable to IVIg, these expectations may not have been met, causing patients to discontinue fSCIg earlier. Interestingly, in this study, patients who continued with fSCIg after study closure were more satisfied compared to their previous IVIg treatment because of the independence and flexibility of administration.

Study limitations include the relatively small number of patients, a common challenge in studies on rare disorders such as MMN. We were able to contact 48 MMN patients of whom 18 (38%) participated and clinical characteristics of participants and non-participants were not significantly different. Furthermore, the study design was a prospective cohort and not a randomized controlled blinded trial. However, the route of administration of fSCIg did not allow a blinded study design, and as IVIg is standard of care this would limit the possibility withholding patients from immunoglobulin treatment. Moreover, we believe this study design was adequate for our aim to explore whether fSCIg could replace IVIg in individual patients, and whether it could serve as an alternative route of administration in a relatively rare disorder.

In conclusion, our study shows that safety of fSCIg is comparable to IVIg. Overall muscle strength, disability and treatment satisfaction remained unchanged after switch to fSCIg. Therefore. fSCIG could be a favorable option in patients who prefer self-treatment and more independency, and in patients who experience systemic adverse events on IVIg or have difficult intravenous access.

References

1. Van Asseldonk JT, Franssen H, Van den Berg-Vos RM, Wokke JH, Van den Berg LH (2005) Multifocal motor neuropathy. Lancet Neurol 4 (5):309-319. doi:10.1016/S1474-4422(05)70074-0
2. Vlam L, van der Pol WL, Cats EA, Straver DC, Piepers S, Franssen H, van den Berg LH (2011) Multifocal motor neuropathy: diagnosis, pathogenesis and treatment strategies. Nat Rev Neurol 8 (1):48-58. doi:10.1038/nrneurol.2011.175
3. Joint Task Force of the E, the PNS (2010) European Federation of Neurological Societies/Peripheral Nerve Society guideline on management of multifocal motor neuropathy. Report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society--first revision. J Peripher Nerv Syst 15 (4):295-301. doi:10.1111/j.1529-8027.2010.00290.x
4. van Schaik IN, van den Berg LH, de Haan R, Vermeulen M (2005) Intravenous immunoglobulin for multifocal motor neuropathy. Cochrane Database Syst Rev (2):CD004429. doi:10.1002/14651858.CD004429.pub2
5. Markvardsen LH, Harbo T (2017) Subcutaneous immunoglobulin treatment in CIDP and MMN. Efficacy, treatment satisfaction and costs. J Neurol Sci 378:19-25. doi:10.1016/j.jns.2017.04.039
6. Harbo T, Andersen H, Hess A, Hansen K, Sindrup SH, Jakobsen J (2009) Subcutaneous versus intravenous immunoglobulin in multifocal motor neuropathy: a randomized, single-blinded cross-over trial. Eur J Neurol 16 (5):631-638. doi:10.1111/j.1468-1331.2009.02568.x
7. Eftimov F, Vermeulen M, de Haan RJ, van den Berg LH, van Schaik IN (2009) Subcutaneous immunoglobulin therapy for multifocal motor neuropathy. J Peripher Nerv Syst 14 (2):93-100. doi:10.1111/j.1529-8027.2009.00218.x
8. Cocito D, Merola A, Peci E, Mazzeo A, Fazio R, Francia A, Valentino P, Liguori R, Filosto M, Siciliano G, Clerici AM, Lelli S, Marfia GA, Antonini G, Cecconi I, Nobile-Orazio E, Lopiano L, Scig, Chronic Dysimmune Neuropathies Italian N (2014) Subcutaneous immunoglobulin in CIDP and MMN: a short-term nationwide study. J Neurol 261 (11):2159-2164. doi:10.1007/s00415-014-7444-2
9. Wasserman RL, Melamed I, Stein MR, Gupta S, Puck J, Engl W, Leibl H, McCoy B, Empson VG, Gelmont D, Schiff RI, Igsc wrSG (2012) Recombinant human hyaluronidase-facilitated subcutaneous infusion of human immunoglobulins for primary immunodeficiency. J Allergy Clin Immunol 130 (4):951-957 e911. doi:10.1016/j.jaci.2012.06.021
10. Wasserman RL, Melamed I, Stein MR, Engl W, Sharkhawy M, Leibl H, Puck J, Rubinstein A, Kobrynski L, Gupta S, Grant AJ, Ratnayake A, Richmond WG, Church J, Yel L, Gelmont D (2016) Long-Term Tolerability, Safety, and Efficacy of Recombinant Human Hyaluronidase-Facilitated Subcutaneous Infusion of Human Immunoglobulin for Primary Immunodeficiency. J Clin Immunol 36 (6):571-582. doi:10.1007/s10875-016-0298-x
11. Sanford M (2014) Human immunoglobulin 10 % with recombinant human hyaluronidase: replacement therapy in patients with primary immunodeficiency disorders. BioDrugs 28 (4):411-420. doi:10.1007/s40259-014-0104-3
12. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. J Stat Soft 67 (1):1-48. doi:10.18637/jss.v067.i01 %& 1
13. Sala TP, Crave JC, Duracinsky M, Lepira Bompeka F, Tadmouri A, Chassany O, Cherin P (2018) Efficacy and patient satisfaction in the use of subcutaneous immunoglobulin immunotherapy for the treatment of auto-immune neuromuscular diseases. Autoimmun Rev 17 (9):873-881. doi:10.1016/j.autrev.2018.03.010
14. Hachulla E, Benveniste O, Hamidou M, Mouthon L, Schleinitz N, Lozeron P, Leger JM, Vial C, Viala K (2017) High dose subcutaneous immunoglobulin for idiopathic inflammatory myopathies and dysimmune peripheral chronic neuropathies treatment: observational study of quality of life and tolerance. Int J Neurosci 127 (6):516-523. doi:10.1080/00207454.2016.1204544

15. Racosta JM, Sposato LA, Kimpinski K (2017) Subcutaneous versus intravenous immunoglobulin for chronic autoimmune neuropathies: A meta-analysis. Muscle Nerve 55 (6):802-809. doi:10.1002/mus.25409

16. Al-Zuhairy A, Jakobsen J, Andersen H, Sindrup SH, Markvardsen LK (2019) Randomized trial of facilitated subcutaneous immunoglobulin in multifocal motor neuropathy. Eur J Neurol. doi:10.1111/ene.13978

17. Gardulf A, Andersen V, Bjorkander J, Ericson D, Froland SS, Gustafson R, Hammarstrom L, Jacobsen MB, Jonsson E, Moller G, et al. (1995) Subcutaneous immunoglobulin replacement in patients with primary antibody deficiencies: safety and costs. Lancet 345 (8946):365-369

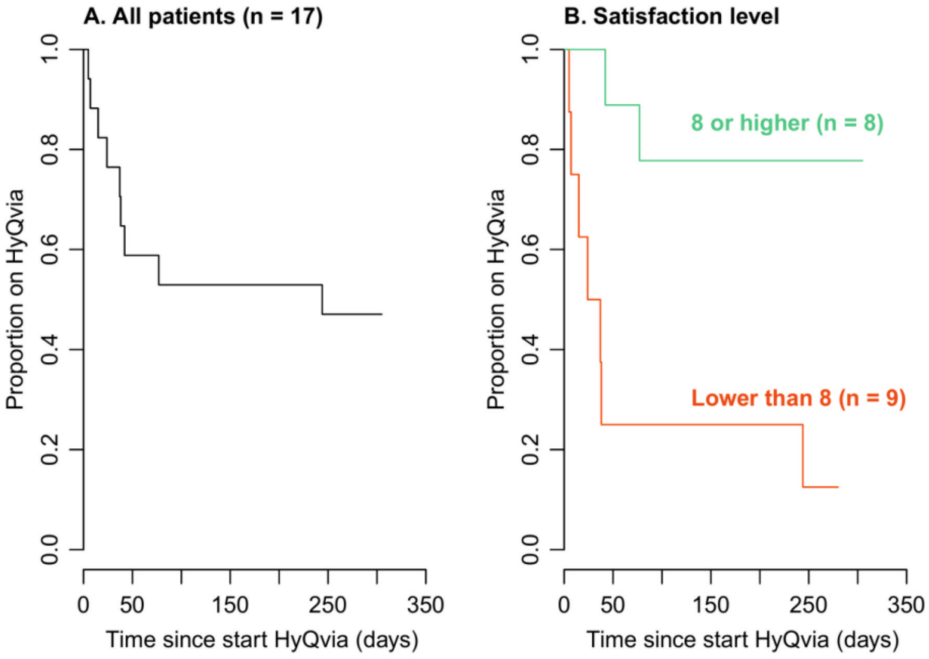
18. Cocito D, Serra G, Paolasso I, Barila DA, Lopiano L, Cattel L (2012) Economic and quality of life evaluation of different modalities of immunoglobulin therapy in chronic dysimmune neuropathies. J Peripher Nerv Syst 17 (4):426-428. doi:10.1111/j.1529-8027.2012.00444.x

19. Lazzaro C, Lopiano L, Cocito D (2014) Subcutaneous vs intravenous administration of immunoglobulin in chronic inflammatory demyelinating polyneuropathy: an Italian cost-minimization analysis. Neurol Sci 35 (7):1023-1034. doi:10.1007/s10072-014-1632-9

20. Harbo T, Andersen H, Jakobsen J (2010) Long-term therapy with high doses of subcutaneous immunoglobulin in multifocal motor neuropathy. Neurology 75 (15):1377-1380. doi:10.1212/WNL.0b013e3181f735ce

21. Hadden RD, Marreno F (2015) Switch from intravenous to subcutaneous immunoglobulin in CIDP and MMN: improved tolerability and patient satisfaction. Ther Adv Neurol Disord 8 (1):14-19. doi:10.1177/1756285614563056

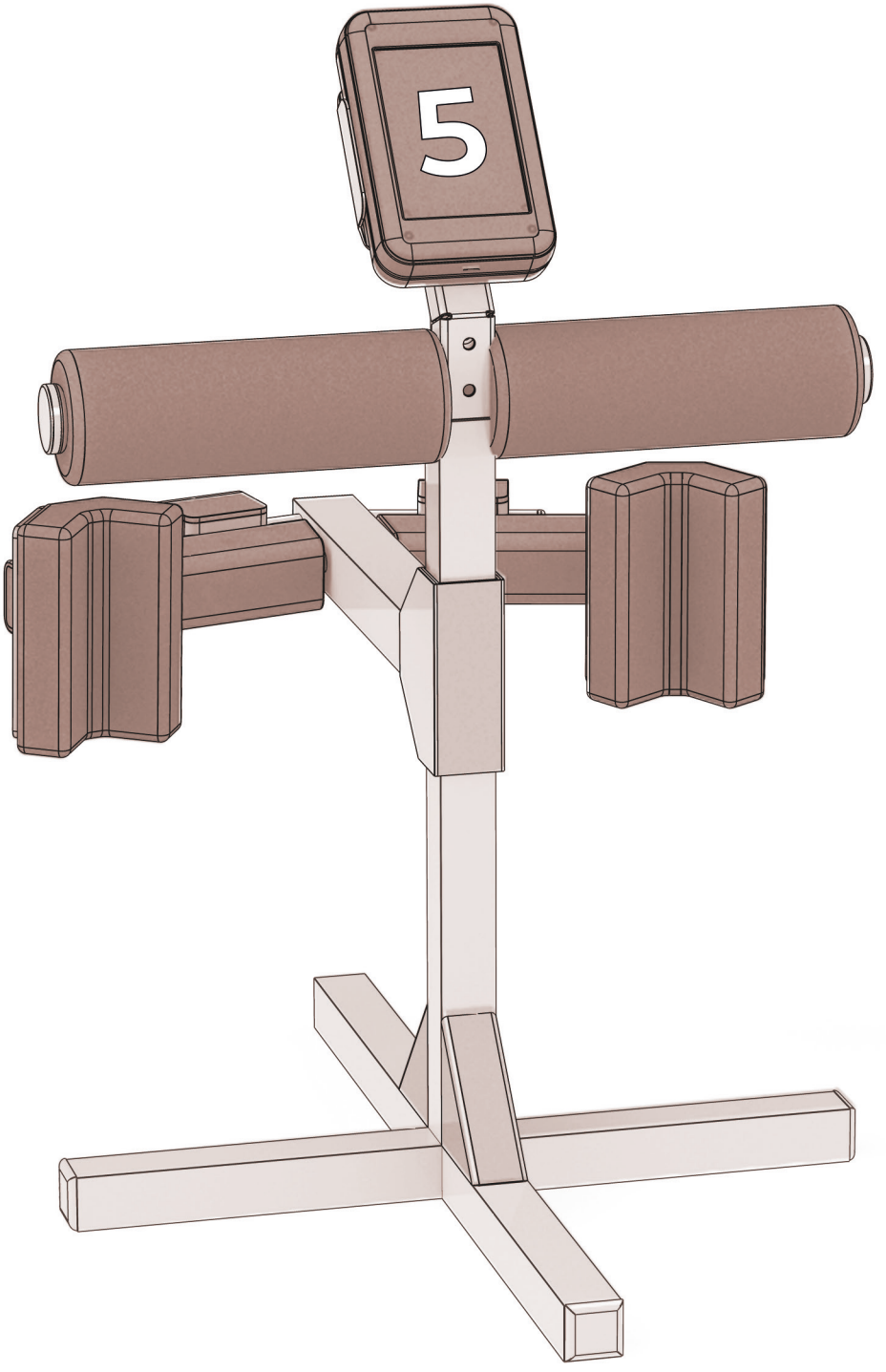
Supplemental figure 1. Proportion of patients remaining on fSCIg treatment



(A) Kaplan-Meier curve of the proportion of patients on fSCIg treatment, the median time on fSCIg treatment was 244 days (n=17); for patients that continued with fSCIg 267 days and patients that discontinued with fSCIg 37 days. (B) For each patient, the average treatment satisfaction score on IVIg was calculated during phase 1 (visit 1-3) and assessed in a Cox proportional hazards models (HR 0.31 95% CI 0.12 – 0.83, *p*-value = 0.007). To visualize its effects, we created two subgroups (green line; higher satisfaction level on IVIg and red line; lower satisfaction level on IVIg) based on the median of this satisfaction level.
fSCIg = Human Immune Globulin 10% with recombinant human Hyaluronidase, IVIg = intravenous immunoglobulins

Part 2

CURRENT LIMITATIONS OF CLINICAL OUTCOME MEASUREMENTS IN MOTOR NEURON DISORDERS


5

Chapter 5

USING THE ALSFRS-R IN MULTICENTRE CLINICAL TRIALS FOR AMYOTROPHIC LATERAL SCLEROSIS: POTENTIAL LIMITATIONS IN CURRENT STANDARD OPERATING PROCEDURES

Jaap Bakers
Adriaan de Jongh
Tommy Bunte
Lindsay Kendall
Steve Han
Noam Epstein
Arseniy Lavrov
Anita Beelen
Anne Visser-Meily
Leonard van den Berg
Ruben van Eijk

Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration 2022

Abstract

Objective: Uniform data collection is fundamental for multicentre clinical trials. We aim to determine the variability, between ALS trial centres, in the prevalence of a sudden increase in the revised ALS functional rating scale (ALSFRS-R) score, and its associations with individual patient and item characteristics.

Methods: We used data from two multicentre studies to estimate the prevalence of a sudden increase in the ALSFRS-R score, defined as an increase of 5 points or more between two consecutive, monthly visits. For each patient with a 5-point or more increase, we evaluated the individual contribution of each ALSFRS-R item.

Results: Longitudinal ALSFRS-R scores, originating from 114 trial centres enrolling a total of 1,240 patients, were analysed. A 5-point or more increase in ALSFRS-R total score was found in 151 (12.2%) patients, with prevalences per study centre ranging from 0% to 83%. Bulbar onset, faster disease progression at enrolment, and a lower ALSFRS-R score at baseline were associated with a sudden 5-point or more increase in the ALSFRS-R total score. ALSFRS-R items 2 (saliva), 9 (stairs), 10 (dyspnoea), and 11 (orthopnoea) were the primary drivers when a 5-point or more increase occurred.

Conclusions: Sudden 5-point or more increases in ALSFRS-R total scores between two consecutive visits are relatively common. These sudden increases were not found to occur with equal frequency in trial centres; which may underscore the need for amending existing standard operating procedures and monitoring of data quality during the study, in multicentre research.

Introduction

Amyotrophic lateral sclerosis (ALS) is characterized by unrelenting functional loss over time with extensive variation between patients (1). Currently, the revised ALS functional rating scale (ALSFRS-R) is the most commonly used primary endpoint in clinical trials (2). Regulators (2-4) encourage the use of the ALSFRS-R which had proven validity and reliability (5, 6). The ALSFRS-R contains 12 items evaluating different aspects of the patient's daily physical functioning and symptomatology, and the score has been shown to be related to overall survival time (7).

Despite being widely adopted in both clinical trials and care, rating of the ALSFRS-R in a clinical trial context may not be straightforward and requires specific training (8, 9). Item categories may be interpreted differently depending on the rater. For example, the natural history for item 10 (dyspnoea) should be a uniformly declining function over time as ALS is a progressive disorder and the ALSFRS-R is intended to monitor disease progression (5). At a certain point in the disease, however, non-invasive ventilation may be initiated, resolving symptoms of dyspnoea. This increases the score of item 10, which may falsely indicate a true improvement in the patient's respiratory function, but also change the interpretation of item 10 as it now reflects the residual presence of symptoms under respiratory support.

The above scenario should be resolved, preferably using standard operating procedures (SOPs), and by training evaluators to ensure adequate and uniform scoring strategies across raters and centres. Nevertheless, training may be suboptimal, or the SOP may overlook certain clinical scenarios, which could result in unnatural sudden changes in ALSFRS-R scores. Long-term improvements in the ALSFRS-R, i.e. reversals, have been reported (10-12), but the prevalence of sudden increases in ALSFRS-R trajectories remains unknown. These may highlight limitations in the current SOPs and scoring strategies, which could impact the accuracy of patient monitoring. In this study, therefore, we aim to determine the prevalence of sudden increases in ALSFRS-R between two subsequent visits, and evaluate variability between centres in clinical trials. In addition, we explore which patient- and item-related characteristics are associated with the prevalence of these sudden increases in the total score.

Methods

Individual participant data

For this study, we used data from two multicentre clinical trials to estimate the prevalence of sudden increases in ALSFRS-R total scores. The EMPOWER clinical trial was a randomized placebo-controlled clinical trial that evaluated the safety and efficacy of dexpramipexole in patients with ALS (13). A total of 942 patients were enrolled in 80 trial centres across 11 countries between March 2011 and September 2011. The primary outcome was the Combined Assessment of Function and Survival (CAFS), a joint-rank score of the ALSFRS-R and survival time, at 12 months (14). The ALSFRS-R was measured at monthly intervals for at least 12 months. The second trial assessed the safety and efficacy of ozanezumab compared to placebo in patients with ALS (15). A total of 303 patients were enrolled in 34 trial centres across 11 countries between December 2012 and November 2013. The primary outcome was the joint-rank analysis of function (ALSFRS-R) and survival at week 48, with ALSFRS-R scores obtained at monthly intervals. Both studies concluded a lack of efficacy; therefore, anonymised individual patient data from both the placebo and active arms were used in the current study. For both studies, raters and centres were certified by ALSFRS-R outcome measure training, and employed SOP guidance as provided by the Northeast ALS Consortium (NEALS) (16).

Statistical analysis

We considered an increase of 5 points or more between two consecutive monthly visits to be an unnatural, sudden change. The cut-off was based on test-retest reliability data published previously (8), indicating that scores within patients may vary up to 4.3 points due to random variability. Per patient, we calculated the sequential difference between two longitudinal ALSFRS-R measurements. To illustrate, if the ALSFRS-R total score was 43 at screening, 42 at baseline, and 40 at month 1, the sequential difference between visits was -1 (baseline - screening) and -2 (month 1- baseline). Subsequently, we evaluated whether a patient encountered any sequential difference equal to, or larger than +5 points and flagged these patients as having an unnatural, sudden change. Finally, we determined the number of patients with at least one sudden increase of 5 points or more per trial and per trial centre. As a sensitivity analysis, we adjusted the sequential difference between visits for the time between two visits, due to the fact that visits may not occur exactly at monthly intervals (adjusted difference = difference / time between visits in months). A patient

was subsequently flagged as having an unnatural, sudden change if the adjusted difference was larger than or equal to +5 points per month.

To distinguish the variability in prevalence between centres from random noise and potential underperformance of a particular centre, for each centre we calculated the probability of observing the number of patients with a 5-point or more increase out of the total number of patients enrolled in that centre, given the average background prevalence observed in the other trial centres. For example, one can calculate that the probability of observing 2 patients with a 5-point or more increase out of the 4 enrolled patients per centre, with a background prevalence of 15%, based on the binomial distribution, is 0.098. Centres with a probability less than 0.05 were flagged as potential outliers. In case of significant variability between centres, logistic regression models were used to evaluate whether the centre prevalence of patients with a sudden increase depended on the centre's number of enrolled patients.

Descriptive statistics were used to compare patients with and without a sudden 5-point or more increase. Baseline data were summarized using the mean and standard deviation (SD) for continuous variables, or number and percentage for categorical variables. Means or proportions were compared using Student's *t* or Chi-square tests, respectively. Finally, for each patient with a 5-point or more increase, we evaluated the individual contribution of each ALSFRS-R item by calculating the change per item and expressing it as a proportion of the total change.

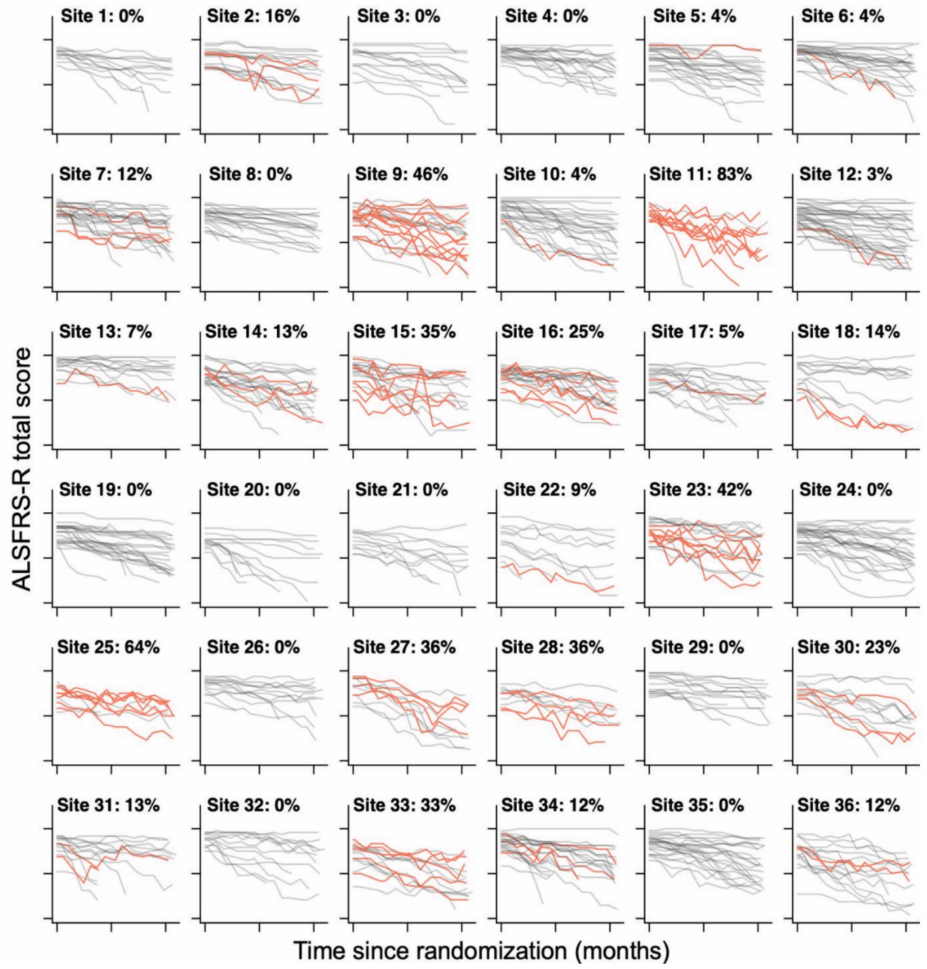
Results

In total, 14,297 longitudinal ALSFRS-R scores were analysed; these originated from 114 trial centres enrolling a total of 1,240 patients. Five patients in the dexpramipexole trial were excluded from the analysis as only one ALSFRS-R measurement was available. In the dexpramipexole trial, the number of enrolled patients per centre varied from 1 to 37 per centre (median 9 patients per centre), whereas in the ozanezumab trial, this ranged from 2 to 23 per centre (median 8 patients per centre).

Prevalence of sudden increases in the dexpramipexole and ozanezumab trials
Pooled across trials, we identified a total of 151 patients with at least one 5-point or more increase between two consecutive, monthly visits. In the dexpramipexole trial, 123 out of 937 (13.1%) patients had at least one 5-point or more increase, and in the ozanezumab trial, 28 out of 303 (9.2%) patients, resulting in an average prevalence of 12.2% (95% CI 10.4% to 14.2%). This percentage was similar when

adjusted for the time between visits (12.1% in dexpramipexole and 9.9% in the ozanezumab trial). Importantly, among trial centres there was extensive variability in the prevalence of sudden increases, illustrated in **Figure 1** for the 36 largest trial centres of the dexpramipexole trial. Prevalence per centre ranged from 0% to 83% in the dexpramipexole trial and from 0% to 40% in the ozanezumab trial. In **Figure e1** (dexpramipexole) and **Figure e2** (ozanezumab) we present the observed prevalence for each centre.

Figure 1. Centre variability in the prevalence of a 5-point or greater increase in ALSFRS-R total score

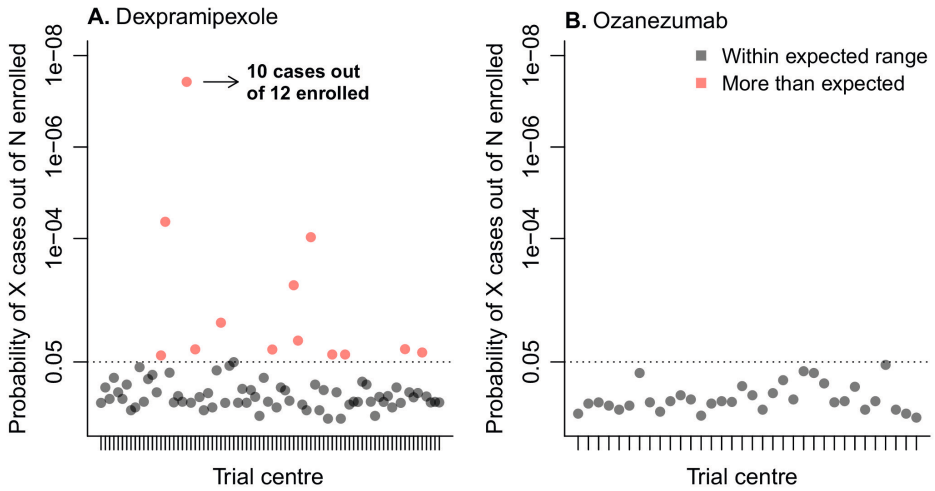


Center variability in the prevalence of a 5-point or greater increase in ALSFRS-R total score. Raw 12-month ALSFRS-R data from the 36 trial centers with the largest number of enrolled patients in the dexpramipexole trial. Per center, the number of patients with a 5-point or more increase in ALSFRS-R total score are highlighted in red. The percentage per center indicates the proportion of patients with an increase, which ranges from 0% to 83%.

Exploring between-centre variability: chance vs. underperformance

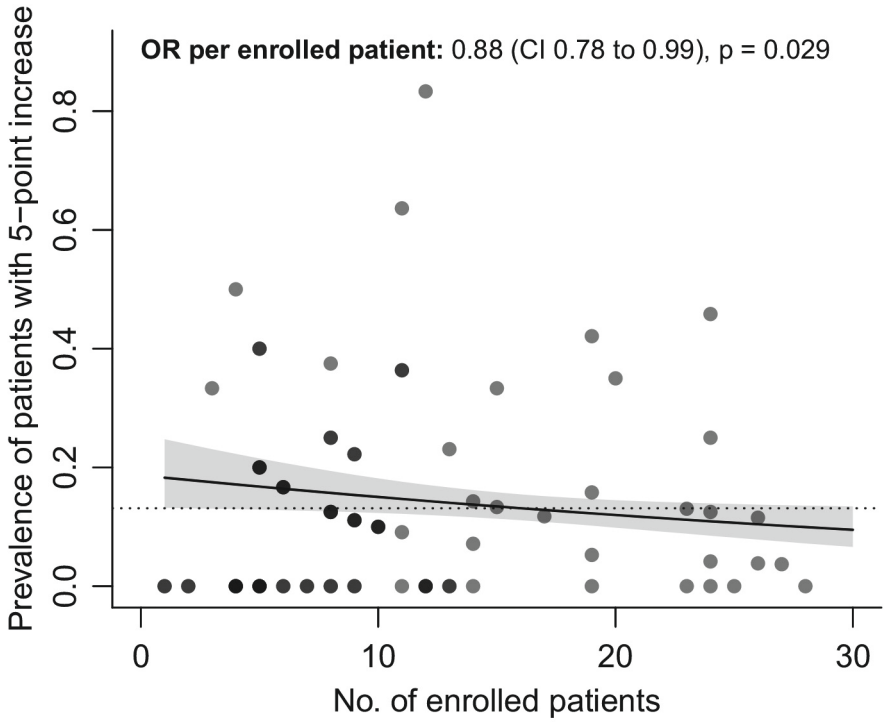
To distinguish the variability between centres from random noise and potential underperformance of a particular centre, for each centre we calculated the probability of observing the number of sudden increases given the average background prevalence observed in the other trial centres and the number of enrolled patients. These probabilities are presented in **Figure 2**. To illustrate: the probability of observing 10 sudden increases in 12 enrolled patients is 4 in one-hundred million, which makes it highly unlikely that this number of sudden increases is due to chance, and may suggest underperformance of a trial centre compared to the other centres. In total, we identified 13 (16.3%) sites in the dexpramipexole trial that had a probability lower than 5%, suggesting a high likelihood of systematic differences between centres. The variability between centres in the dexpramipexole trial was related to the number of enrolled patients (odds ratio per patient 0.88, 95% CI 0.1678 to 0.2099, p = 0.029, **Figure 3**), where sites that enrolled a higher number of patients had, on average, a lower prevalence of patients with a 5-point or more increase. In contrast, between-centre variability in the ozanezumab trial fell within the expected range, suggesting that centres were performing similarly and no clear underperforming centres could be identified.

Figure 2. Probability of observing the number of patients with a 5-point or more increase per trial centre



Probability of observing the number of patients with a 5-point or more increase per trial center. To distinguish the variability in prevalence between centers from random noise and potential underperformance of a particular center, for each center we calculated the probability of observing a particular number of patients with a 5-point or more increase (cases) out of the total number of patients enrolled in that center, given the average background prevalence observed in other trial centers. Centers with a probability less than 5% (dotted line) were flagged as potential outliers.

Figure 3. Number of enrolled patients *vs.* prevalence of sudden 5-point or more increase



Number of enrolled patients vs. prevalence of sudden 5-point or more increase. Relationship between the number of enrolled patients per trial center who participated during the dexpramipexole trial, and their association with the prevalence of a sudden 5-point or more increase. Darker dots represent overlapping centers. Solid line: regression line estimate with 95% confidence interval. Dashed line: average prevalence in the dexpramipexole trial. OR: odds ratio; CI: confidence interval; No: number.

Patient characteristics associated with potential measurement errors

In **Table 1** we provide the individual characteristics of patients with and without a 5-point or more increase. Patients with high baseline ALSFRS-R scores have fewer 5-point or more increases during follow-up, possibly as they are less likely to gain 5 points or more in their ALSFRS-R total score. Patients with bulbar onset and faster disease progression (expressed as ΔFRS)(7) were more likely to have a 5-point or more increase.

Table 1. Baseline characteristics of patients with and without a 5-point or more increase in ALSFRS-R total score

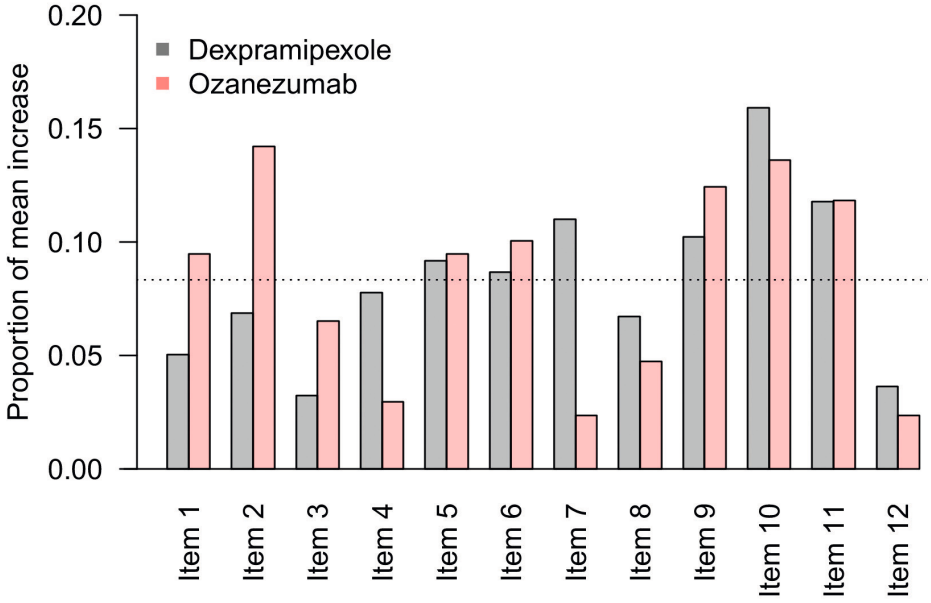
Characteristic	Dexpramipexole trial			Ozanezumab trial		
	Patients without a ≥5 point increase *(N = 814)*	Patients with a ≥5 point increase *(N = 123)*	P-value	Patients without a ≥5 point increase *(N = 275)*	Patients with a ≥5 point increase *(N = 28)*	P-value
Age at screening, years	57 (11)	58 (12)	0.150	55 (11)	58 (12)	0.33
Sex, female	280 (34%)	55 (45%)	0.034	90 (33%)	13 (46%)	0.21
Site of symptom onset, bulbar	174 (21%)	43 (35%)	0.001	52 (19%)	13 (46%)	0.002
Symptom duration, months	15 (5)	16 (5)	0.37	18 (7)	18 (6)	0.63
ΔFRS, points per month	-0.72 (0.57)	-0.86 (0.54)	0.011	-0.59 (0.38)	-0.82 (0.38)	0.003
ALSFRS-R total score (max. 48)	38 (5)	36 (5)	< 0.001	38 (5)	35 (5)	< 0.001
Bulbar (max. 12)	10 (2)	9 (2)	< 0.001	10 (2)	9 (3)	0.003
Fine (max. 12)	9 (3)	8 (3)	0.073	8 (3)	8 (3)	0.39
Gross (max. 12)	8 (3)	8 (3)	0.34	8 (3)	7 (2)	0.135
Respiratory (max. 12)	11 (1)	11 (2)	0.030	12 (1)	11 (2)	0.062
SVC, %predicted	89 (17)	89 (18)	0.79	95 (18)	91 (19)	0.27

Data are expressed as mean (SD) or number (%). Abbreviations: ALSFRS-R, ALS Functional Rating Scale-Revised; SVC, slow vital capacity. ΔFRS = mean change score of the amyotrophic lateral sclerosis functional rating scale revised. The purpose of the reported p-values is solely explorative, to flag potential differences between patient populations.

Item characteristics associated with potential measurement errors

Finally, in **Figure 4** we show the average change scores for each individual ALRFRS-R item in which no sudden increase occurred, averaged over the measurements that were flagged as a 5-point or more increase. Despite differences in observed frequency of sudden increases between the dexpramipexole and ozanezumab trials, items 2, 9, 10, and 11 were consistently identified as the primary drivers of a 5-point or more increase in ALSFRS-R total scores; this may reflect potential scoring difficulties for these questions.

Figure 4. Contribution of individual items to a sudden 5-point or more increase



Contribution of individual items to a sudden 5-point or more increase. Mean, proportional contribution of individual ALSFRS-R items to a sudden 5-point or more increase. For example, in the dexpramipexole trial, there were 123 patients with a 5- point or more increase during follow-up with a mean increase in total score of 6.3 points, of which 0.3 points (5%) were due to an increase in Item 1. If each item was equally responsible for the mean increase in total score, one would expect that each item would be accountable for 1/12 (8.3%, dashed line).

Discussion

In this study, we show the relatively common occurrence of sudden, large increases in ALSFRS-R total scores between two consecutive monthly measurements. These sudden increases did not occur with equal frequency in trial centres, which underscores the potential presence of an external cause, such as a difference in scoring strategies among trial centres, or dissimilarities between raters. We found several patient- and item characteristics that were associated with the prevalence of sudden increases. Identified patients characteristics were bulbar onset, faster disease progression at enrolment, and a lower ALSFRS-R score at baseline. ALSFRS-R items that were associated sudden increases were items 2 (saliva), 9 (stairs), 10 (dyspnoea), and 11 (orthopnoea). Given the identified patient- and item related characteristics, an important source of these sudden increases may be related to the initiation of symptomatic interventions, especially for respiratory and bulbar symptomatology. Although the study staff of both clinical trials were well-trained, our results indicate

that the current SOPs may leave room for improvement and highlight the potential benefit of real-time monitoring of data quality to ensure SOP conformity.

The unnatural, large, sudden increases in the ALSFRS-R total score, and especially the imbalance in distribution among trial centres, suggest that these increases are most likely the result of a limitation in the ALSFRS-R itself rather than due to a biological mechanism. Given that several items in the bulbar and respiratory domains were marked as important drivers for sudden increases. The vulnerability of these two domains to sudden increases is further substantiated by the finding that the increases occurred significantly more often in patients with low respiratory and bulbar scores at baseline. Especially for these domains several symptomatic treatment options are available, a main challenge for the ALSFRS-R scoring is how to handle symptomatic interventions. The ALSFRS-R was developed to monitor disease progression and to quantify the efficacy of experimental treatments in clinical trials. From a clinical trial perspective, therefore, it would be preferable if an improvement in ALSFRS-R score resulted only from an experimental intervention. It is thus important to separate the effect caused by a potential symptomatic intervention from the effect caused by the experimental intervention; furthermore, initiation of a symptomatic intervention should reflect natural disease progression.

The initiation of a symptomatic treatment might not only cause a sudden increase in ALSFRS-R (e.g. salivation therapy improving the patient condition from severe drooling to no salivary excess), but may also be a reflection of day-to-day variation in the patient's symptomatology. In this study, we did not look at small sudden changes in the ALSFRS-R items and subdomains, but given that the random variation for the subdomains ranges from 1.6 to 2.4 points (8), a few items conincidently improving between two consecutive visits could also result in a 5-point or more increase. Just as with the symptomatic interventions, the effect of these natural improvements should be minimized, so that the natural trajectory of the ALSFRS-R becomes a uniformly declining function over time. This highlights not only the importance of facilitating uniform scoring strategies, but also of continuously evaluating the accuracy of the ALSFRS-R items (17, 18). A targeted adjustment of the ALSFRS-R SOP might be justified. To ensure broad consensus, this requires a collaborative effort between large ALS trial networks, such as the Northeast ALS Consortium (NEALS) (16), Trial Research Initiative to Cure ALS (TRICALS) (19), and the Motor Neurone Disease group Australia (20).

Although sudden increases might be related to limitations in the ALSFRS-R, underperformance of individual trial centres may play a role. By calculating the probability of the proportion of sudden increases, we were able to get an impression of which individual centres could have been flagged with suspected underperformance during trial conduct. The results demonstrate that the number and the degree of deviation of the outlier centres was higher in the dexpramipexole trial, compared to the ozanezumab trial. These differences could very well be due to an improvement in training and refined standardization of the ALSFRS-R scoring strategies, as the ozanezumab study was conducted four years after the dexpramipexole study (in particular in case of overlapping sites or raters). However, the ozanezumab study additionally employed a central in-stream blinded monitoring system during the study, to identify outlier efficacy data values at patient or site level triggering data queries to the sites. Interestingly, we found that high proportions of sudden increases in trial centres of the dexpramipexole trial occurred more often in centres with a low number of enrolled patients. This finding is consistent with existing literature that points out that factors such as reaching enrolment goals may be related to centre-related performance in data quality, highlighting the importance of recognising centres of excellence via disease networks (17, 21).

Our study has several limitations. First, a true improvement cannot be entirely ruled out in individual cases (10). For example, dietary supplements or other experimental treatments may have led to a real improvement in function (11). Second, although our analysis indicated that the number of enrolled patients per trial centre was an explanatory factor for the occurrence of sudden increases, the available data did not allow us to analyse other centre characteristics that were potentially associated with sudden increases, such as previous trial experience. However, our results, supported by previous literature, indicate that preliminary selection and interim assessment of participating trial centres, could potentially contribute to improvement of data quality (17). Finally, the influence of different raters for the same patient, and the influence of unknown placebo effects, as a source of unwanted variability could not be estimated. However, longitudinal scoring by the same rater, possibly supported by video review (22) of expert raters, could contribute to optimising data quality. Since adjusted SOPs cannot prevent all sources of variation, for example inadequate training of raters, video review and other methods for monitoring of data quality (including real-time monitoring) are likely to be of important added value.

In conclusion, the results of this study suggest that sudden increases in consecutive ALSFRS-R total scores occur relatively frequently in multicentre studies. We found that these sudden increases did not occur with equal frequency in trial centres. In addition, multiple ALSFRS-R items were related to sudden increases, especially score for the items in the bulbar and respiratory domains, which can be impacted by available symptomatic treatments. Patients with a bulbar onset, a low ALSFRS-R baseline score and a faster disease progression were more likely to have a sudden increase. To facilitate adequate and uniform handling of improvements due to symptomatic treatment, a targeted adjustment of the SOP, and corresponding skill-training is warranted. In addition, multicentre research could benefit from methodology to monitor for data quality, as well as interim video reviews by expert raters.

References

1. van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, et al. Amyotrophic lateral sclerosis. Lancet. 2017;390(10107):2084-98.
2. van Eijk RPA, Kliest T, van den Berg LH. Current trends in the clinical trial landscape for amyotrophic lateral sclerosis. Curr Opin Neurol. 2020;33(5):655-61.
3. Food Drug Administration Center for Drugs Evaluation Research. Guidance for industry: amyotrophic lateral sclerosis: developing drugs for treatment 2019 [Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/amyotrophic-lateral-sclerosis-developing-drugs-treatment-guidance-industry.
4. European Medicines Agency. Guideline on clinical investigation of medicinal products for the treatment of amyotrophic lateral sclerosis (ALS) 2016 [Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-amyotrophic-lateral-sclerosis_en.pdf.
5. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J Neurol Sci. 1999;169(1-2):13-21.
6. Gordon PH, Miller RG, Moore DH. Alsfrs-R. Amyotroph Lateral Scler Other Motor Neuron Disord. 2004;5 Suppl 1:90-3.
7. Kimura F, Fujimura C, Ishida S, Nakajima H, Furutama D, Uehara H, et al. Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. Neurology. 2006;66(2):265-7.
8. Bakker LA, Schroder CD, Tan HHG, Vugts S, van Eijk RPA, van Es MA, et al. Development and assessment of the inter-rater and intra-rater reproducibility of a self-administration version of the ALSFRS-R. J Neurol Neurosurg Psychiatry. 2020;91(1):75-81.
9. Franchignoni F, Mandrioli J, Giordano A, Ferro S, Group E. A further Rasch study confirms that ALSFRS-R does not conform to fundamental measurement requirements. Amyotroph Lateral Scler Frontotemporal Degener. 2015;16(5-6):331-7.
10. Bedlack RS, Vaughan T, Wicks P, Heywood J, Sinani E, Selsov R, et al. How common are ALS plateaus and reversals? Neurology. 2016;86(9):808-12.
11. Harrison D, Mehta P, van Es MA, Stommel E, Drory VE, Nefussy B, et al. "ALS reversals": demographics, disease characteristics, treatments, and co-morbidities. Amyotroph Lateral Scler Frontotemporal Degener. 2018;19(7-8):495-9.
12. Vasta R, D'Ovidio F, Canosa A, Manera U, Torrieri MC, Grassano M, et al. Plateaus in amyotrophic lateral sclerosis progression: results from a population-based cohort. Eur J Neurol. 2020;27(8):1397-404.
13. Cudkowicz ME, van den Berg LH, Shefner JM, Mitsumoto H, Mora JS, Ludolph A, et al. Dexpramipexole versus placebo for patients with amyotrophic lateral sclerosis (EMPOWER): a randomised, double-blind, phase 3 trial. Lancet Neurol. 2013;12(11):1059-67.
14. Berry JD, Miller R, Moore DH, Cudkowicz ME, van den Berg LH, Kerr DA, et al. The Combined Assessment of Function and Survival (CAFS): a new endpoint for ALS clinical trials. Amyotroph Lateral Scler Frontotemporal Degener. 2013;14(3):162-8.
15. Meininger V, Genge A, van den Berg LH, Robberecht W, Ludolph A, Chio A, et al. Safety and efficacy of ozanezumab in patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled, phase 2 trial. Lancet Neurol. 2017;16(3):208-16.
16. Northeast Amyotrophic Lateral Sclerosis Consortium. [Available from: https://www.neals.org/.
17. Dombernowsky T, Haedersdal M, Lassen U, Thomsen SF. Criteria for site selection in industry-sponsored clinical trials: a survey among decision-makers in biopharmaceutical companies and clinical research organizations. Trials. 2019;20(1):708.

18. Walker R, Morris DW, Greer TL, Trivedi MH. Research staff training in a multisite randomized clinical trial: Methods and recommendations from the Stimulant Reduction Intervention using Dosed Exercise (STRIDE) trial. Addict Res Theory. 2014;22(5):407-15.

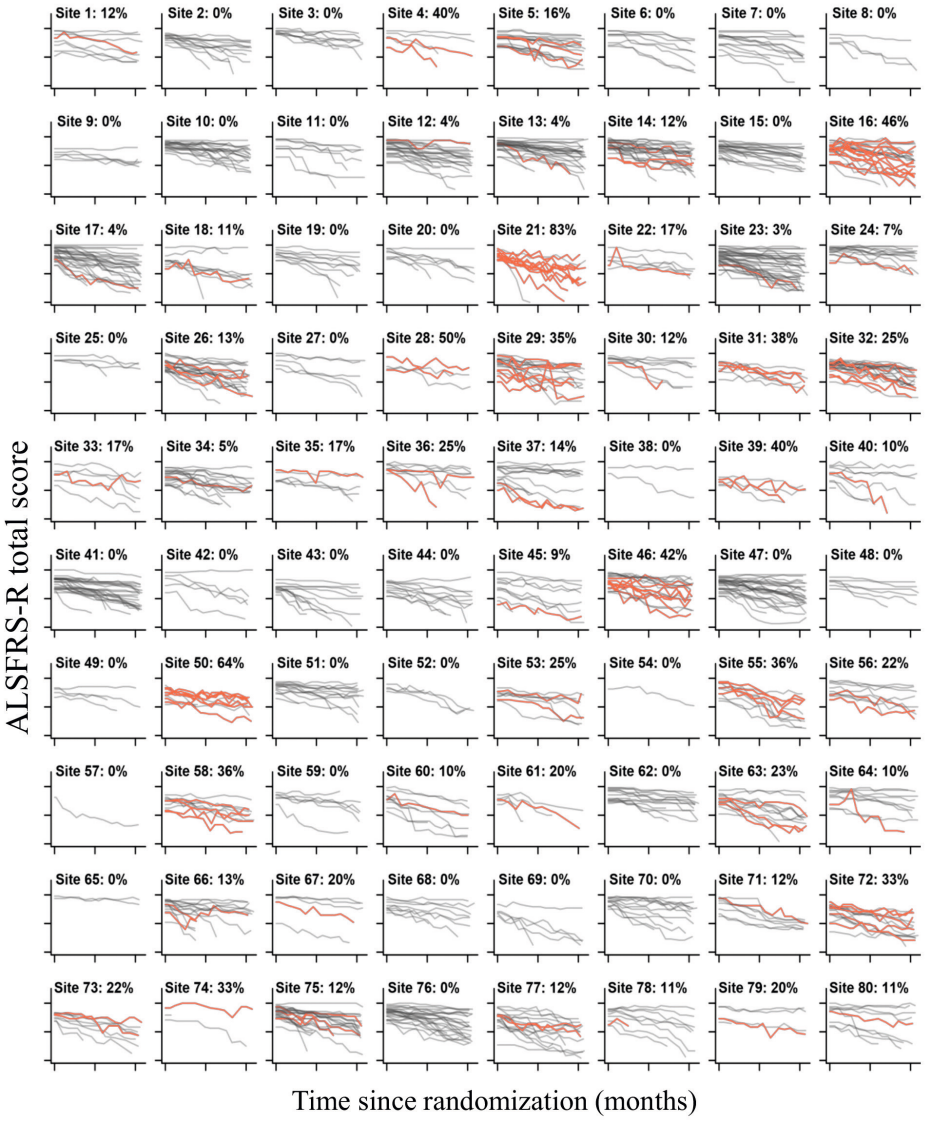
19. van Eijk RPA, Kliest T, McDermott CJ, Roes KCB, Van Damme P, Chio A, et al. TRICALS: creating a highway toward a cure. Amyotroph Lateral Scler Frontotemporal Degener. 2020;21(7-8):496-501.

20. Motor Neurone Disease group Australia [Available from: https://www.sydney.edu.au/brain-mind/our-research/forefront-ageing-and-neurodegeneration/motor-neurone-disease.html.

21. Gehring M, Taylor RS, Mellody M, Casteels B, Piazzi A, Gensini G, et al. Factors influencing clinical trial site selection in Europe: the Survey of Attitudes towards Trial sites in Europe (the SAT-EU Study). BMJ Open. 2013;3(11):e002957.

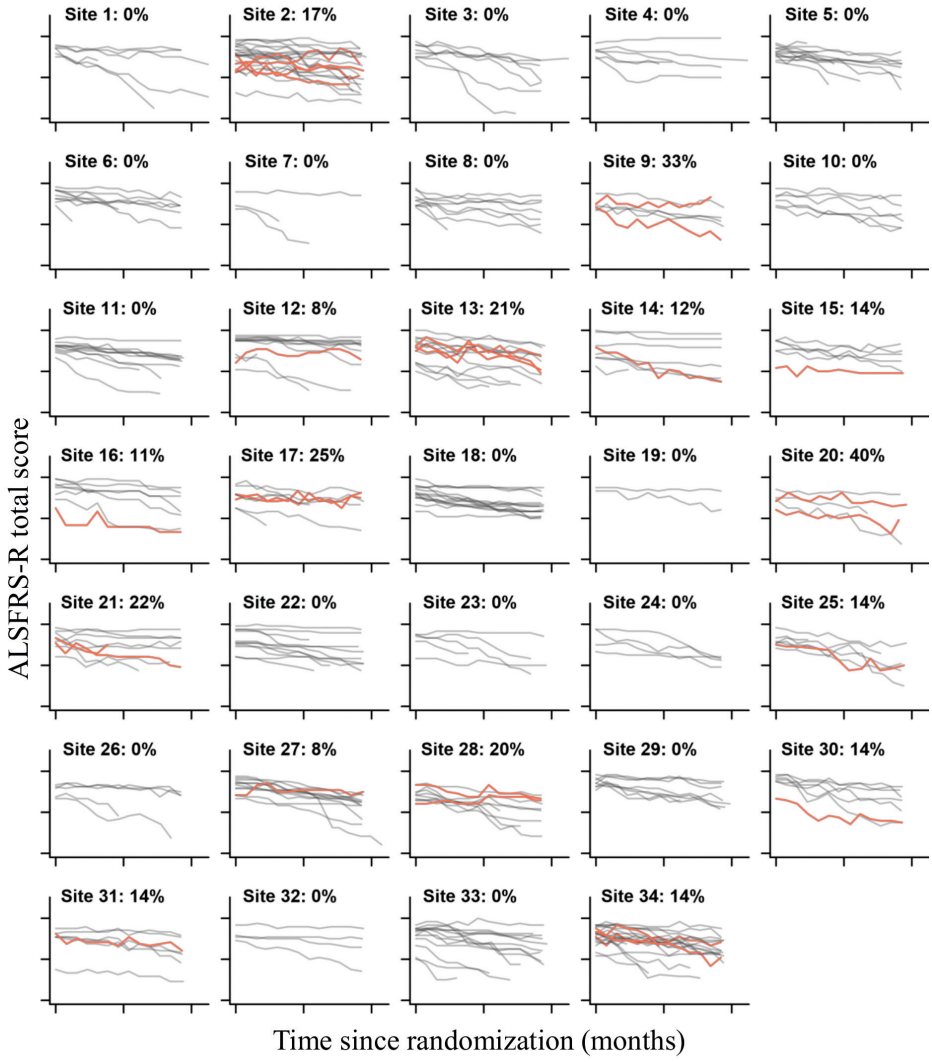
22. Sutherland AE, Stickland J, Wee B. Can video consultations replace face-to-face interviews? Palliative medicine and the Covid-19 pandemic: rapid review. BMJ Support Palliat Care. 2020;10(3):271-5.

Figure e1. Oversight of the prevalence of a 5-point or greater increase in ALSFRS-R total score for all participating centres of the *dexpramipexole* study

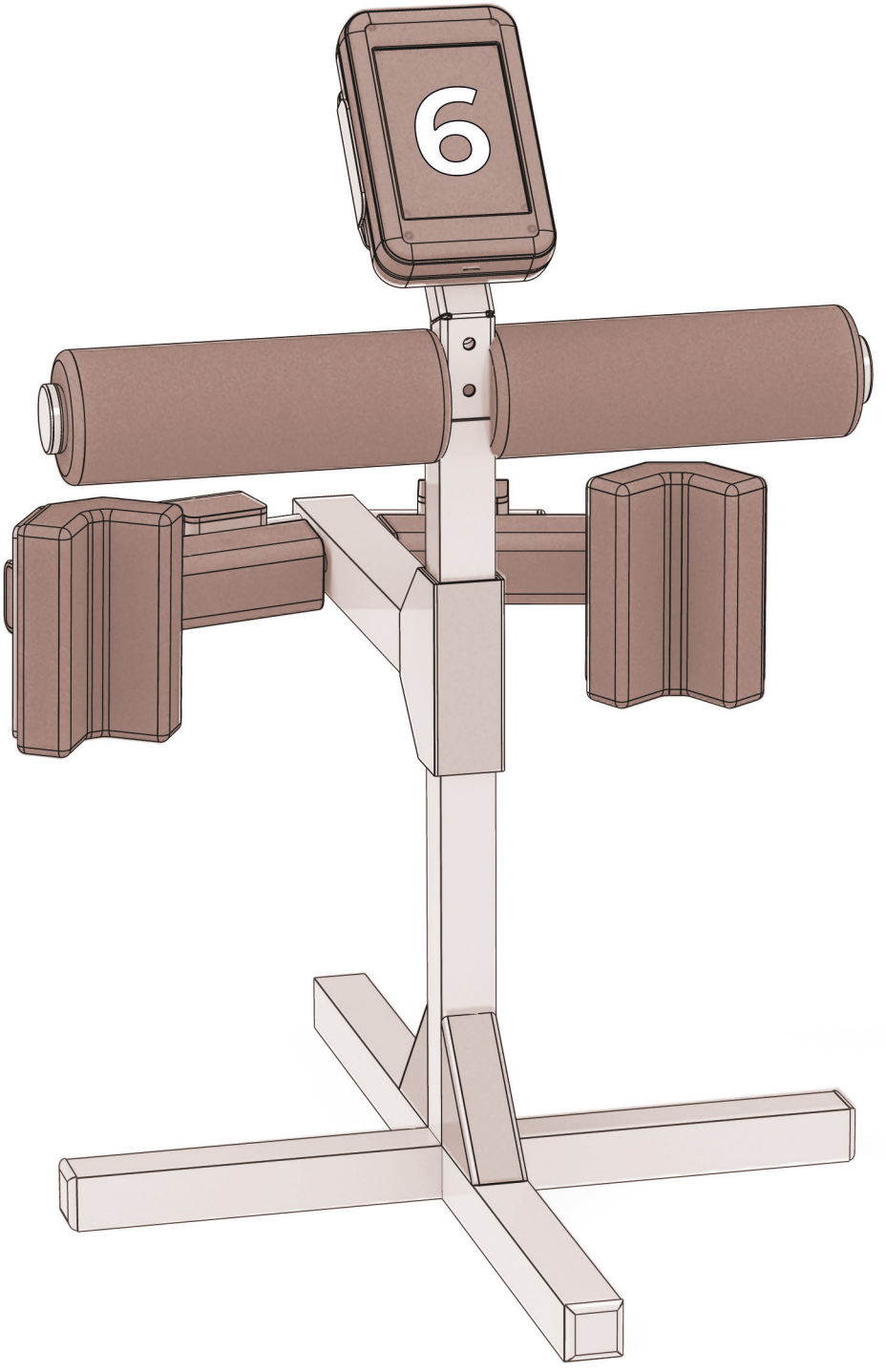


Per centre, the number of patients with a 5-point or more increase in ALSFRS-R total score are highlighted in *red*.

Figure e2. Oversight of the prevalence of a 5-point or greater increase in ALSFRS-R total score for all participating centres of the *ozanezumab* study.



Per centre, the number of patients with a 5-point or more increase in ALSFRS-R total score are highlighted in *red*.


6

Chapter 6

IMPLICATIONS OF SPIROMETRIC REFERENCE VALUES FOR AMYOTROPHIC LATERAL SCLEROSIS

Ruben van Eijk
Jaap Bakers
Michael van Es
Marinus Eijkemans
Leonard van den Berg

Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration 2019

Abstract

Objective: Spirometry is commonly used as screening tool for respiratory insufficiency in neuromuscular diseases. Despite the well-known effects of reference standards on spirometric outcomes, its standardization is overlooked in current guidelines. We aim to illustrate the effect of spirometric reference values on prognostication, medical decision-making and trial eligibility in the applied setting of amyotrophic lateral sclerosis (ALS).

Methods: We selected 4,651 patients with 32,022 FVC measurements from the PRO-ACT dataset. The FVC estimates were standardized according to five reference standards: Knudson '76, Knudson '83, ECSC, NHANES III and GLI-2012. (Generalized) linear mixed-effects and Cox proportional hazard models were used to evaluate longitudinal patterns and time-to-event outcomes.

Results: The mean population %predicted FVC varied between 78.5% (95% CI 78.0-79.1) and 88.5% (95% CI 87.9-89.1). The unstandardized litres provided the worst fit on the survival data (AIC 20573, c-index 0.760), whereas the GLI provided the best fit (AIC 20374, c-index 0.780, $p < 0.001$). The mean population rate of decline in %predicted FVC could vary as much as 11.4% between reference standards. The median time-to-50% predicted FVC differed by 2.9 months between recent (14.5 months, 95% CI 14.4 – 16.1) and early reference standards (17.4 months, 95% CI 16.1 – 18.2).

Conclusions: Independent of technique, device or evaluator, spirometric reference values affect the utility of spirometry in ALS. Standardization of reference values is of the utmost importance to optimize clinical decision-making, improve prognostication, enhance between-centre comparison and unify patient selection for clinical trials.

Introduction

Reduced pulmonary function is a common feature in COPD,[1] heart failure,[2] and neuromuscular disorders.[3] Spirometric reference standards are used to compare a patient's pulmonary function to a healthy population of a similar age, body height, sex and ethnicity.[4] To illustrate, an absolute forced vital capacity (FVC) of 3.84L could be either 90% or 67% from normal for two patients with a predicted FVC of 4.25L or 5.75L, respectively, given their morphological characteristics. Absolute spirometric measurements are, therefore, of little value in clinical practice. Importantly, the reference value for an individual patient depends heavily on the applied reference standard.[5] As a consequence, variability between reference standards may considerably affect the patients' spirometric results and alter clinical decisions.[6]

Spirometry is commonly used as screening tool for respiratory insufficiency in progressive neuromuscular diseases such as amyotrophic lateral sclerosis (ALS).[7] The end-stage of ALS is characterized by weakening of the diaphragm musculature, which results in a progressive decline in lung function and finally death. Spirometry plays, therefore, a central role in the clinical care for patients with ALS and is a primary predictor of survival time.[8] Specific cut-off values for spirometric measures are used to guide the timing of non-invasive ventilation (NIV),[7] as decision-tool for gastrostomy placement or as eligibility criterion for clinical trial participation.[9]

Despite the significant efforts to standardize the operational procedures of spirometry, standardization of references values is overlooked in the current guidelines for ALS (e.g. the NICE 2016,[10] AAN 2009[11] and EFNS 2011[12]). The variability in predictions between spirometric reference standards is, nevertheless, well known.[5] It is, therefore, surprising that there is currently no guidance for spirometric reference standards in ALS. The lack of standardization of reference values could have implications for patients and ALS-oriented research settings. This study, therefore, aims to illustrate the effect of spirometric reference values on prognostication, medical-decision making and trial eligibility, and ultimately provides a basis for standardization of spirometric reference values in ALS.

Methods

Study population

Data for this study originated from the open-access PRO-ACT database (version Dec. 2015). PRO-ACT contains data for 10,731 individuals from 23 ALS clinical trials

performed over the past 20 years, is IRB-approved and uses solely anonymised data; individual trials within PRO-ACT cannot be traced.[13] All subjects provided their consent during trial participation. In order to assess the effect of different spirometric reference standards on clinical endpoints, we excluded those patients without follow-up time or missing demographic data regarding sex, age, height or ethnicity. For each patient we extracted the FVC values in litres.

Study design, outcomes and variables

We retrospectively determined the predicted FVC in litres according to five reference standards: Knudson '76,[14] Knudson '83,[15] European Community for Steel and Coal (ECSC),[16] National Health and Nutrition Examination Survey (NHANES) III,[17] and GLI-2012.[4] The Knudson '76, '83 and ECSC reference standards are still frequently used in ALS clinics (*personal communication*) and ALS clinical trials (e.g. the Ceftriaxone trial used Knudson '83,[18] whereas the Xaliproden trials were using ECSC),[19] despite their known disadvantages.[5] The GLI-2012, the most recent standard and based on >74,000 global control subjects,[4] has been endorsed by both the American Thoracic Society (ATS) and European Respiratory Society (ERS) and will probably supersede previous reference standards.[5] The observed FVC was standardized according to the prediction per reference standard and expressed as percentage from normal (%predicted FVC). Survival time was defined as the time from trial inclusion until death from any cause. Patients who remained alive during the trial were censored after their last follow-up visit. The severity of dyspnoea symptoms was assessed using item 10 of the revised ALS functional rating scale (ALSFRS-R).

Statistical analysis

We used linear mixed effects models (LME) to evaluate the longitudinal patterns of decline in FVC, as described elsewhere.[20] Using the longitudinal sample size framework provided by Edland, we calculated the number of patients required to detect a 25% reduction in the rate of decline with 80% power after 18 months of follow-up and a quarterly visiting scheme.[20] Subsequently, we analysed the predictive value of each reference standard in predicting survival time using Cox proportional hazard (PH) models. Predictive performance and model fit were evaluated with the Concordance statistic (C-statistic) and Akaike Information Criterion (AIC), respectively. All Cox PH models were adjusted for the following predictors: age at randomization, treatment arm, ΔFRS (ALSFRS-R at randomization – 48 / symptom duration), Body Mass Index (BMI), site of symptom onset and diagnostic delay.[8, 21] Missing data in any of the covariates were handled by creating multiple imputed datasets, a procedure described in more detail elsewhere.[9] Finally,

we calculated for each individual the time to reach a specific %predicted FVC cut-off value (e.g. < 80%). Kaplan-Meier curves were used to calculate the median time to reach the defined cut-off value. In addition, a generalized (logistic) LME was used to assess the longitudinal correlation between symptoms of dyspnoea and the probability of obtaining a %predicted FVC below the cut-off value. The GLI-2012 predicted FVC values were calculated using the R package *rspiro* (version 0.1, Lytras T, 2017).[4] (Generalized) LME models were fitted using the *(g)lmer* function (lme4, version 1.1-18-1).[22]

Results

In total, complete baseline information was available for 4,651 patients; their baseline characteristics are given in **Table 7.1**. The average number of longitudinal FVC (L) measurements was 6.9 (total 33,296 measurements); the total follow-up time was 4,701 person-years, during which 1398 deaths occurred (12-month survival since trial enrolment of 76.1% [95% CI 74.8% - 77.5%]). Depending on which reference standard was applied, the mean population %predicted FVC varied between 78.5% (95% CI 78.0-79.1) and 88.5% (95% CI 87.9-89.1); mean difference of 10.0% (95% CI 9.8-10.1, *p*-value < 0.001). **Figure 7.1** provides the individual differences between reference standards as measure of disagreement, which revealed a clear difference between males and females: the median disagreement for males was 8.4% (IQR 3.3%) *vs.* 14.3% (IQR 5.1%) for females (*p*-value < 0.001). The mean disagreement between reference standards, irrespective of gender, increased by 1.57% (95% CI 1.50-1.64) per 10% increase in %predicted FVC (*p*-value < 0.001).

Table 7.1. Characteristics of patients at baseline

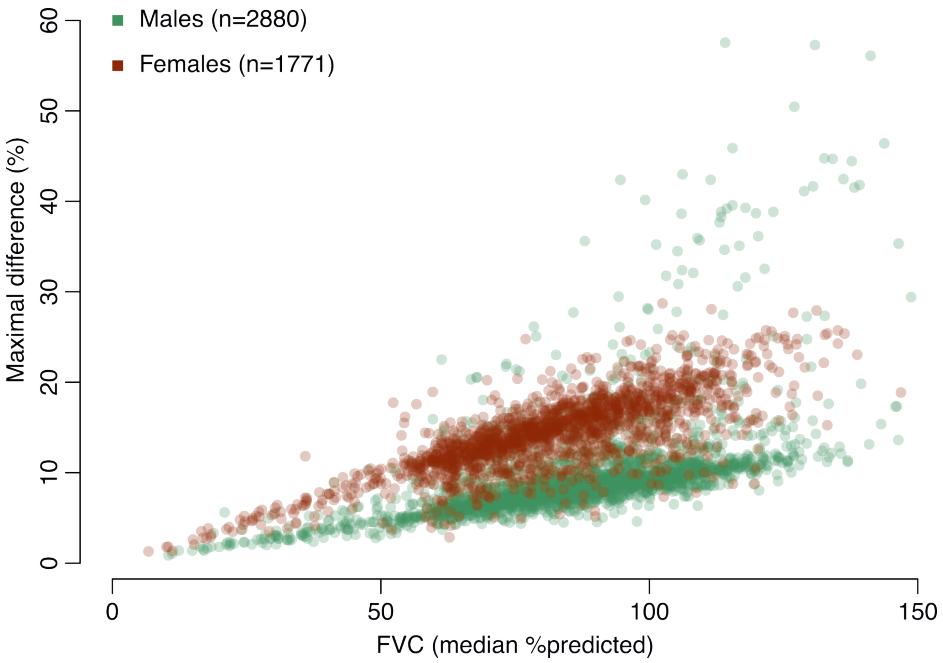
	PRO-ACT (N = 4651)
Age at enrolment, *years*	56 (12)
Sex Male Female	62% (2880) 38% (1771)
Race Caucasian African-American Other	97.6% (4540) 1.8% (82) 0.6% (29)
Site of symptom onset Bulbar Other	21% (982) 79% (3667)
Symptom duration, *months*	24 (18)

Table 7.1. *(continued)*

	PRO-ACT (N = 4651)
Diagnostic delay, *months*	12 (9)
Body Mass Index (BMI), *kg/m²*	25.8 (4.7)
Forced vital capacity, *litres*	3.30 (1.12)
Overall	3.82 (1.01)
Males	2.47 (0.71)
Females	
Forced vital capacity, *%predicted*	86.2% (21.2)
Knudson 1976	85.4% (22.1)
Knudson 1983	88.5% (21.4)
ECSC 1993	78.5% (19.1)
NHANES 1999	80.4% (19.4)
GLI 2012	

Data are given in mean (SD) or % (n).

Figure 7.1. Disagreement in %predicted FVC between reference standards at baseline



For each patient we calculated the maximal difference in %predicted FVC between two reference standards as measure of disagreement. The median %predicted is the median value of the five reference standards.

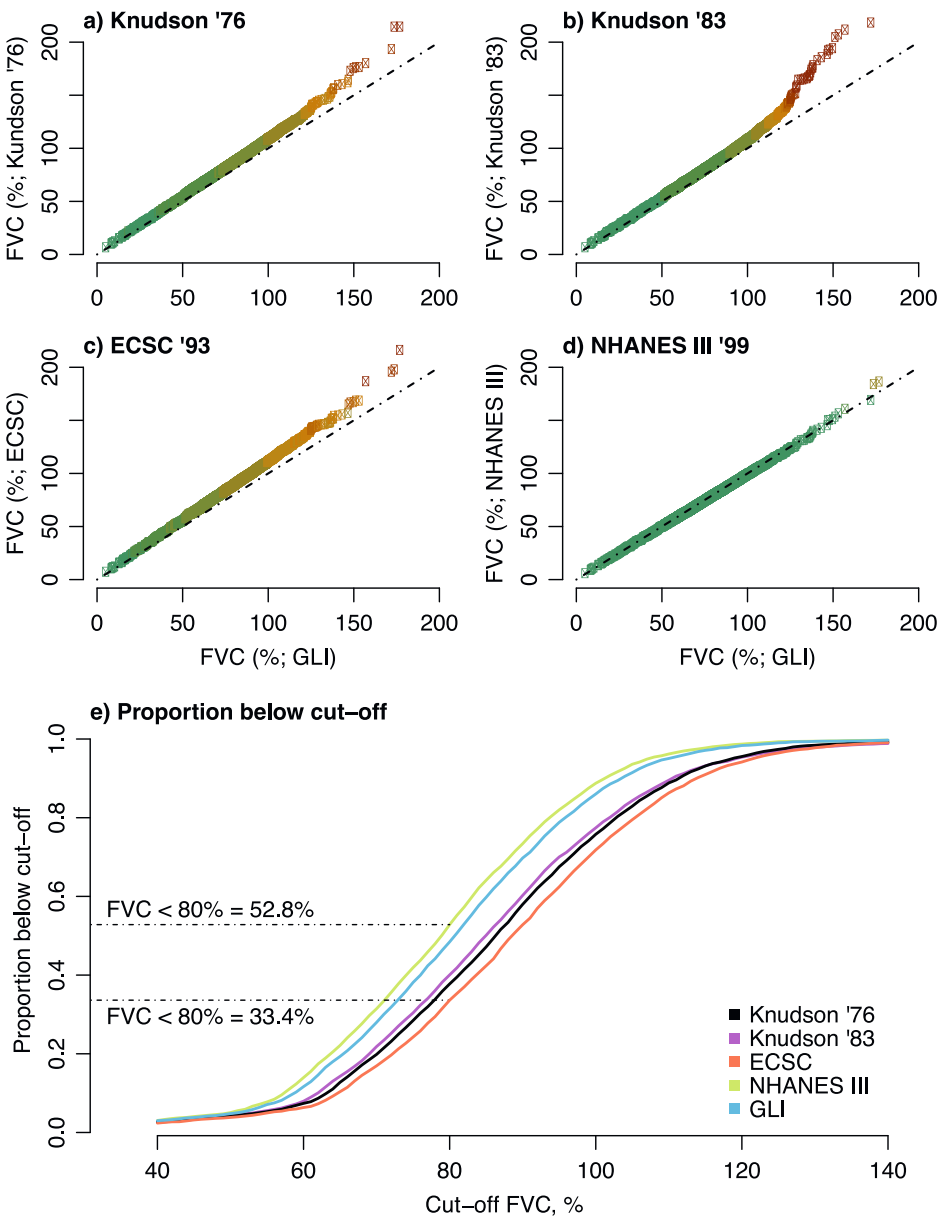
The NHANES III and GLI reference standards had a strong correlation (Pearson *r* 0.997, *p*-value < 0.001), whereas older standards showed a larger deviation from the GLI

(**Figure 7.2A-D**). As non-invasive ventilation (NIV) referral or trial eligibility is often based on a specific cut-off [9], the cumulative proportion of patients below a certain cut-off is given in **Figure 7.2E**. When trial eligibility is based on, for example, an 80% FVC, 52.8% and 33.6% of the PRO-ACT cohort would be ineligible under NHANES III and ECSC, respectively. Interestingly, the ineligible population (i.e. FVC < 80%) under NHANES III has a 12-month survival of 65.1% (95% CI 63.1% - 67.2%); under ECSC it is 58.2% (95% CI 55.6% - 60.9%, $p < 0.001$ for survival difference). This suggests that, depending on the reference standard, a different population is selected, which could affect the generalizability and between-trial comparability.

The effect of each reference standard on the predictive performance for survival time and variability over time are given in **Tables 7.2** & **7.3**. The unstandardized litres provided the worst fit on the survival data (AIC 20573, c-index 0.760), whereas the GLI provided the best fit (AIC 20374, c-index 0.780, $p < 0.001$). In the longitudinal data, NHANES III exhibited the lowest between-patient variability over time (2.04%) and would be the most sensitive measure to detect a given treatment effect (**Table 7.3**). Interestingly, the mean monthly rate of decline in % FVC could vary as much as 11.4% between reference standards (i.e. -2.81% ECSC *vs.* -2.49% NHANES III).

Finally, we estimated the median time-to-referral for non-invasive ventilation defined as predicted FVC < 50% or < 80% (**Figure 7.3A+C**).[23, 24] Overall, results for Knudson '76 & '83 were similar to the ECSC (*results not shown*). The median time-to-50% predicted FVC was the shortest when using NHANES III (14.5 months, 95% CI 14.4 – 16.1), which was 2.9 months earlier than the ECSC (17.4 months, 95% CI 16.1 – 18.2). This difference was 2.1 months (95% CI: 0.2 – 3.0) for males and 3.3 months (95% CI: 2.0 – 4.6) for females. A similar pattern was seen when applying the 80% cut-off, with a median time of 8.1 (95% CI 7.8 - 8.4) for NHANES III and 12.8 (95% CI 12.3 – 14.3) for ECSC. Overall, the difference between GLI and NHANES III was minimal. **Figures 7.3B+D** reveal the association between symptoms of dyspnoea and proportion of FVC measurements less than 50% or 80%. For the 50% cut-off, symptoms have an approximate linear relationship with the proportion of patients failing the threshold with a minimal difference between standards. Nevertheless, 15.4% - 23.2% of the patients with severe symptomatology (always dyspnoeic) still have a %predicted FVC above 50%. Interestingly, 72.4% and 60.9% of the patients under NHANES and GLI, respectively, fail the 80% cut-off without any symptomatology, whereas this is only 26.2% for de ECSC. Overall, results for Knudson '76 & '83 were similar to the ECSC (*results not shown*).

Figure 7.2. Quantile-Quantile plots of the different reference standards



Quantile-Quantile plots of each reference standard with the GLI as reference distribution. Ideally, two reference standards would provide similar estimates, resulting in a straight line (dashed-line). The colours represent the deviation from the ideal line (green less than 2.5% deviation, red more than 25% deviation). (**E)** Cumulative proportions of patients below FVC cut-off values in PRO-ACT.

Table 7.2. Multivariate Cox proportional hazard models per reference standard

Model	**AIC**	**C-statistic**	**Hazard ratio**
Unadjusted litres	20573	0.760	-
Knudson (1976)	20393	0.778	0.74 (0.72 - 0.76)
Knudson (1983)	20428	0.774	0.75 (0.72 - 0.77)
ECSC (1993)	20383	0.780	0.74 (0.72 - 0.76)
NHANES III (1999)	20380	0.779	0.71 (0.69 - 0.74)
GLI (2012)	20374	0.780	0.72 (0.69 - 0.74)

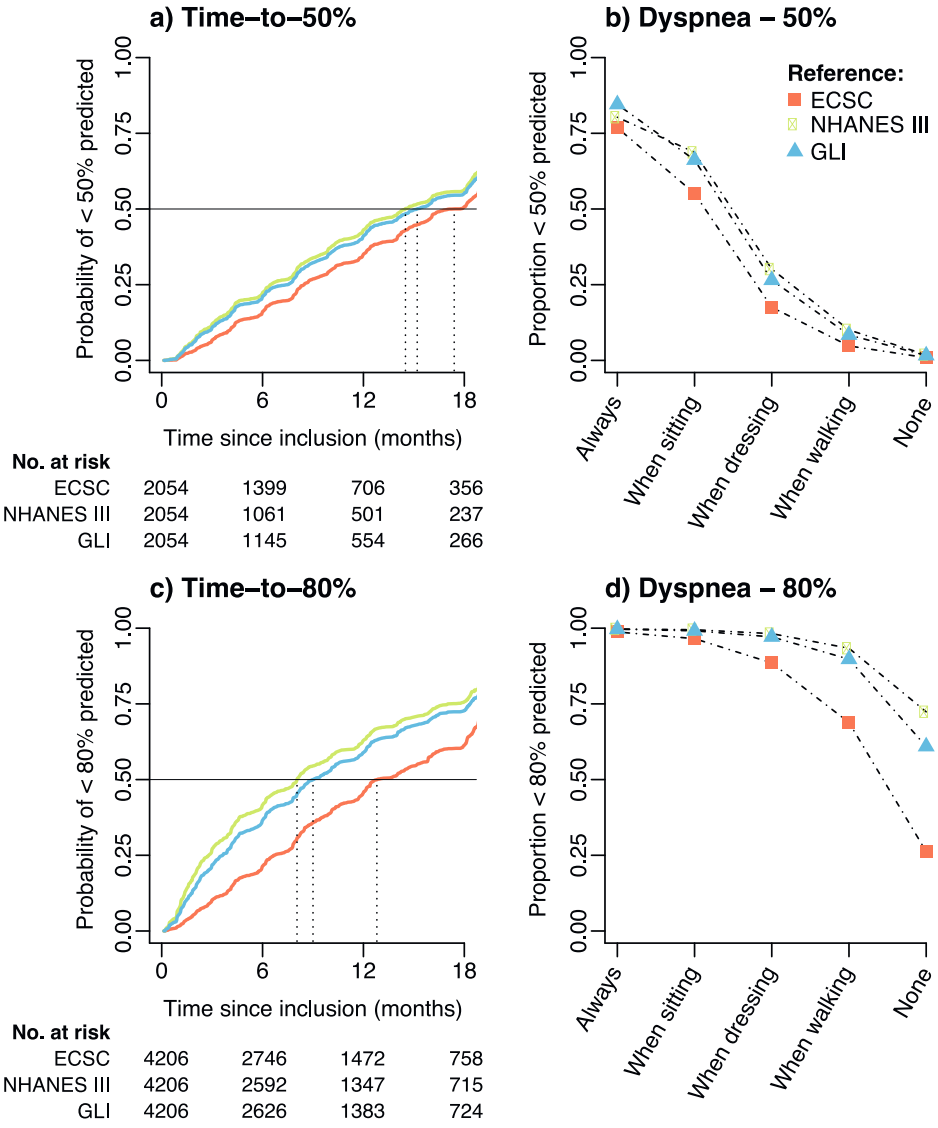
Models were corrected for age, treatment allocation PRO-ACT, ALSFRS-R slope, BMI, site of onset and diagnostic delay. Estimated hazard ratios are per 10% increase in predicted FVC.
AIC = Akaike Information Criterion; C-statistic = concordance statistic.

Table 7.3. Longitudinal linear mixed models per reference standard

Model	Baseline	Baseline variance	Slope (*Monthly*)	Slope variance	Required sample size
Unadjusted litres	3.31 L	1.14 L	-0.10 L	0.09 L	370
Knudson (1976)	86.38%	21.23%	-2.73%	2.25%	356
Knudson (1983)	85.57%	22.12%	-2.69%	2.22%	358
ECSC (1993)	88.70%	21.49%	-2.81%	2.32%	358
NHANES III (1999)	78.71%	19.17%	-2.49%	2.04%	352
GLI (2012)	80.58%	19.49%	-2.55%	2.09%	354

Sample size calculations were based on a 25% reduction in the rate of decline in %FVC after 18 months (quarterly follow-up) with 80% power and a two-sided alpha of 5%.

Figure 7.3. Kaplan-Meier curves of referral criteria and the association with symptoms of dyspnoea



In red the ECSC, in green the NHANES III and in blue the GLI. Time-to-50% or 80% is defined as the time from trial inclusion to the first FVC measurement below 50% or 80% predicted, estimated in those patients whose FVC was higher than 50% or 80% at inclusion (N = 4,206 and N = 2,054, respectively). **(B+D)** Generalized linear mixed-effects model (logistic) with FVC dichotomized (<50% vs. ≥50% or <80% vs. ≥80%) as function of ALSFRS item 10 (Dyspnoea) in 4,528 patients with 32,955 matched ALSFRS – FVC data-points. Overall, results for Knudson '76 & '83 were similar to the ECSC (results not shown).

Discussion

In this study, we assessed the impact of spirometric reference standards on clinical and research-oriented settings for patients with ALS. Independent of technique, device or evaluator, there are important differences between reference standards that could considerably affect medical decision-making, prognostication, patient counselling and patient selection for clinical trials. The timing of referral to a respiratory care clinic is affected by the applied reference values. Moreover, the comparability of clinical research is obscured when investigators use different standards for study eligibility or when treatment efficacy is based on the respiratory rate of decline. Standardizing the use of spirometric reference standards in both clinical and research-oriented settings is, therefore, of the utmost importance to optimize the utility of spirometry.

Despite the known differences between spirometric reference standards,[5] standardization of spirometric reference standards receives surprisingly little attention in ALS. Our results not only scrutinize the differences between reference standards, but also reveal how the variability in predicted FVC could considerably affect real-world decisions. To illustrate: a female patient (62-year-old, 1.68m, Caucasian, FVC 2.43L) could obtain an 82.5% predicted FVC at the neurologist (using ECSC), whereas this could be 69.0% at the physiotherapist (using NHANES III). This sudden 13.5% drop may erroneously lead to activation of additional care and cause unnecessary psychological distress. In addition, the %predicted FVC is part of a prognostic tool used for patient counselling (8). In this case, a 13.5% difference in FVC (HR 0.99), while keeping other factors constant, falsely predicts a 14.5% increased risk of death during follow-up. Importantly, this difference is solely caused by a systematic difference in reference standards and not due to a true difference in pulmonary function. Between-technician differences and other technical concerns may further inflate this bias. In addition, our results suggest essential differences between references standards in the timing of referral to specialised respiratory care (e.g. NIV initiation). This is of particular importance considering that NIV improves overall survival,[25] and the suggestive evidence for a beneficial effect of early NIV initiation.[23, 24] The GLI-2012 and NHANES provide more conservative estimates, which may result in a timelier referral as compared to Knudson '76, '83 or ESCS (**Figure 7.3**). Moreover, from the same figure it can be seen that a considerable proportion (~25%) of the patients with severe dyspnoea symptoms still has a %predicted FVC > 50%. This

highlights an important limitation of spirometry in ALS and emphasizes the need to consider both symptomology and spirometry in clinical settings.

The %predicted FVC is one of the primary eligibility criteria in ALS clinical trials and a common outcome for evaluating drug efficacy.[9, 26] Patients are selected for clinical trials based on a specific %predicted FVC (e.g. ≥ 80%).[27] As the %predicted is related to the applied reference standard,[5] the reference standard exerts a considerable effect on the exclusion rate in clinical trials. In our example, the difference in exclusion rates was considerable (33.4% under NHANES III *vs.* 52.8% under ECSC) and the selected trial populations differed in their overall survival. This observation has considerable consequences for the comparability of trial populations and the generalizability of results.[28] This becomes even more critical when one considers the systematic underreporting of reference standards in ALS clinical trials: of the 37 randomized placebo-controlled trials using lung function either as inclusion criterion or efficacy endpoint, only three reported the applied reference standard (2 Knudson '83 & 1 ECSC).[9] This indicates that, for 92% of the clinical trials, it is not known which reference standard was used to determine eligibility or efficacy. In the end, this obscures clinicians' ability to translate trial outcomes to clinical practice.

Our study has limitations that should be considered. First, we used an open-source database of ALS clinical trials. It is well known that trial participants differ from the general population and fast-progressing patients are likely underrepresented.[9, 29] The estimated time to reach a certain %predicted FVC is, therefore, overestimated. Nevertheless, our imperative was not to obtain real-world estimates of the time to reach a critical limit, but merely to illustrate the large effect of the reference standard on population characteristics and outcomes. Moreover, the PRO-ACT dataset contains data over the full-range of %predicted FVC (**Figure 7.1**) and the relative differences between spirometric reference standards will not change in population-based datasets. An important consideration is that respiratory insufficiency is generally defined as a carbon dioxide pressure (PCO_2) exceeding 45 mmHg,[30] an endpoint that could not be evaluated in PRO-ACT. It would be worthwhile to assess the associations between reference standards with PCO_2 and to optimize cut-off criteria for referral to specialized respiratory care. Finally, the implementation of the GLI-2012 in clinical practice is not straightforward due to the underlying mathematical model.[4] Important work to mediate the implementation has been conducted by the GLI-2012 research group.[5] We have extended the implementation tools by providing an

additional web-based tool for clinicians to determine the %predicted FVC according to the GLI reference standard (http://reactive.tricals.org).

In conclusion, our results show how spirometric reference values can considerably affect the utility of spirometry in ALS and, potentially, other neurological diseases. Similar effects may be observed in other outcomes that depend on reference standards such as muscle strength testing or other markers of respiratory function. Our results emphasize the need for the standardization of spirometric reference, which is currently overlooked in guidelines. The GLI-2012 and NHANES III differed minimally and showed the strongest associations with survival and patient-reported symptoms. Standardization of reference standards may optimize clinical decision-making, improve prognostication, enhance between-centre comparisons and unify patient selection for clinical trials.

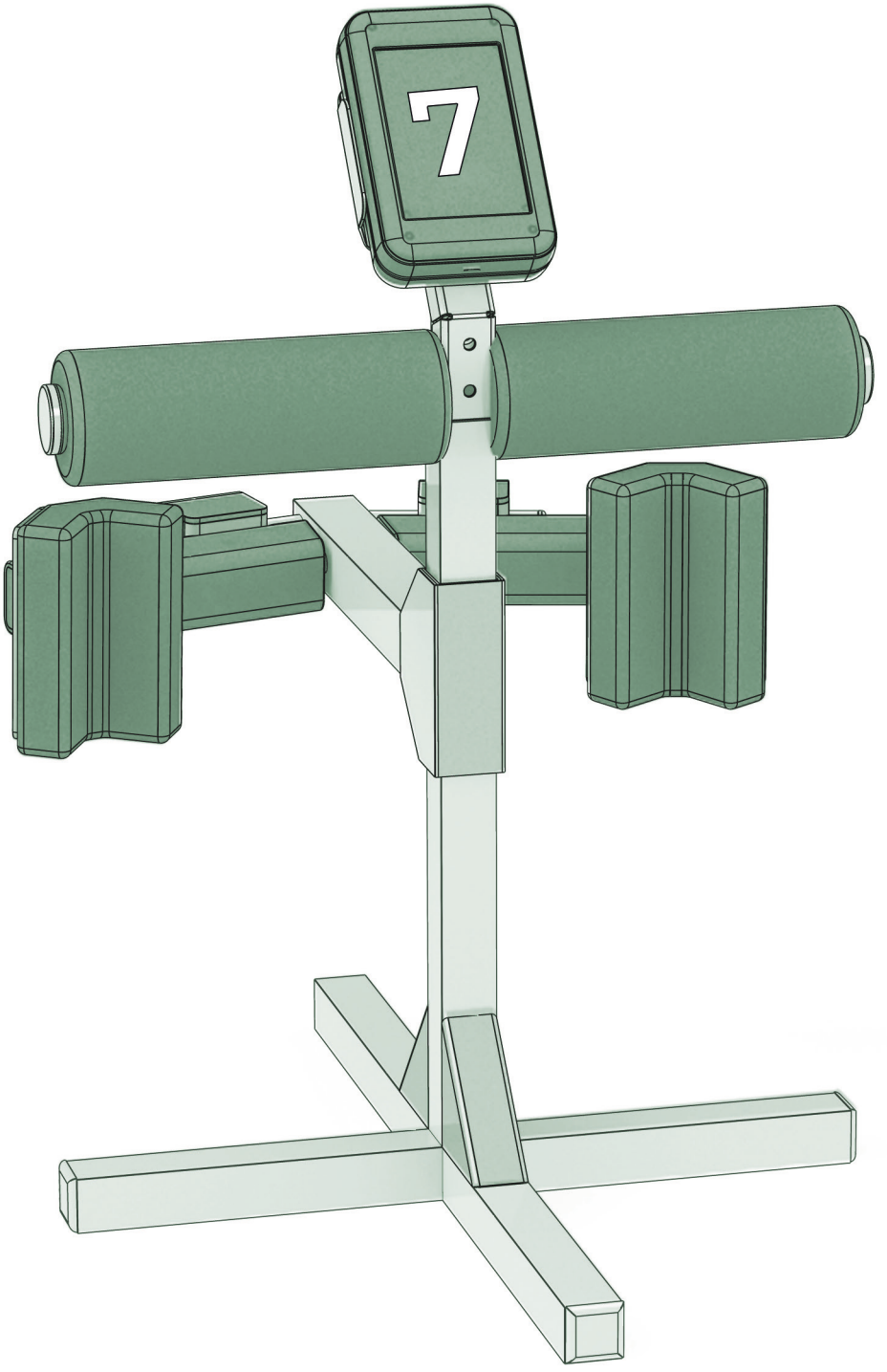
References

1. Lange P, Celli B, Agusti A, et al. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. New Engl J Med 2015;373:111-122.
2. Kelley RC, Ferreira LF. Diaphragm abnormalities in heart failure and aging: mechanisms and integration of cardiovascular and respiratory pathophysiology. Heart Fail Rev 2017;22:191-207.
3. Ambrosino N, Carpene N, Gherardi M. Chronic respiratory care for neuromuscular diseases in adults. Eur Respir J 2009;34:444-451.
4. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J 2012;40:1324-1343.
5. Quanjer PH, Brazzale DJ, Boros PW, Pretto JJ. Implications of adopting the Global Lungs Initiative 2012 all-age reference equations for spirometry. Eur Respir J 2013;42:1046-1054.
6. Crapo RO. Pulmonary-function testing. N Engl J Med 1994;331:25-30.
7. Tilanus TBM, Groothuis JT, TenBroek-Pastoor JMC, et al. The predictive value of respiratory function tests for non-invasive ventilation in amyotrophic lateral sclerosis. Respir Res 2017;18:144.
8. Westeneng HJ, Debray TPA, Visser AE, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. Lancet Neurol 2018;17:423-433.
9. van Eijk RPA, Westeneng HJ, Nikolakopoulos S, et al. Refining eligibility criteria for amyotrophic lateral sclerosis clinical trials. Neurology 2019.
10. National Institute for Health and Care Guidance. Motor neurone disease: assessment and management. https://www.nice.org.uk/guidance/ng42, 2016: Accessed 1 May 2019.
11. Miller RG, Jackson CE, Kasarskis EJ, et al. Practice parameter update: the care of the patient with amyotrophic lateral sclerosis: drug, nutritional, and respiratory therapies (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2009;73:1218-1226.
12. EFNS Task Force on Management of Amyotrophic Lateral Sclerosis, Andersen PM, et al. EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS)--revised report of an EFNS task force. Eur J Neurol 2012;19:360-375.
13. Atassi N, Berry J, Shui A, et al. The PRO-ACT database: design, initial analyses, and predictive features. Neurology 2014;83:1719-1725.
14. Knudson RJ, Slatin RC, Lebowitz MD, Burrows B. The maximal expiratory flow-volume curve. Normal standards, variability, and effects of age. Am Rev Respir Dis 1976;113:587-600.
15. Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B. Changes in the normal maximal expiratory flow-volume curve with growth and aging. Am Rev Respir Dis 1983;127:725-734.
16. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J Suppl 1993;16:5-40.
17. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med 1999;159:179-187.
18. Cudkowicz ME, Titus S, Kearney M, et al. Safety and efficacy of ceftriaxone for amyotrophic lateral sclerosis: a multi-stage, randomised, double-blind, placebo-controlled trial. Lancet Neurol 2014;13:1083-1091.

19. Meininger V, Bensimon G, Bradley WR, et al. Efficacy and safety of xaliproden in amyotrophic lateral sclerosis: results of two phase III trials. Amyotroph Lateral Scler Other Motor Neuron Disord 2004;5:107-117.
20. van Eijk RPA, Eijkemans MJC, Ferguson TA, Nikolakopoulos S, Veldink JH, van den Berg LH. Monitoring disease progression with plasma creatinine in amyotrophic lateral sclerosis clinical trials. J Neurol Neurosurg Psychiatry 2018;89:156-161.
21. Chio A, Logroscino G, Hardiman O, et al. Prognostic factors in ALS: A critical review. Amyotroph Lateral Scler 2009;10:310-323.
22. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. 2015 2015;67:48.
23. Jacobs TL, Brown DL, Baek J, Migda EM, Funckes T, Gruis KL. Trial of early noninvasive ventilation for ALS: A pilot placebo-controlled study. Neurology 2016;87:1878-1883.
24. Vitacca M, Montini A, Lunetta C, et al. Impact of an early respiratory care programme with non-invasive ventilation adaptation in patients with amyotrophic lateral sclerosis. Eur J Neurol 2018;25:556-e533.
25. Bourke SC, Tomlinson M, Williams TL, Bullock RE, Shaw PJ, Gibson GJ. Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomised controlled trial. Lancet Neurol 2006;5:140-147.
26. Andrews JA, Meng L, Kulke SF, et al. Association Between Decline in Slow Vital Capacity and Respiratory Insufficiency, Use of Assisted Ventilation, Tracheostomy, or Death in Patients With Amyotrophic Lateral Sclerosis. JAMA Neurol 2018;75:58-64.
27. Edaravone Writing Group. Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. Lancet Neurol 2017;16:505-512.
28. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365:82-93.
29. Chio A, Canosa A, Gallo S, et al. ALS clinical trials: do enrolled patients accurately represent the ALS population? Neurology 2011;77:1432-1437.
30. Weinberger SE, Schwartzstein RM, Weiss JW. Hypercapnia. N Engl J Med 1989;321:1223-1231.

Part 3

INNOVATIONS OF CLINICAL OUTCOME MEASUREMENTS IN MOTOR NEURON DISORDERS


7

Chapter 7

PORTABLE FIXED DYNAMOMETRY: TOWARDS REMOTE MUSCLE STRENGTH MEASUREMENTS IN PATIENTS WITH MOTOR NEURON DISEASE

Jaap Bakers
Leonard van den Berg
Toju Ajeks
Maxine Holleman
Jill Verhoeven
Anita Beelen
Anne Visser-Meily
Ruben van Eijk

Journal of Neurology 2020

Abstract

Background: We aimed to determine (1) the reliability of a newly developed portable fixed dynamometer (PFD) as compared to the hand-held dynamometer (HHD) in patients with motor neurone disease (MND), (2) the PFD's ability to reduce possible examiner-induced ceiling effects and (3) the relationship between HHD and PFD measurements with loss of function.

Methods: Reliability of isometric muscle strength of the quadriceps was measured in patients with ALS and non-neurological controls using the HHD and PFD. Reliability was estimated by the intraclass correlation coefficient (ICC) and standard error of measurement (SEM) using linear mixed effects models, and the Bland-Altman method of agreement.

Results: In total, 45 patients with MND and 43 healthy controls were enrolled in this study. The ICC of the PFD was excellent and similar in both patients and controls ($ICC_{Patients}$ 99.5% *vs.* $ICC_{Controls}$ 98.6%) with a SEM of 6.2%. A strong examiner-induced ceiling effect in HHD was found when the participant's strength exceeded that of examiner. Employing the PFD increased the range of muscle strength measurements across individuals nearly 2-fold from 414 to 783 Newton. PFD measurements were approximately linearly related to the reported loss of function.

Conclusions: Portable fixed dynamometry may significantly reduce examiner-induced ceiling effects, optimize the standardization of muscle strength testing, and maximize reliability. Ultimately, PFD may improve the delivery of care due to its potential for unsupervised, home-based assessments and reduce the burden to the patient of participating in clinical trials for MND or other neuromuscular diseases.

Introduction

Progressive muscle weakness is the hallmark of motor neurone disease (MND).[1] Muscle strength testing has, therefore, a central role in monitoring MND progression.[2-4] Isometric muscle strength testing using the Hand Held Dynamometer (HHD) is preferred to the Medical Research Council (MRC) scale due to its increased objectivity and sensitivity.[3,5-7] Despite its user-friendliness, portability and cost-effectiveness, the HHD's reliability depends on the technique and strength of the examiner.[8-13] These limitations become especially apparent in strong muscle groups such as the quadriceps, resulting in possible ceiling effects and a reduced sensitivity for quantifying MND progression in early disease stages.[14,15]

Reducing examiner variability may, therefore, significantly reduce measurement error and optimize the sensitivity of muscle strength testing in MND. Fixed dynamometry (i.e. fixation of the dynamometer in a rigid structure) has been shown to alleviate the limitations of the HHD, but currently available systems still require a trained examiner and hospital visits.[16-22]

Given the current transition to home-based assessments (i.e. remote monitoring),[23,24] and the accompanying need for reliable and unsupervised measurements of disease progression,[25,26] we developed a portable fixed dynamometer (PFD). The PFD was developed to evaluate quadriceps strength, because, although there is a gradual rate of decline, the function of this muscle is preserved for a relatively long time.[27] The quadriceps is, therefore, potentially a sensitive muscle group for objective measurement of disease progression in MND. However, as the quadriceps is one of the strongest muscles of the human body, its assessment is challenging, which leads to high variability among examiners when using the HHD.[28]

In this study, we aimed to determine (1) the reliability of PFD as compared to the HHD, (2) the PFD's ability to reduce possible examiner-induced ceiling effects and (3) the relationship between HHD and PFD measurements with loss of function.

Methods

Study population

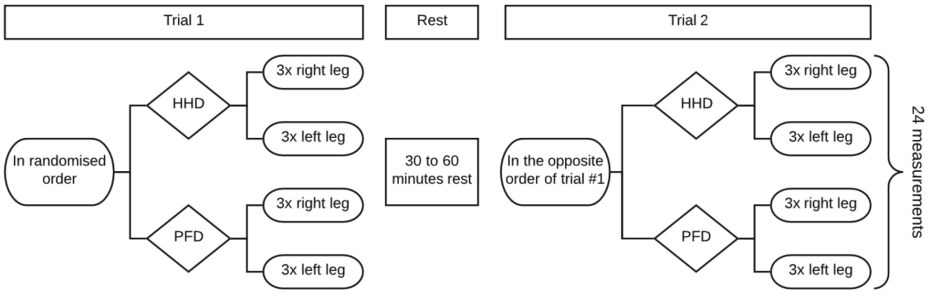
Study participants consisted of two groups: (1) participants with a diagnosis within the motor neurone disease spectrum (i.e. Amyotrophic Lateral Sclerosis, Progressive Muscular Atrophy or Primary Lateral Sclerosis),[29] and (2) controls without

a neurological condition. Participants were excluded from the study if they met one of the following criteria: (1) less than MRC 2 quadriceps strength in both legs, (2) recent or current pain in knee joint or quadriceps muscle, (3) not able to follow test instructions from the examiner, or (4) having another non-MND disorder that affects muscle strength. All patients with MND were recruited from the outpatient clinic of the University Medical Centre Utrecht, The Netherlands. Control participants were recruited from personnel and students from the department of rehabilitation and geriatrics. The study was approved by the Medical Ethics Committee of the UMCU (protocol number 18-243). All study participants gave written informed consent to participate in this study.

Procedures and measurement techniques

Two examiners (T.G.A. and M.J.H) were certified for isometric HHD muscle testing after (1) completing the 'Treatment Research Initiative to Cure ALS' (TRICALS) e-course 'Isometric muscle testing in ALS', and (2) satisfactorily completing supervised HHD testing of five control participants. After the registration of patient information and collection of MND-specific characteristics (e.g. ALS functional rating scale [ALSFRS-R] and respiratory functioning), reliability was assessed in two separate trials on the same day (**Figure 1**). Thirty minutes before each trial, participants were requested to refrain from engaging in any strenuous activities. Each trial consisted of six measurements per leg, three with the HHD and three with the PFD. Participants were seated on a chair with back support, hips and knees were kept at 90 degrees. In order to rule out the influence of arm function, arms were placed in the lap. The starting sequence of the assessment (HHD or PFD) was randomized in order to minimize the effect of fatigue. All trials within the same participant were conducted by the same examiner, verbally instructing and motivating the patient.

Figure 1. Overview of study procedures



Hand-held dynamometry

The HHD (MicroFET 2, HOGGAN Scientific) assessments consisted of three isometric 'break contractions',[30] approximately 10 seconds apart. The HHD was placed one centimetre proximal to the midline between the malleoli. If necessary, a towel roll was placed under the knee to prevent the foot from touching the ground. During the maximal contraction, the examiner not only attempted to offer sufficient resistance, but also strived to give a gentile break in the opposite direction of the isometric strength. The score of each measurement, as well as the ability to perform a 'break' (classified as *break* OR *unable-to-break*), was registered.

Portable fixed dynamometer

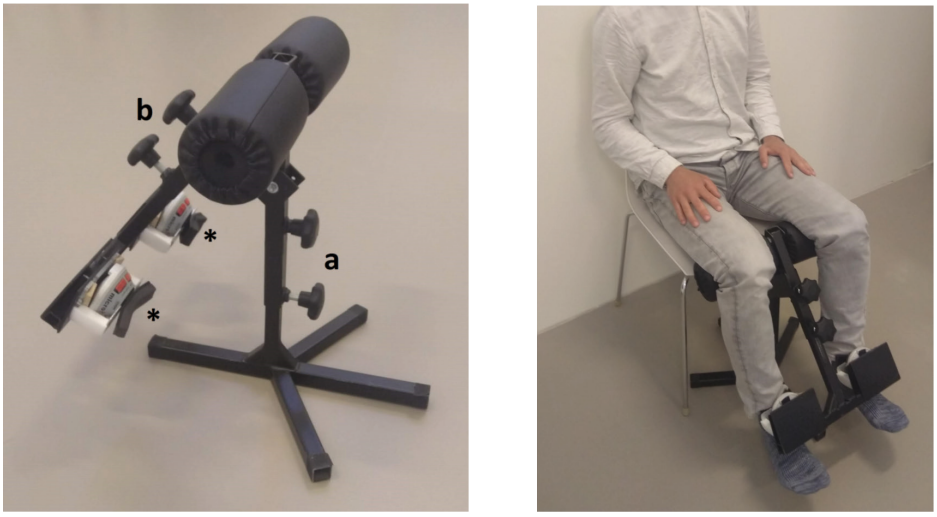
In order to be able to measure muscle strength of the quadriceps muscle using a portable, but fixed device, we constructed a simple rigid construction with two HHD holders. In Figure **2** we illustrate the PFD with the HHDs fixed in a rigid, but portable construction. The PFD can be easily placed in front of a (wheel)chair and standardizes the starting position of the knee joint at 90 degrees. The vertical arm of the PFD (indicated by *a,* **Figure 2**) can be adjusted to prevent the feet from touching the ground. The arm length of the PFD (indicated by *b*) is adjusted to the length of the participant's lower leg and positions the HHD. We evaluated muscle strength in a fashion identical to that of the HHD. Importantly, in contrast to the HHD, and due to the nature of the fixed construction, the PFD tests consisted of an isometric 'make contraction' as opposed to a 'brake contraction'. [30]

Statistical analysis

Data were summarized using the mean with standard deviation (SD) for continuous variables and number with percentage for categorical variables. Mean differences (MD) were calculated between cases and controls using an independent t-test. For each participant, per trial (i.e. trial 1 or 2) and per method (i.e. HHD or PFD), we selected the highest muscle strength score from the two values which were most similar of the three measurements taken. Reliability was assessed using (1) the intraclass correlation coefficient (ICC) and its associated standard error of the measurement (SEM), and (2) the Bland-Altman method of agreement.[31,32] Due to the heteroscedastic nature of the data, we applied a [10]log-scale transformation, and calculated the mean difference between trials. The ICC was estimated using a linear mixed effects model, incorporating only a fixed intercept and random intercept per subject. The ICC was then calculated as the percentage of the total variation (i.e. sum of the between-subject and within-subject variation) that can be

explained by between-subject differences (i.e. the between-subject variation); 95% confidence intervals were obtained by means of bootstrapping (n = 1000). The SEM was calculated by taking the square root of the within-subject variation and back-transformed as described elsewhere.[32]

Figure 2. Prototype of the Portable Fixed Dynamometer



A rigid framework with two holders for the MicroFET dynamometers (*) was created to remove the need for an examiner when evaluating quadriceps strength. With the vertical arm (a), the height was adjusted to prevent the feet from touching the ground. The diagonal arm (b) enabled adjustment of the dynamometer pad to one centimeter proximal to the midline between the malleoli.

To determine the presence of a ceiling effect, we assessed the relationship between HHD and PFD measurements with a linear mixed model. This relationship was modelled using a natural spline with four knots. Finally, we compared the relationship between HHD and PFD measurements with functional loss, using item 8 (walking) of the ALSFRS-R. All analyses were conducted using R. Linear mixed models were fitted and bootstrapped using the *lmer* and *bootMer* functions (R package lme4, version 1.1–21), respectively.[33]

Results

Between 4th April 2018 and 8th March 2019, 88 Dutch participants were enrolled in this study: 45 patients with MND and 43 non-neurological controls; their baseline characteristics are presented in **Table 1**. The age distribution ranged from 22 to 94 years. Patients with MND had a significantly lower quadriceps strength compared to

the control population ($p < 0.001$). Group difference in quadriceps strength (MND vs controls) was considerably larger on PFD (-113N, 95% CI -52N to -174N, standardized: 0.792) as compared to HHD (-60N, 95% CI 23N to 97N, standardized: 0.679).

Table 1. Baseline characteristics of study population

Characteristic	**Patients** *(N = 45)*	**Controls** *(N = 43)*
Age (years)		
Median	62	47
Range	30 to 84	22 to 94
Males	27 (60%)	22 (51%)
Body mass index (kg/m^2)	25 (2)	24 (3)
Muscle strength, average (N)		
HHD	216 (94)	275 (80)
PFD	237 (126)	350 (159)
Muscle strength, range (N)		
HHD	15 to 373	128 to 414
PFD	8 to 508	91 to 783
MND Subtype		
ALS	41 (91%)	-
PMA	3 (7%)	-
PLS	1 (2%)	-
Bulbar onset	11 (24%)	-
FVC, %predicted - GLI2012	78 (20)	-
Symptom duration (months)		
Median	26	-
Range	8 - 311	-
Diagnostic delay (months)		
Median	12	-
Range	3 - 157	-
Riluzole use	37 (82%)	-
ALSFRS-R total score (SD)	35 (7)	-
ΔFRS (points per month)		
Median	0.41	-
Range	0.01 – 1.94	-

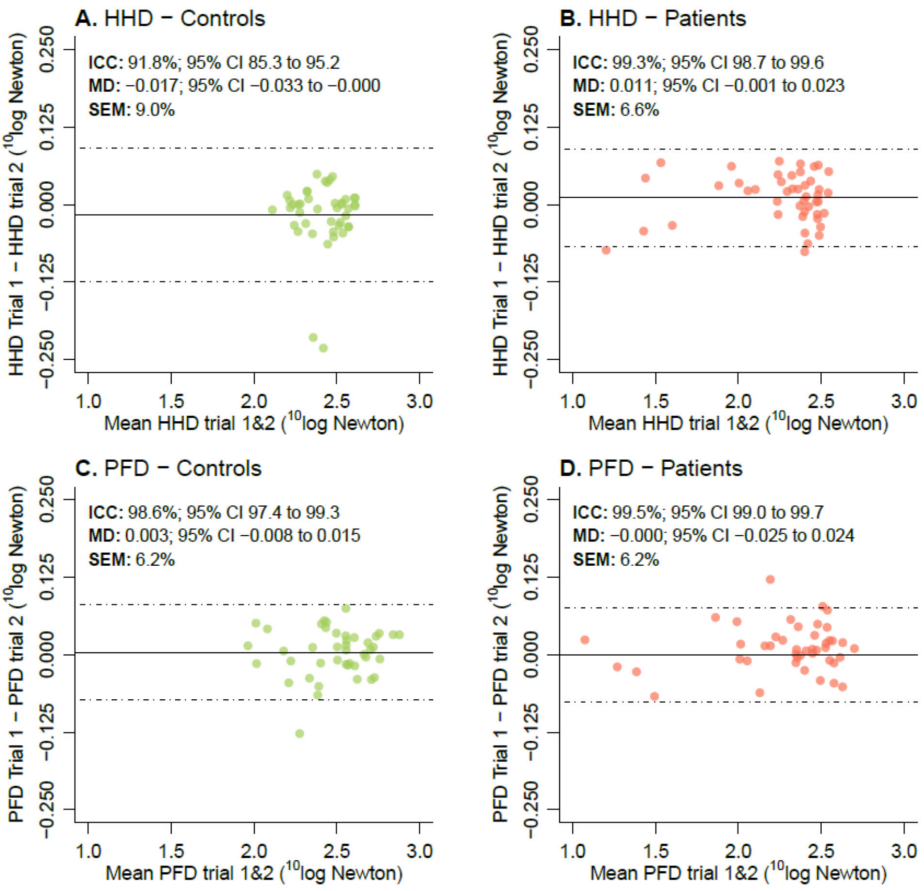
Data are in mean (SD) or no. (%). HHD = hand-held dynamometry; PFD = portable fixed dynamometry; MND = motor neurone disease; ALS = amyotrophic lateral sclerosis; PMA = progressive muscular atrophy; PLS = primary lateral sclerosis; ALSFRS-R = revised ALS functional rating scale; ΔFRS = 48 – ALSFRS-R total score / symptom duration.[46] Prognostic subgroups are based on the ENCALS prediction model.[47]

Reliability HHD and PFD

In **Figure 3** we provide the measurements of reliability. Due to the heteroscedastic nature of the data, we applied a ^{10}log transformation on the muscle strength scores.

For both the HHD and PFD, no systematic differences were found between trials 1 and 2 as indicated by their mean difference (MD, i.e. mean trial 1 – trial 2). The ICC of the PFD was excellent and similar in both patients and controls (ICC $_{Patients}$ 99.5% *vs.* ICC $_{Controls}$ 98.6%). The back-transformed SEM of the PFD was 6.2% of the mean strength in Newton, meaning that test-retest values may vary by as much as ±12.5% of their mean (i.e. approximately 2 times the SEM).[32] Interestingly, the ICC of the HHD was excellent, albeit considerably lower in controls (ICC $_{Patients}$ 99.3% *vs.* ICC $_{Controls}$ 91.8%).

Figure 3. Bland-Altman plots for test-retest reliability

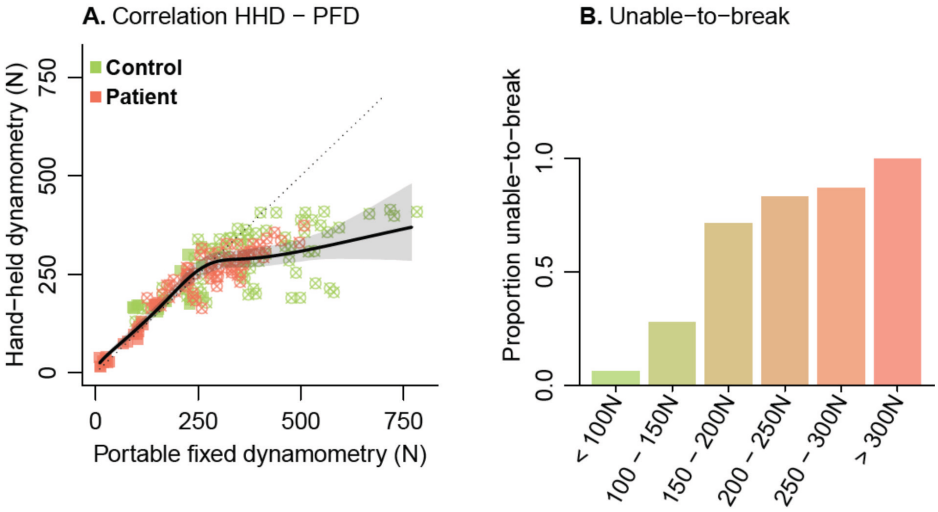


Bland-Altman plots of the HHD (**A-B**) and PFD (**C**-**D**) for non-neurological controls and patients with MND, respectively. Due to the heteroscedastic nature of the data, we applied a ^{10}log-transformation. MD = mean difference between trials 1 and 2; ICC = Intraclass Correlation Coefficient; SEM = Standard error of measurement, expressed as percentage of the mean on the original scale (Newton).[32]

Examiner-induced ceiling effects in Hand-Held Dynamometry

Figure 4A shows the relationship between HHD and PFD measurements. For lower values up to 200N, the PFD and HHD show visually a level of high agreement, resulting in nearly identical strength values. Above 200 to 250 N, the proportion of *unable-to-break* (**Figure 4B**) increases and the correlation between PFD and HHD weakens (as reflected by a deviation from the dashed line). As patients, on average, had lower muscle strength than controls, the correlation coefficient between HHD and PFD was high in patients (Pearson's *r* 0.94, 95% CI 0.89 to 0.97), whereas in controls it was considerably lower (Pearson's *r* controls: 0.71, 95% CI 0.52 to 0.83).

Figure 4. Association between hand-held and portable fixed dynamometry



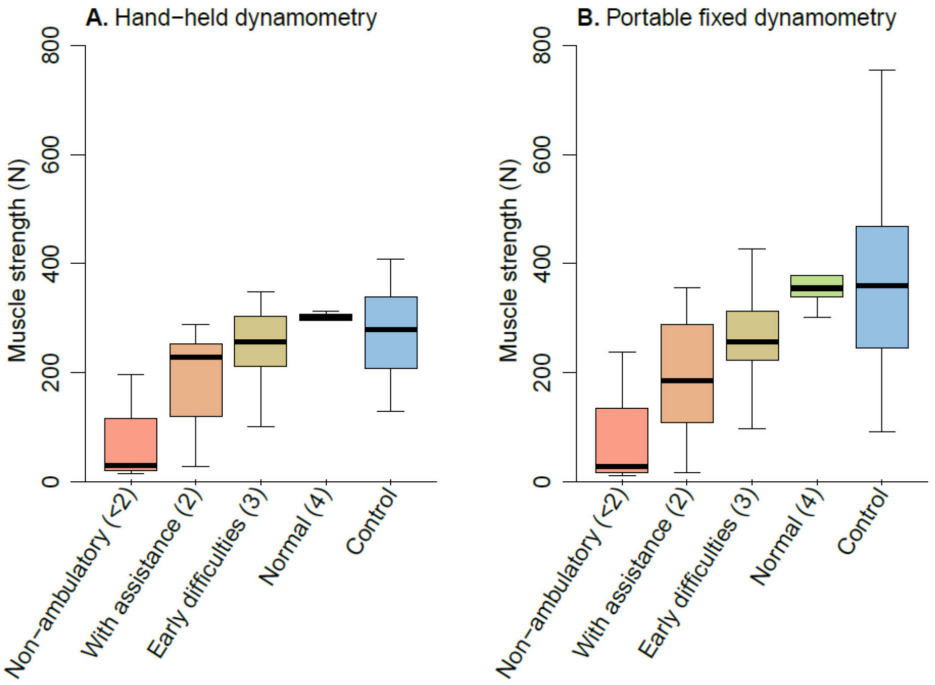
(**A**) Association between HHD and PFD measurements. The dotted line represents a correlation of 1 (i.e. HHD measurement = PFD measurement and vice-versa). The black solid line represents the observed association, estimated using a linear mixed effects model. The dots with crosses are data points classified as *unable-to-break*. (**B**) Muscle strength based on hand-held dynamometry data was categorized and per sub-group we determined the proportion *unable-to-break* A clear pattern emerges with the assessor no longer being able to measure the full muscle strength from 300N upwards (i.e. a proportion of *unable-to-break* of 1).

Relationship with loss of function

Both HHD and PFD measurements in patients were strongly related with question 8 (walking) of the ALSFRS-R (**Figure 5**). Similar to **Figure 4A,** a ceiling effect in HHD measurements can be seen in both patients and controls. In patients the range of HHD measurements seemed to be limited as walking ability increased from early ambulatory difficulties to normal functioning. This ceiling effect was absent

in PFD measurements, where muscle strength was approximately linearly related to the reported functional loss. In the controls, the range of muscle strength was considerably larger for PFD (8N to 783N) compared to HHD (15N to 414N).

Figure 5. Relationship between functional loss and muscle strength measurements



Boxplot of the HHD and PFD measurements in 45 patients with MND and 43 non-neurological controls (blue). Functional loss in MND patients was categorized using self-reported function scores of the ALS functional rating scale item 8 (walking); the score range was 0 to 4, where 0 indicates no leg muscle function (not included) and 4 indicates normal walking function.

Dicussion

In this study, we showed that (1) the PFD achieves a high level of precision with an excellent reliability over a wider range in muscle strength measurements, (2) the PFD has an advantage over hand-held isometric strength measurements as it reduces examiner-induced ceiling effects and (3) the PFD measurements have an approximately linear relationship with functional loss in MND. Fixating a dynamometer in a simple and portable framework opens the door to standardized self-assessments by patients in their homes, and may eventually decrease the number of hospital visits and reduce the burden to participate in clinical trials.

Optimizing muscle strength testing is important in order to optimize the evaluation of efficacy of new MND therapies and to contribute to the delivery of remote care.[20] In the field of voluntary muscle strength testing, the Biodex system is considered to be the gold standard.[21,34,35] Similar to the Biodex, other systems have been developed that fixate a loading cell or dynamometer in a rigid framework (e.g. Maximum Voluntary Isometric Contraction [MVIC], Accurate Test of Limb Isometric Strength [ATLIS]). The MVIC and ATLIS have proven to be reliable in patients with ALS with an excellent intra-rater test variability ranging from 8.6% to 8.9% for the assessment of quadriceps muscle strength .[16,36] These rigid frameworks are, however, not portable, require visits to the out-patient department, are relatively expensive and are still operated by a trained examiner. Other methods that are applied to fixate dynamometers are belt or clamp fixations and stabilization devices. [18,34,37-41] Although these have the advantage of being portable and less expensive, unsupervised use has led to inaccurate measurements and patient discomfort.[42]

Our results show that the examiner-induced ceiling effect in HHD measurements is an important source of variability. This is a critical observation as the HHD is a common endpoint in both exploratory and confirmatory clinical trials for ALS.[5,6] The ceiling effect prevents the investigator from determining the patient's true strength if the examiner is no longer able to overcome the participant's muscle strength.[10,15,21,22] This may become particularly problematic in longitudinal settings if patients are assessed by multiple examiners, with each examiner being able to withstand a different amount of force (commonly around 200 to 300 Newton).[14,15] More importantly, site personnel training and their experience are unlikely to fully eliminate these effects, leading to persistent between-examiner and -site variability. Although the ceiling effects are irrelevant for relatively weak muscle groups, the true strength of major muscle groups like the quadriceps could be significantly underestimated with HHD. Particularly in asymptomatic stages of the disease (as was shown in **Figure 5**), this could mean important signals of early disease progression or therapeutic efficacy are missed.

Fixating the HHD in a rigid framework could, therefore, reduce the effect of examiner strength on muscle strength assessments. As is indicated by our results, employing a rigid framework around the HHD increases the range of muscle strength measurements across individuals nearly 2-fold, from 414 to 783 Newton. On an individual level, these results are critical as they suggest that the PFD may track the progression curve of quadriceps strength better, especially in early disease. Moreover, in our study the increased range of the PFD better reflected the group-

level difference between cases and controls as compared to the HHD. The examiner-induced ceiling effect of HHD measurements suggests that group-level differences may increase when using the PFD. In addition to the potential increase in sensitivity of the PFD to detect changes over time, larger effect sizes (e.g. standardized group-differences) could have significant benefits for clinical trials in terms of sample size. Importantly, the reliability of the PFD remained practically unchanged; implementing the PFD may, therefore, help standardize muscle testing protocols in multi-Centre settings and reduce site-variability.

Our study does, however, have limitations. As comparison of the PFD to the HHD was limited to strength assessment of the quadriceps, it remains to be established how well the PFD performs compared to the gold standard (i.e. Biodex). Previous research indicated a good correlation between HHD and the Biodex for measurements of the quadriceps.[21,43,44] Given the strong correlation between HHD and PFD, we expect that the PFD and Biodex will have an equivalent correlation.

The PFD is currently applied to one muscle group and might, therefore, not capture the full extent of motor function loss in patients with MND (e.g. arm weakness). Extensive muscle strength testing is time-consuming and increases patient burden which may lead to significant attrition over time.[2,45] It is, therefore, critical to minimize the number of assessments, while obtaining sufficient information for clinical decision-making or monitoring disease progression. Dedicated longitudinal studies are required to use a data-driven approach to determine which muscle groups provide complementary information in addition to quadriceps strength monitoring. An important aspect to consider is the current transition to home-based assessments (i.e. remote monitoring); [23-26] portability, cost-effectiveness and the potential for unsupervised, user-friendly assessments should have a prominent place in any future iteration of the PFD.

In conclusion, our study reveals the value of fixed dynamometry in reducing examiner-induced ceiling effects and optimizing the standardization of muscle strength testing in order to maximize reliability. The PFD may be better than other devices in tracking disease progression in individual patients and revealing group-level differences. Ultimately, extending the PFD to home-based settings in MND or other neuromuscular diseases, could improve the delivery of remote care, optimize trial efficiency, and reduce the burden to the patient of participating in clinical trials. A prerequisite for independent remote use is further development of the PFD by integrating the load cells and introducing a patient-friendly user interface.

References

1. National Institute for Health and Care Excellence: Clinical Guidelines; Motor Neurone Disease: Assessment and Management. London 2016.
2. Cudkowicz M, Qureshi M, Shefner J. Measures and markers in amyotrophic lateral sclerosis. *NeuroRx* 2004; **1**: 273-283.
3. Paganoni S, Cudkowicz M, Berry JD. Outcome measures in amyotrophic lateral sclerosis clinical trials. *Clin Investig (Lond)* 2014; **4**: 605-618.
4. Rutkove SB. Clinical Measures of Disease Progression in Amyotrophic Lateral Sclerosis. *Neurotherapeutics.* 2015; **12** :384-393.
5. Cudkowicz ME, Titus S, Kearney M, et al. Safety and efficacy of ceftriaxone for amyotrophic lateral sclerosis: a multi-stage, randomised, double-blind, placebo-controlled trial. *Lancet Neurol.* 2014; **13**; 1083-1091.
6. Cudkowicz ME, van den Berg LH, Shefner JM, et al. Dexpramipexole versus placebo for patients with amyotrophic lateral sclerosis (EMPOWER): a randomised, double-blind, phase 3 trial. *Lancet Neurol.* 2013; **12**: 1059-1067.
7. Meininger V, Genge A, van den Berg LH, et al. Safety and efficacy of ozanezumab in patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Neurol.* 2017; **16**: 208-216.
8. Kolber MJ, Cleland JA. Strength testing using hand-held dynamometry. *Physical Therapy Reviews.* 2005;**10**:99-112.
9. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review. *PM & R : the journal of injury, function, and rehabilitation.* 2011;**3**:472-479.
10. Keating JL, Matyas TA. The influence of subject and test design on dynamometric measurements of extremity muscles. *Phys Ther.* 1996;**76**:866-889.
11. Lu TW, Hsu HC, Chang LY, Chen HL. Enhancing the examiner's resisting force improves the reliability of manual muscle strength measurements: comparison of a new device with hand-held dynamometry. *J Rehabil Med.* 2007;**39**:679-684.
12. Wikholm JB, Bohannon RW. Hand-held Dynamometer Measurements: Tester Strength Makes a Difference. *J Orthop Sports Phys Ther.* 1991;**13**:191-198.
13. Lu TW, Chien HL, Chang LY, Hsu HC. Enhancing the examiner's resisting force improves the validity of manual muscle strength measurements: application to knee extensors and flexors. *J Strength Cond Res.* 2012;**26**:2364-2371.
14. Beck M, Giess R, Wurffel W, Magnus T, Ochs G, Toyka KV. Comparison of maximal voluntary isometric contraction and Drachman's hand-held dynamometry in evaluating patients with amyotrophic lateral sclerosis. *Muscle & nerve.* 1999;**22**:1265-1270.
15. Visser J, Mans E, de Visser M, et al. Comparison of maximal voluntary isometric contraction and hand-held dynamometry in measuring muscle strength of patients with progressive lower motor neuron syndrome. *Neuromuscul Disord.* 2003;**13**:744-750.
16. Andres PL, Skerry LM, Munsat TL, et al. Validation of a new strength measurement device for amyotrophic lateral sclerosis clinical trials. *Muscle & nerve.* 2012;**45**:81-85.
17. Bui KL, Mathur S, Dechman G, Maltais F, Camp P, Saey D. Fixed Handheld Dynamometry Provides Reliable and Valid Values for Quadriceps Isometric Strength in People With Chronic Obstructive Pulmonary Disease: A Multicenter Study. *Phys Ther.* 2019.
18. Kollock RO, Jr., Onate JA, Van Lunen B. The reliability of portable fixed dynamometry during hip and knee strength assessments. *J Athl Train.* 2010;**45**:349-356.

19. Mentiplay BF, Perraton LG, Bower KJ, et al. Assessment of Lower Limb Muscle Strength and Power Using Hand-Held and Fixed Dynamometry: A Reliability and Validity Study. *PLoS One.* 2015;10(10):e0140822.

20. Andres PL, Allred MP, Stephens HE, et al. Fixed dynamometry is more sensitive than vital capacity or ALS rating scale. *Muscle & nerve.* 2017;**56**:710-715.

21. Martin HJ, Yule V, Syddall HE, Dennison EM, Cooper C, Aihie Sayer A. Is hand-held dynamometry useful for the measurement of quadriceps strength in older people? A comparison with the gold standard Bodex dynamometry. *Gerontology.* 2006;**52**:154-159.

22. Toonstra J, Mattacola CG. Test-retest reliability and validity of isometric knee-flexion and -extension measurement using 3 methods of assessing muscle strength. *J Sport Rehabil.* 2013;**22**.

23. Helleman J, Van Eenennaam R, Kruitwagen ET, et al. Telehealth as part of specialized ALS care: feasibility and user experiences with "ALS home-monitoring and coaching". *Amyotroph Lateral Scler Frontotemporal Degener.* 2020:1-10.

24. Hobson EV, Baird WO, Bradburn M, et al. Using telehealth in motor neuron disease to increase access to specialist multidisciplinary care: a UK-based pilot and feasibility study. *BMJ Open.* 2019;**9**:e028525.

25. Rutkove SB, Qi K, Shelton K, Liss J, Berisha V, Shefner JM. ALS longitudinal studies with frequent data collection at home: study design and baseline data. *Amyotroph Lateral Scler Frontotemporal Degener.* 2019;**20**:61-67.

26. van Eijk RPA, Bakers JNE, Bunte TM, de Fockert AJ, Eijkemans MJC, van den Berg LH. Accelerometry for remote monitoring of physical activity in amyotrophic lateral sclerosis: a longitudinal cohort study. *J Neurol.* 2019;**266**:2387-2395.

27. Rushton DJ, Andres PL, Allred P, Baloh RH, Svendsen CN. Patients with ALS show highly correlated progression rates in left and right limb muscles. *Neurology.* 2017;89(2):196-206.

28. Shefner JM, Liu D, Leitner ML, et al. Quantitative strength testing in ALS clinical trials. *Neurology.* 2016;**87**:617-624.

29. van Es MA, Hardiman O, Chio A, et al. Amyotrophic lateral sclerosis. *Lancet.* 2017;**390**:2084-2098.

30. van der Ploeg RJ, Oosterhuis HJ. The "make/break test" as a diagnostic tool in functional weakness. *J Neurol Neurosurg Psychiatry.* 1991;**54**:248-251.

31. Ludbrook J. Confidence in Altman-Bland plots: a critical review of the method of differences. *Clin Exp Pharmacol Physiol.* 2010;**37**:143-149.

32. Euser AM, Dekker FW, le Cessie S. A practical approach to Bland-Altman plots and variation coefficients for log transformed variables. *J Clin Epidemiol.* 2008;**61**:978-982.

33. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *2015.* 2015;**67**:48.

34. Hogrel JY, Benveniste O, Bachasson D. Routine monitoring of isometric knee extension strength in patients with muscle impairments using a new portable device: cross-validation against a standard isokinetic dynamometer. *Physiol Meas.* 2020;**41**:015003.

35. Roussel MP, Hebert LJ, Duchesne E. Intra-Rater Reliability and Concurrent Validity of Quantified Muscle Testing for Maximal Knee Extensors Strength in Men with Myotonic Dystrophy Type 1. *J Neuromuscul Dis.* 2019;**6**:233-240.

36. Andres PL, Hedlund W, Finison L, Conlon T, Felmus M, Munsat TL. Quantitative motor assessment in amyotrophic lateral sclerosis. *Neurology.* 1986;**36**:937-941.

37. Jackson SM, Cheng MS, Smith AR, Jr., Kolber MJ. Intrarater reliability of hand held dynamometry in measuring lower extremity isometric strength using a portable stabilization device. *Musculoskelet Sci Pract.* 2017;**27**:137-141.

38. Koblbauer IF, Lambrecht Y, van der Hulst ML, et al. Reliability of maximal isometric knee strength testing with modified hand-held dynamometry in patients awaiting total knee arthroplasty: useful in research and individual patient settings? A reliability study. *BMC Musculoskelet Disord*. 2011;**12**:249.

39. Kolber MJ, Beekhuizen K, Cheng MS, Fiebert IM. The reliability of hand-held dynamometry in measuring isometric strength of the shoulder internal and external rotator musculature using a stabilization device. *Physiother Theory Pract*. 2007;**23**:119-124.

40. Shin S, Lee K, Song C. Relationship of body composition, knee extensor strength, and standing balance to lumbar bone mineral density in postmenopausal females. *J Phys Ther Sci*. 2016;**28**:2105-2109.

41. Thorborg K, Bandholm T, Holmich P. Hip- and knee-strength assessments using a hand-held dynamometer with external belt-fixation are inter-tester reliable. *Knee Surg Sports Traumatol Arthrosc*. 2013;**21**:550-555.

42. Hansen EM, McCartney CN, Sweeney RS, Palimenio MR, Grindstaff TL. Hand-held Dynamometer Positioning Impacts Discomfort During Quadriceps Strength Testing: A Validity and Reliability Study. *Int J Sports Phys Ther*. 2015;**10**:62-68.

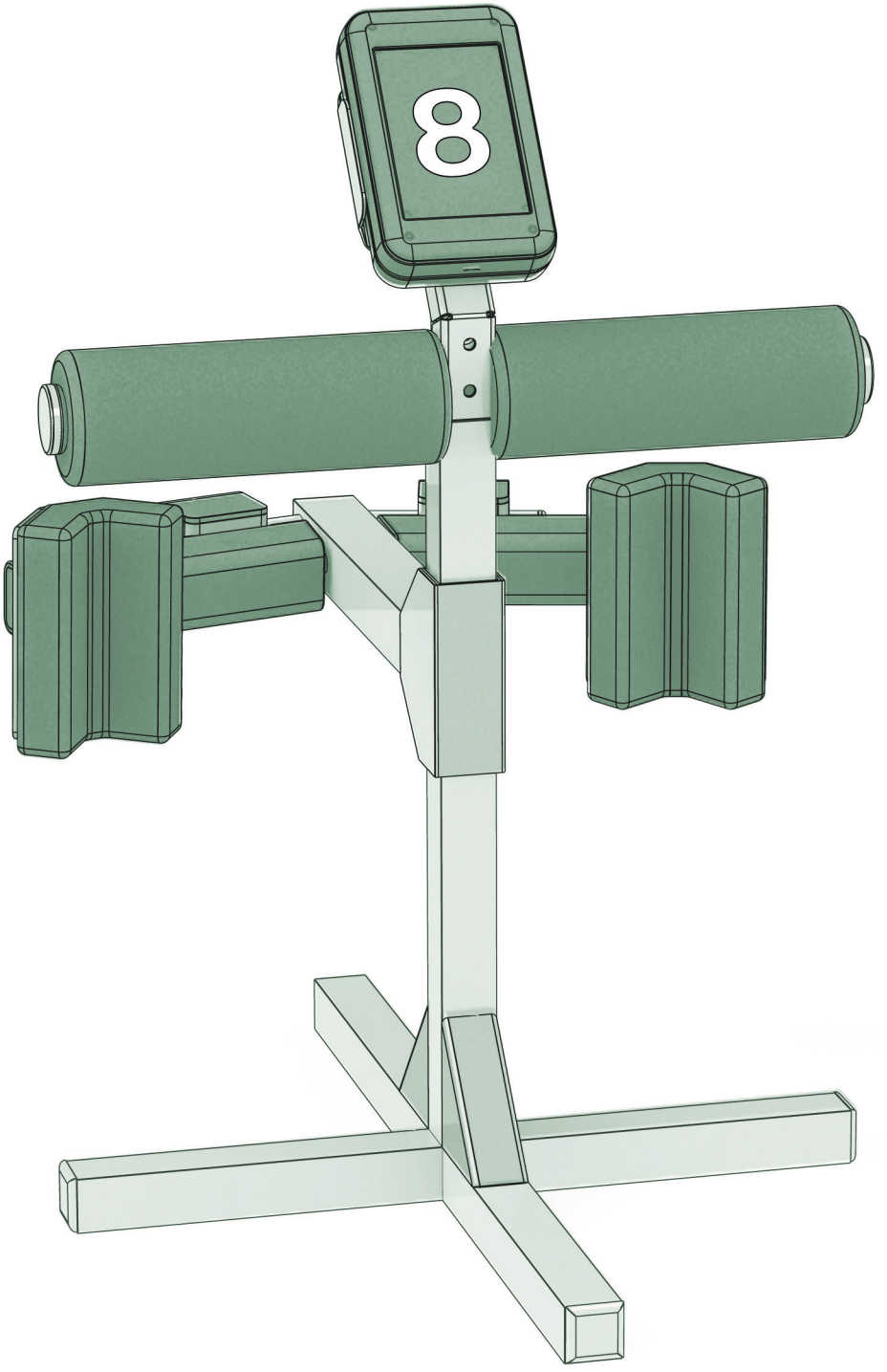
43. Bohannon RW. Hand-held compared with isokinetic dynamometry for measurement of static knee extension torque (parallel reliability of dynamometers). *Clin Phys Physiol Meas*. 1990;**11**:217-222.

44. Reed RL, Den Hartog R, Yochum K, Pearlmutter L, Ruttinger AC, Mooradian AD. A comparison of hand-held isometric strength measurement with isokinetic muscle strength measurement in the elderly. *J Am Geriatr Soc*. 1993;**41**:53-56.

45. Great Lakes ALSSG. A comparison of muscle strength testing techniques in amyotrophic lateral sclerosis. *Neurology*. 2003;**61**:1503-1507.

46. Kimura F, Fujimura C, Ishida S, et al. Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. *Neurology*. 2006;**66**:265-267.

47. Westeneng HJ, Debray TPA, Visser AE, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. *Lancet Neurol*. 2018;**17**:423-433.


8

Chapter 8

ACCELEROMETRY FOR REMOTE MONITORING OF PHYSICAL ACTIVITY IN AMYOTROPHIC LATERAL SCLEROSIS: A LONGITUDINAL COHORT STUDY

Ruben van Eijk
Jaap Bakers
Tommy Bunte
Arianne Fockert
Marinus Eijkemans
Leonard van den Berg

Journal of Neurology 2019

Abstract

Objective: In this study we aim to determine the value of remote, accelerometer-based monitoring of physical activity in patients with ALS.

Methods: This longitudinal cohort study was conducted in a home-based setting; all study materials were sent by mail. Patients wore the ActiGraph during waking hours for 7 days every 2 to 3 months and provided information regarding their daily functioning (ALSFRS-R). We defined four accelerometer-based endpoints that either reflect the average daily activity or quantify the patient's physical capacity.

Results: A total of 42 patients participated; the total valid monitoring period was 9,288 hours with a 93.0% adherence rate. At baseline, patients were active 27.9% (range 11.6% to 52.4%.) of their time; this declined by 0.64% (95% 0.43 – 0.86, $p < 0.001$) per month. Accelerometer-based endpoints were strongly associated with the ALSFRS-R (r 0.78, 95% CI 0.63 – 0.92, $p < 0.001$), but showed less variability over time than the ALSFRS-R (coefficient of variation 0.64 to 0.81 *vs.* 1.06, respectively). Accelerometer-based endpoints could reduce sample size by 30.3% for 12-month trials and 44.6% for 18-month trials; for trials lasting less than 9 months, the ALSFRS-R resulted in smaller sample sizes.

Conclusions: Accelerometry is an objective method for quantifying disease, which could obtain real-world insights in the patient's physical functioning and may personalize the delivery of care. In addition, remote monitoring provides patients with the opportunity to participate in clinical trials from home, paving the way to a patient-centric clinical trial model.

Introduction

The progressive and debilitating nature of amyotrophic lateral sclerosis (ALS) restricts patients from participating in clinical trials. Trial participation is burdensome due to the frequent clinical assessments, laboratory tests and hospital visits. Consequently, there is a selective enrollment of patients, which obscures the collection of safety and efficacy data on the majority of patients.[1-4] In addition, the extensive heterogeneity between patients complicates the quantification of disease progression.[4] This affects the design of clinical trials and their ability to detect treatment responses.

Therefore, there is an increasing interest in remote monitoring of efficacy endpoints.[5] Remote monitoring maximizes the collection of information outside clinical visits and could make in-clinic visits superfluous. Numerous studies are revealing the feasibility of remote monitoring of cardiac,[6] respiratory,[7] neurological,[8] physical or homeostatic parameters.[9-11] Giving patients the opportunity to participate in clinical trials from home is intriguing, especially for debilitating disorders like ALS. As ALS leads to progressive functional loss, remote monitoring of physical activity (i.e. *accelerometry*) could be an inexpensive method to assess a patient's progression rate objectively. Using remote markers of disease progression may reduce the overall trial burden and potentially allow more patients to participate in clinical trials.

Currently, there are no data regarding accelerometer-based monitoring in patients with ALS. It remains, therefore, unknown whether accelerometry can accurately reflect disease progression or whether it can improve current trial endpoints. In this study, we provide an initial step towards validating remote monitoring of disease progression in patients with ALS.

Methods

Study population and procedures

Patients were recruited from the Treatment Research Initiative to Cure ALS (TRICALS) database. The TRICALS database is a web-based international patient registry for patients with motor neuron disease (MND). The database holds approximately 300 to 350 active Dutch patients at any given time. For this study, all active TRICALS patients were approached by e-mail and invited to participate. Patients were required to have a diagnosis within the MND spectrum (i.e. ALS, progressive muscular atrophy (PMA) or primary lateral sclerosis (PLS)); no additional eligibility criteria were applied. The study physician (R.P.A.v.E) reviewed the medical records for all participating

patients to confirm their diagnosis and to classify patients into five prognostic groups according to the ENCALS survival model, as described elsewhere.[22] Subsequently, patients were sent the ActiGraph GT9X Link (ActiGraph LLC, Pensacola, FL), a small (0.5 × 3.5 × 1 cm), lightweight (14 g) tri-axial accelerometer. The ActiGraph was worn on the right hip in the anterior axillary line using a belt clip during waking hours for 7 days. It was initialized to collect data at a sampling rate of 30 Hz. In addition, patients were asked to keep a wear time log and to provide information regarding their daily functioning (revised ALS functional rating scale, ALSFRS-R), weight and mood (Hospital Anxiety and Depression Scale, HADS). All study materials were sent and returned by mail every two to three months for a maximum of seven measurements (T0 – T6). This study was approved by the Medical Ethics Committee of the UMCU (16/606). All study participants gave written informed consent to be approached digitally for research purposes and consented to participate in this study.

Accelerometer data

The ActiLife (version 6.13.3) software was used to extract the raw accelerometer data from the ActiGraph. The 30 Hz data were summarized in 10-second epochs with application of the Low Frequency Extension (LFE) algorithm. The LFE algorithm increases the sensitivity for capturing lower-intensity activities (e.g. sleeping, studying or watching television), which was hypothesized to be of relevance for elderly and neurologically impaired patients.[12, 13] Raw accelerometer files were processed to remove the recorded activity during mail transportation and to identify *non-wear* periods. Non-wear periods were identified using the non-wear time classification algorithm reported by Choi L *et al.*[14] Due to extremely low activity levels of far progressed patients (e.g. ALSFRS-R < 15), we defined a *non-wear* period conservatively as a consecutive period of no activity for 150 minutes. Finally, we calculated per day the *total wear time* in hours. In order to obtain an accurate estimate of the mean activity during a day, days with less than eight hours of total wear time were excluded from the analysis.

Accelerometer-based outcomes

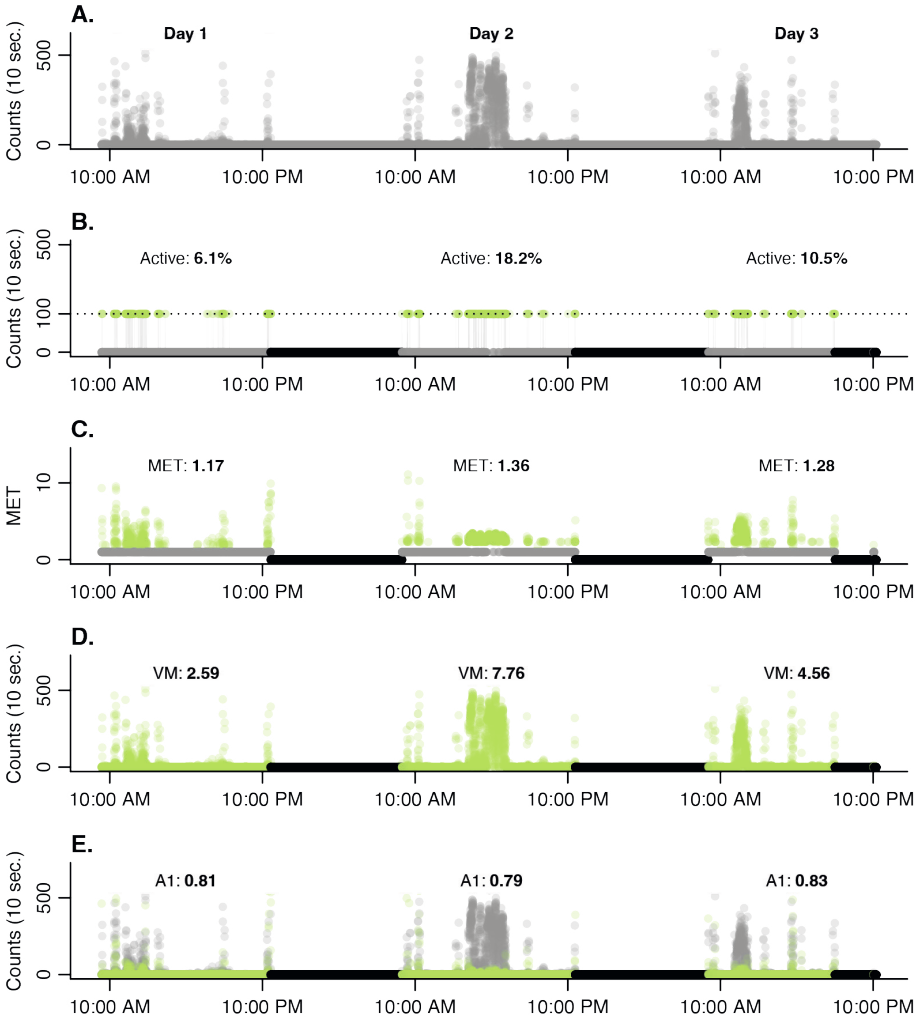
The summarized and processed accelerometer data consists of approximately 10,000 to 30,000 observations per measurement (**Figure 6.1A**). Observations are expressed as activity counts per 10 seconds. An activity count is based on the *vector magnitude*, i.e. the squared sum of the tri-axial data: $\sqrt{x^2 + y^2 + z^2}$, where *x*, *y*, *z* are the vertical, forward and sideway axes, respectively. Due to the extent of the data, we defined four different outcomes to summarize daily activity into a single value: (1) %active; (2)

MET score; (3) daily VM and (4) daily A1 (**Figure 6.1B-E**). For the *%active*, we estimated the proportion of the vector magnitude counts that exceeded the 100 counts per minute threshold (**Figure 6.1B**), which is consistent with more than sedentary activity (sedentary <100, light < 760, moderate-to-vigorous < 2020 and vigorous ≥ 2020 counts per minute).[15] For the second outcome, we translated the vector magnitude counts to Metabolic Equivalent of Task estimates (MET), which were summarized by calculating the average daily *MET score* (**Figure 6.1C**).[16] Both summary statistics, %active and MET score, are reflecting the average daily activity of a patient, i.e. what a patient *does* during the day. They do, however, not directly indicate what a patient physically *can* do. This additional information can be partially extracted from the variation in vector magnitude counts. If the variation in vector magnitude counts is large, but the average vector magnitude small, this indicates that the patient is physically capable of making strong (i.e. high vector magnitude) movements, but chooses not to (e.g. due to a lack of motivation or fatigue). We defined, therefore, an outcome based on the average daily vector magnitude count and its variation (*daily VM,* i.e. average x standard deviation of the vector magnitude, **Figure 6.1D**). The *daily VM* was estimated on the log (*ln*) scale due to its (zero-inflated) Poisson distribution. Similar to the daily VM, we evaluated a fourth outcome solely based on the variation in vertical axis (i.e. movement against gravity, *y*-axis), hereafter referred to as *daily A1* (**Figure 6.1E**). From **Figure 6.1** it becomes clear that each endpoint results in different daily summaries with important differences in day-to-day variation. For example, the %actives ranges in this illustration between 6.2% and 18.1%, while the daily A1 ranges only between 0.79 and 0.83. This could have important consequences for the sensitivity of each end point.

Sample size calculation

At the time of study initiation (Q4 2016), no data were available for accelerometer-based outcomes in patients with MND. We hypothesized that physical activity was strongly correlated with the functional status measured by the ALSFRS-R. The rate of decline and nuisance parameters (i.e. variance-covariance matrix) were, therefore, based on the longitudinal patterns of the ALSFRS-R total score using data from the LITRA study.[17] The longitudinal sample size calculation assumed, conservatively, a 12-month follow-up period with quarterly measurements. In total, 34 patients were needed to detect 0.60 points per month decline in ALSFRS-R (i.e. the lower 25th percentile of individual ALSFRS-R slopes) with 90% power and an alpha of 5%.[18] With an expected 10 - 20% attrition rate per year due to death or study withdrawal, we recruited 42 patients.

Figure 6.1. Raw accelerometer data of a single measurement and illustration of outcomes



Non-wear periods (black) were identified in raw accelerometer data (**A**). For the wear periods we defined four outcomes: (1) %active, (2) Metabolic Equivalent (MET), (3) vector magnitude (VM) and (4) A1. (**B**) %active; the activity count (y-axis) was split based on a 100 counts per minute cut-off and we calculated the proportion of being active (i.e. >100 counts, green).[15] (**C**) MET; the activity counts were recoded to MET (grey = MET 1) and averaged. When a patient is inactive (i.e. lying), the MET is 1 (grey).[16] (**D**) VM; the average daily activity count (mean; mu) was multiplied by the daily variation in activity counts (standard deviation; sd). (**E**) A1; instead of using the composite of three accelerometer axis (grey), we extracted only the vertical axis (i.e. movement against gravity; A1, green). The A1 was defined as the daily variation in the vertical axis (sd). The four outcomes resulted in different daily summaries with important differences in day-to-day variation. For example, the %active ranges from 6.1% to 18.2%, whereas the A1 ranges only from 0.79 to 0.83. This could have important consequences for the sensitivity to detect differential disease progression.

Statistical analysis

The primary aim of the analysis was to assess the longitudinal rates of decline in daily activity or disease progression. Linear mixed effects (LME) models were used to estimate the mean population rate of decline; all LME models were fitted with a fixed effect for time and a random intercept and slope for time per individual. To quantify the heterogeneity in rates of decline between individuals, we calculated the Coefficient of Variation (CoV) per outcome. The CoV was defined as the variation in slopes (i.e. random effect of time) divided by the mean rate of decline. Similar LME models were used to assess the longitudinal correlation between physical (accelerometer) activity and clinical markers of disease severity (i.e. ALSFRS-R or King stage, as estimated from the ALSFRS-R).[18, 19] We used standardized outcomes in the LME models in order to express the longitudinal associations as correlation coefficients with a similar interpretation as Pearson's *r*. Finally, we evaluated the effect of each endpoint on trial design based on longitudinal sample size calculations as described in more detail elsewhere.[18, 20] LME models were fitted using the R *lmer* function (lme4, version 1.1-18-1).[21] The R *PhysicalActivity* library (version 0.2-2, 2018) was used to process the raw accelerometer data.[14] Results were considered significant when alpha was less than 0.05.

Results

Patient population and feasibility

Between the 7th of October 2016 and the 1st of November 2018, forty-two Dutch patients participated in this prospective longitudinal cohort study; their baseline characteristics are given in **Table 6.1**. Despite the lack of eligibility criteria, the study population consisted primarily of patients with a relatively good prognosis.[22] The total follow-up time was 503.2 months; on average, each patient was observed for 12.0 months (interquartile range from 5.9 to 18.1) and produced an average of 4.9 measurements. A total of 15 patients died during follow-up (overall 18-month survival 71.5%, CI 58.4% to 87.4%). Patients rated the burden to wear the ActiGraph on a scale of 0 to 10, where 0 indicates no burden, as low: mean 1.3 (95% CI 0.7 – 1.9, range: 0 – 7). Three patients rated the burden ≥ 5: two patients were afraid to lose the ActiGraph, whereas one female patient reported limited clothing options (e.g. could not wear a dress). Overall, the burden was similar for males and females (p = 0.78). In total, 694 valid ActiGraph wear-time days were available for analysis with a total monitoring

period of 9,288 hours and a mean daily monitoring time of 13.4 hours/day. The wear-time adherence of 93.0% was excellent (694 ≥8-hour periods out of the 746 days).

Table 6.1. Characteristics of the patients at baseline

Characteristic	**Overall** *(n = 42)*
Age, mean (SD), years	60 (12)
Males, No. (%)	31 (74)
MND Subtype, No. (%)	
ALS	39 (93)
PMA	3 (7)
PLS	0 (0)
Bulbar onset, No. (%)	7 (17)
Symptom duration, *months*	
Median	25
Range	7 – 218
Diagnostic delay, *months*	
Median	8
Range	2 – 130
Riluzole use, No. (%)	30 (75)
Body-mass index, mean (SD), kg/m^2	25 (3)
ALSFRS-R total score, mean (SD)	36 (8)
ΔFRS, *points per month*	
Median	0.34
Range	0.05 – 1.24
Prognostic subgroup, No. (%)	
Very long	16 (38)
Long	14 (33)
Intermediate	11 (26)
Short	1 (2)
Very short	0 (0)

MND = motor neuron disease; ALS = amyotrophic lateral sclerosis; PMA = progressive muscular atrophy; ALSFRS-R = revised ALS functional rating scale; ΔFRS = 48 - ALSFRS-R score / disease duration.[23]

Longitudinal change in physical activity and daily functioning

Based on accelerometer data at baseline, patients were active or non-sedentary 27.9% (95% CI 24.8% - 31.1%) of the time (activity count > 100 per minute),[15] with a between-patient variability in baseline activity ranging from 11.6% to 52.4%. **Table 6.2** provides the baseline and longitudinal monthly rates of change (i.e. slope) for the ALSFRS-R and accelerometer-based outcomes. All outcomes exhibited a strong declining trend over time (all *p*-values < 0.001). The average monthly decline in ALSFRS-R was 0.59 points (95% 0.39 – 0.80); the average monthly decline in being active was 0.64%

(95% 0.43 – 0.86). In all outcomes, there was between-patient variability in the rate of decline (i.e. the presence of both fast- and slow-progressing patients, all *p*-values < 0.001). The between-patient variability, expressed as coefficient of variation (CoV), was lower in accelerometer-based outcomes than the ALSFRS-R; range 0.64 to 0.81 *vs.* 1.06. A lower CoV could positively affect sample size calculations and increase the sensitivity to detect differential disease progression.

Table 6.2. Longitudinal rates of change during follow-up

Outcome	Model parameters				Coefficient of variation
	Intercept	Slope [a]	95% CI [b]	*p*-value [b]	
ALSFRS-R					
Total score	36.4	-0.59	-0.80 – -0.39	< 0.001	1.06
Bulbar score	10.2	-0.13	-0.19 – -0.07	< 0.001	1.34
Motor score	15.2	-0.44	-0.60 – -0.28	< 0.001	1.10
Respiratory score	11.2	-0.09	-0.15 – -0.03	0.006	2.02
ActiGraph					
%Active	27.9	-0.64	-0.86 – -0.43	< 0.001	0.81
MET	1.71	-0.018	-0.024 – -0.013	< 0.001	0.64
Vector magnitude (VM)	8.55	-0.19	-0.25 – -0.14	< 0.001	0.77
Vertical axis (A1)	1.65	-0.029	-0.038 – -0.021	< 0.001	0.74

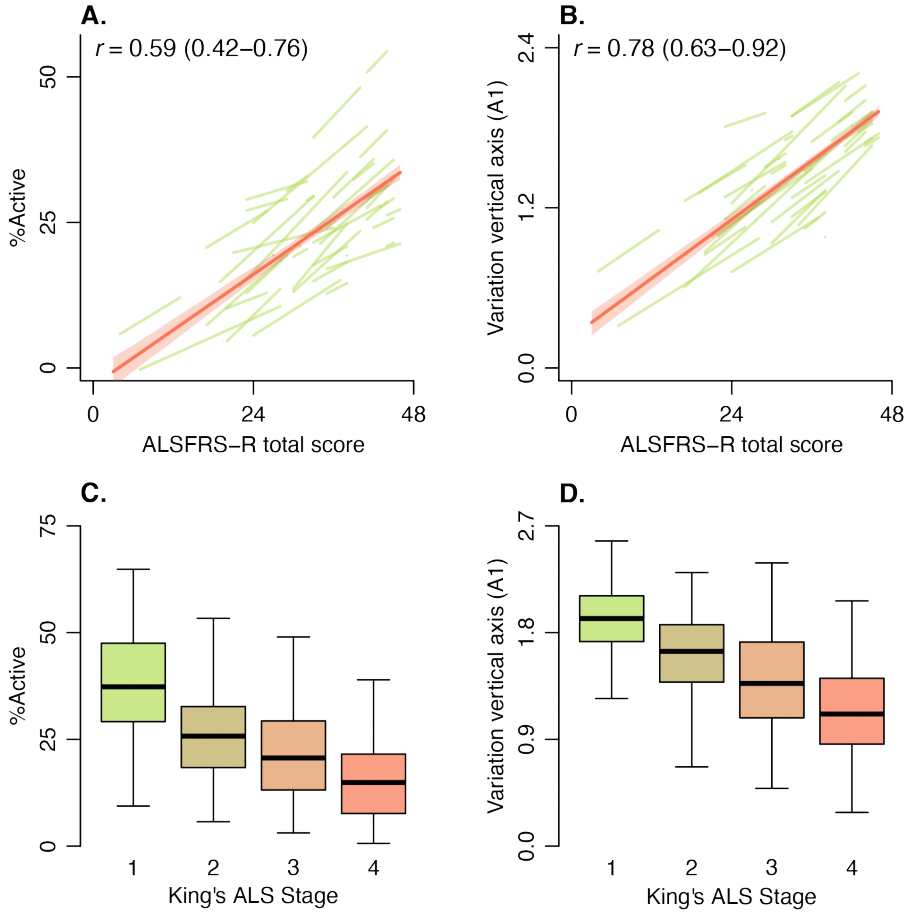
[a] Slope is the mean monthly rate of change during follow-up. [b] 95% CI and *p*-value of slope (indicating whether the rate of change is different from zero). Abbreviations: Coefficient of variation (CoV) = between-patient standard deviation of slope / mean rate of change, a lower value indicates that there is less variation among patients and disease progression can be detected more accurately; CI = confidence interval; ALSFRS-R = revised ALS functional rating scale; MET = metabolic equivalent. Linear mixed models were used to estimate the mean rate of change, CI, *p*-value and CoV.

Correlation with disease progression

Figure 5.2A-B reveals the correlation between the ALSFRS-R and accelerometer-based outcomes. The MET score had the lowest correlation with ALSFRS-R (*r* 0.57; 95% CI 0.43 – 0.71, *p* < 0.001), whereas the variation in vertical axis (i.e. movement against gravity; A1) had the strongest correlation (*r* 0.78, 95% CI 0.63 – 0.92, *p* < 0.001); the correlation for daily VM was *r* 0.75 (95% CI 0.59 – 0.92, *p* < 0.001). The motor domain (i.e. ALSFRS-R item 4-9) was the primary driver of the correlation between A1 and the ALSFRS-R (*r* 0.83; correlation bulbar, fine motor, gross motor and respiratory domains were *r* 0.69, *r* 0.75, *r* 0.74 and *r* 0.55, respectively). A similar association was observed with clinical stage as defined by the King's ALS staging algorithm (**Figure 6.2C-D,** *p* < 0.001).[19] Mean A1 levels were 1.82 (95% CI 1.70 – 1.94) for stage 1, 1.62 (95%

CI 1.51 – 1.72) for stage 2, 1.40 (95% CI 1.30 – 1.51) for stage 3 and 1.12 (95% CI 1.00 – 1.24) for stage 4, suggesting a near linear trend across King's ALS stages.

Figure 6.2. Correlation between accelerometer-based outcomes and disease progression



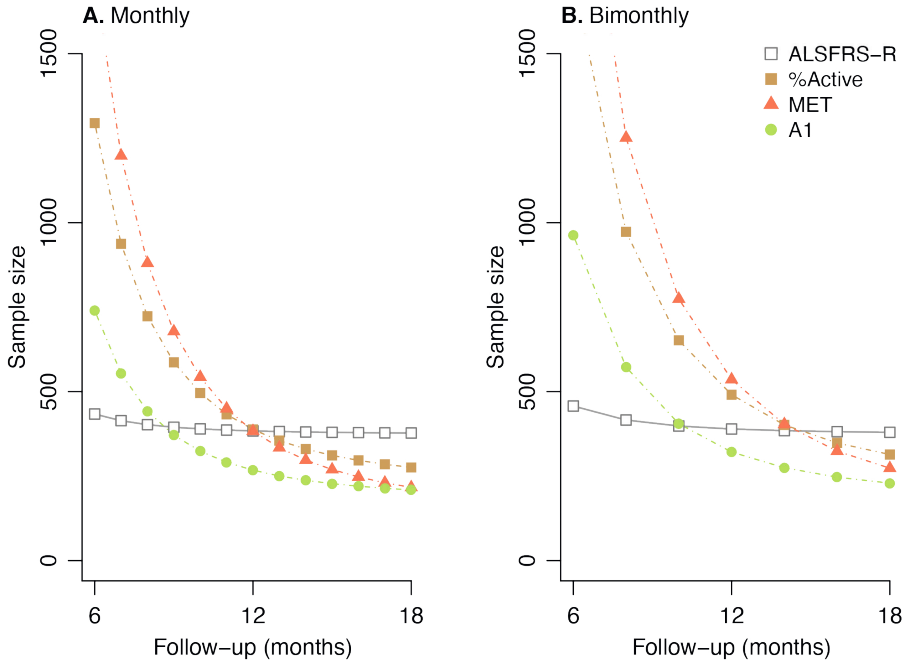
Longitudinal correlation between two accelerometer-based outcomes, %active (**A**) and variation in the vertical axis (A1, **B**), with the ALSFRS-R; r = Pearson correlation. The green lines represent the individual patient correlations. **C+D:** Distribution of %active and A1 within clinical stages defined by the King's ALS staging algorithm.[19]

Consequences for clinical trial design

Finally, for each outcome, we explored the required group size to detect a 25% improvement in the rate of decline with 90% power for various follow-up periods and sampling frequencies (**Table 6.2**). When a monthly sampling interval is used, the accelerometer-based daily VM and A1 outcome outperform the ALSFRS-R when

follow-up duration exceeds 9 months. At 12-months, a 30.3% reduction in sample size is achieved, which increases to 44.6% after 18 months (**Figure 6.3A**). When a bimonthly sampling interval is used, the ALSFRS-R is outperformed after 12 months, resulting in a 17.5% reduction at 12 months, which increases to 39.7% after 18 months (**Figure 6.3B**). The daily VM and A1 were superior outcomes in all settings when compared to the %active and MET. The difference between VM and A1 were minimal (5% to 7% difference in sample size).

Figure 6.3. Longitudinal sample size calculations for three accelerometer-based outcomes



Models from **Table 2** were used for the sample size calculation to detect a 25% reduction in slope with 90% power (per group). We evaluated different scenarios by varying the follow-up duration (x-axis) and using either a monthly (**A**) or bimonthly (**B**) sampling interval. The colors represent the different accelerometer-based endpoints. A1 = variation in vertical axis (i.e. movement against gravity). The sample size calculations are based on the observed slopes in **Table 6.2** and do not account for missing data.[20] It is important to note that in other settings, the absolute sample size varies, but is unlikely to affect the relative differences between outcomes (i.e. absolute sample size are high in this example due to the relatively slow rate of progression of the enrolled population).

Discussion

In this study, we show the feasibility and value of remote, accelerometer-based monitoring of disease progression in patients with ALS. Accelerometer-based outcomes accurately assess the patients' activity level under free-living conditions and provide an objective quantification of the disease progression rate across the disease spectrum. Accelerometry has a high adherence rate and may potentially lead to improvements in clinical trial design, not only by reducing sample sizes, but also by providing patients with the opportunity to participate in clinical trials from home. Reducing the number of hospital visits lowers the burden of trial participation and may better fit the physical abilities of patients with ALS. In the end, remote monitoring of disease progression may potentially increase the number of eligible patients, enhance enrollment rates and improve protocol adherence in ALS clinical trials.

The use of accelerometers, biosensors and medication adherence monitors is receiving increasing interest across all fields of medicine.[5] Digital heath monitoring could increase the efficiency of clinical trials, reduce costs, better reflect patient functioning and evaluate treatment responses in real-world settings.[24-26] Most importantly, remote monitoring of trial participants maximizes the collection of information outside clinical visits and could make in-clinic visits superfluous. This will pave the way to a patient-centric clinical trial model, where the trial is designed around the patient rather than fitting the patient into a clinical infrastructure. Despite these clear advantages, digital biomarkers are not frequently implemented in pivotal clinical trials.[5] Apart from the data complexity and potential ethical limitations,[24] regulatory hurdles may be the main driver of their delayed utilization.[25] The limited standardization of the data capture, auditability and use for digital biomarkers may result in a lower level of consistency and quality as compared to in-clinic measured endpoints. To overcome these hurdles, it is imperative to obtain insight into longitudinal patterns and confirm that digital biomarkers are valid surrogates for classical endpoints.[5, 25]

Interestingly, we found a considerable degree of variation between the four methods to summarize daily activity and their ability to detect treatment responses (**Figure 6.3**). Our results indicate that the daily VM and A1 are more suitable as accelerometer-based endpoints for clinical trials as compared to the %active and MET. An important consideration is to distinguish between accelerometer-based outcomes that reflect what a patient *does* during the day (i.e. the percentage being active or the mean accelerometer count) with those that reflect what a patient *can* do (i.e. the

variation in accelerometer counts like the A1). Although there is a decline in average daily activity as ALS progresses, the endpoints are affected by intrinsic patient-level characteristics, such as culture, life-style and motivation. Similar to a control population,[15] there is a wide variability between patients in the amount of physical activity. By estimating the variability in daily physical activity, one can obtain an estimate of the range of activities a patient is capable of. Our results indicate that end points based on the variation in daily activity levels (e.g. VM or A1) have reduced between-patient variability and an increased sensitivity to detect differential disease progression.

Our study provides an important step towards validating remote monitoring of disease progression in patients with ALS. ALS is pathognomonic for the loss of motor neurons and the accompanying loss of muscle strength and function. The extensive heterogeneity between patients complicates the quantification of disease progression. This affects the design of clinical trials and their ability to detect treatment responses. In addition, the most common marker of disease progression, the ALSFRS-R, is affected by multidimensionality, which may prevent a sensitive assessment of the disease progression rate. Our results indicate that accelerometer-based outcomes approximate the ALSFRS-R, but have considerably less between-patient variability over time. This increases the sensitivity to detect treatment responses and may potentially lead to reductions in sample size and costs for mid- to long-term trials.

Our study has several limitations that should be considered. Similar to the white coat syndrome,[27] activity monitoring might be affected by the Hawthorne effect (i.e. an alteration of behavior due to the awareness of being observed).[28] This effect is problematic if one is interested in the average daily activity, but unlikely to bias the estimated progression rate. Our results indicate that high-frequent or continuous monitoring may further increase the sensitivity of accelerometer-based outcomes. In addition, we only used 0.01% of the available data and a (Bayesian) modeling approach to define the entire dataset may significantly improve outcomes. Interestingly, when comparing the VM (3 axes) and the A1 (only vertical axis), the vertical axis (i.e. anti-gravity movement) seems to be the most important axis for quantifying the rate of disease progression. There was no clear benefit of incorporating additional information from the forward and sideway axes (i.e. rotatory or sideways movements). This may be an important observation, as this may indicate that these axes hold limited information and may inflate the noise in the data. Moreover, and despite the

strong correlations with functional measurements observed in our study, it remains essential to validate accelerometry prospectively with other key efficacy endpoints in ALS clinical trials, such as survival, muscle strength and lung function.

As a final note, ALS disease progression is not solely defined by a loss of (gross) physical functioning, but also involves progression in domains such as bulbar, fine motor, respiratory and cognitive functioning. To fully quantify ALS disease progression there is, therefore, a need to objectively assess multiple domains in ALS, e.g. using speech analysis (e.g. from Aural Analytics), apps evaluating fine motor tasks, actigraphy for gross motor functioning, and remote devices for body composition (e.g. percentage fat and fat-free mass) or pulmonary function. Besides the value of digital technology for clinical trials, the implementation of these technologies could also significantly benefit the delivery of care. Clinical decision algorithms based on remote data sources may be integrated into current healthcare systems to personalize clinical visiting schemes, or to optimize the detection (or prediction) of events such as respiratory failure or wheelchair dependency. In the end, these developments could significantly add to current methods of rating disease progression such as the ALSFRS-R.

In conclusion, in this study, we show the feasibility and value of remote monitoring of disease progression in patients with ALS. Accelerometry provides a non-invasive, remote and objective method for quantifying disease progression and correlates strongly with current outcomes. Accelerometer-based outcomes have the potential to be used as efficacy endpoint and may improve the efficiency of clinical trials. Remote monitoring provides patients with the opportunity to participate in clinical trials from home, paving the road to a patient-centric clinical trial model.

References

1. Chio A, Canosa A, Gallo S, et al. ALS clinical trials: do enrolled patients accurately represent the ALS population? Neurology 2011;77:1432-1437.
2. Bedlack RS, Pastula DM, Welsh E, Pulley D, Cudkowicz ME. Scrutinizing enrollment in ALS clinical trials: room for improvement? Amyotroph Lateral Scler 2008;9:257-265.
3. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365:82-93.
4. van Eijk RPA, Westeneng HJ, Nikolakopoulos S, et al. Refining eligibility criteria for amyotrophic lateral sclerosis clinical trials. Neurology 2019.
5. Perry B, Herrington W, Goldsack JC, et al. Use of Mobile Devices to Measure Outcomes in Clinical Research, 2010-2016: A Systematic Literature Review. Digit Biomark 2018;2:11-30.
6. Steinhubl SR, Waalen J, Edwards AM, et al. Effect of a Home-Based Wearable Continuous ECG Monitoring Patch on Detection of Undiagnosed Atrial Fibrillation: The mSToPS Randomized Clinical Trial. JAMA 2018;320:146-155.
7. Walker PP, Pompilio PP, Zanaboni P, et al. Telemonitoring in Chronic Obstructive Pulmonary Disease (CHROMED). A Randomized Clinical Trial. Am J Respir Crit Care Med 2018;198:620-628.
8. Zhan A, Mohan S, Tarolli C, et al. Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity: The Mobile Parkinson Disease Score. JAMA Neurol 2018;75:876-880.
9. Lipsmeier F, Taylor KI, Kilchenmann T, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018;33:1287-1297.
10. Garg SK, Shah VN, Akturk HK, Beatson C, Snell-Bergeon JK. Role of Mobile Technology to Improve Diabetes Care in Adults with Type 1 Diabetes: The Remote-T1D Study iBGStar((R)) in Type 1 Diabetes Management. Diabetes Ther 2017;8:811-819.
11. Arends J, Thijs RD, Gutter T, et al. Multimodal nocturnal seizure detection in a residential care setting: A long-term prospective trial. Neurology 2018;91:e2010-e2019.
12. Evenson KR, Wen F, Herring AH, et al. Calibrating physical activity intensity for hip-worn accelerometry in women age 60 to 91 years: The Women's Health Initiative OPACH Calibration Study. Prev Med Rep 2015;2:750-756.
13. Korpan SM, Schafer JL, Wilson KC, Webber SC. Effect of ActiGraph GT3X+ Position and Algorithm Choice on Step Count Accuracy in Older Adults. J Aging Phys Act 2015;23:377-382.
14. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011;43:357-364.
15. Matthews CE, Keadle SK, Troiano RP, et al. Accelerometer-measured dose-response for physical activity, sedentary time, and mortality in US adults. Am J Clin Nutr 2016;104:1424-1432.
16. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DR, Jr. Refined two-regression model for the ActiGraph accelerometer. Med Sci Sports Exerc 2010;42:1029-1037.
17. Verstraete E, Veldink JH, Huisman MH, et al. Lithium lacks effect on survival in amyotrophic lateral sclerosis: a phase IIb randomised sequential trial. J Neurol Neurosurg Psychiatry 2012;83:557-564.
18. van Eijk RPA, Eijkemans MJC, Ferguson TA, Nikolakopoulos S, Veldink JH, van den Berg LH. Monitoring disease progression with plasma creatinine in amyotrophic lateral sclerosis clinical trials. J Neurol Neurosurg Psychiatry 2018;89:156-161.
19. Balendra R, Jones A, Jivraj N, et al. Estimating clinical stage of amyotrophic lateral sclerosis from the ALS Functional Rating Scale. Amyotroph Lateral Scler Frontotemporal Degener 2014;15:279-284.

20. Liu G, Liang KY. Sample size calculations for studies with correlated observations. Biometrics 1997;53:937-947.

21. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. 2015 2015;67:48.

22. Westeneng HJ, Debray TPA, Visser AE, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. Lancet Neurol 2018;17:423-433.

23. Kimura F, Fujimura C, Ishida S, et al. Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. Neurology 2006;66:265-267.

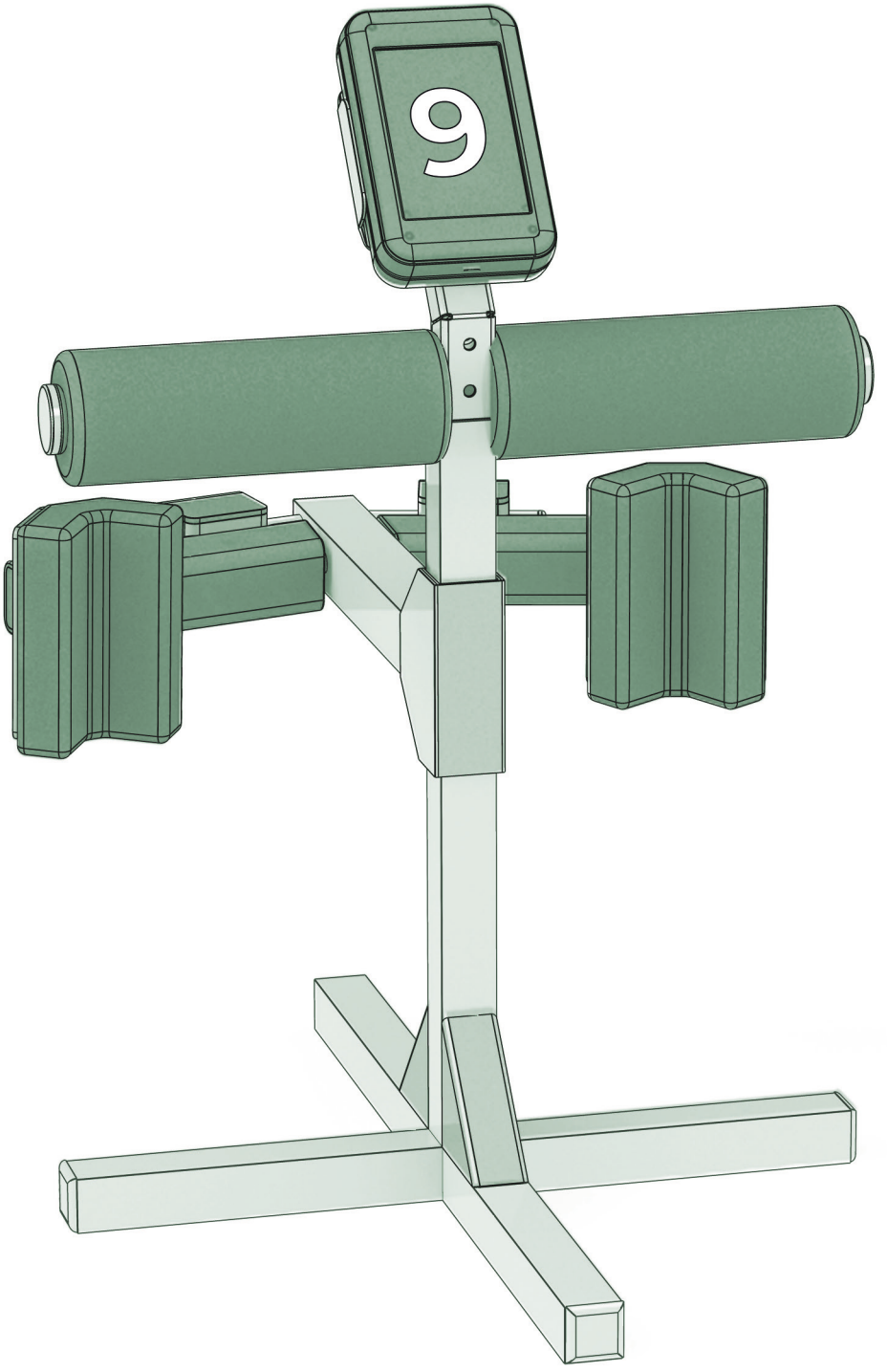
24. Gold M, Amatniek J, Carrillo MC, et al. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. Alzheimers Dement (N Y) 2018;4:234-242.

25. Kothare PA, Jadhav PR, Gupta P, Harrelson JC, Dickmann L. Harnessing the Potential of Emerging Digital Health and Biological Sampling Technologies for Clinical Drug Development: Promise to Reality. Clin Pharmacol Ther 2018;104:1125-1135.

26. Steinhubl SR, McGovern P, Dylan J, Topol EJ. The digitised clinical trial. Lancet 2017;390:2135.

27. Pioli MR, Ritter AM, de Faria AP, Modolo R. White coat syndrome and its variations: differences and clinical impact. Integr Blood Press Control 2018;11:73-79.

28. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. BMJ 2015;351:h4672.


9

Chapter 9

HOME-MONITORING OF VITAL CAPACITY IN PEOPLE WITH MOTOR NEURON DISEASE

Jochem Helleman
Jaap Bakers
Evelien Pirard
Leonard van den Berg
Anne Visser-Meily
Anita Beelen

Journal of Neurology 2021

Abstract

Background: Home-monitoring of spirometry has the potential to improve care for patients with a motor neuron disease (MND) by enabling early detection of respiratory dysfunction and reducing travel burden. Our aim was to evaluate the validity and feasibility of home-monitoring vital capacity (VC) in patients with MND.

Methods: We included 33 patients with amyotrophic lateral sclerosis, progressive muscular atrophy or primary lateral sclerosis who completed a 12-week home-monitoring protocol, consisting of 4-weekly unsupervised home assessments of VC and a functional rating scale. At baseline, during a home visit, patients/caregivers were trained in performing a VC test, and the investigator performed a supervised VC test, which was repeated at final follow-up during a second home visit. Validity of the unsupervised VC tests was evaluated by the differences between supervised and unsupervised VC tests, and through Bland-Altman 95% limits-of-agreement. Feasibility was assessed by means of a survey of user-experiences.

Results: The 95% limits-of-agreement were [−14.3; 11.7] %predicted VC, and 88% of unsupervised VC tests fell within 10 %predicted of supervised VC. 88% of patients experienced VC testing as easy and not burdensome, however, 15% patients did not think their VC test was performed as well as in the clinic. 94% of patients would like home-monitoring of VC in MND care.

Discussion: Unsupervised VC testing at home, with prior face-to-face training, is a valid and time-efficient method for the remote monitoring of respiratory function, and well-accepted by patients with MND and their caregivers.

Introduction

In patients with a motor neuron disease (MND), respiratory failure is the main cause of death (1,2). When patients show signs or symptoms of respiratory dysfunction, as described in clinical guidelines, non-invasive ventilation (NIV) is recommended (3–5). Studies have shown that the use of NIV prolongs survival and improves quality of life (6–8). Regular monitoring of respiratory function is essential to ensure timely detection of respiratory dysfunction so that NIV can be initiated (3–5). In current MND healthcare, respiratory function is monitored during regular visits to a multidisciplinary clinic. Two drawbacks of this type of monitoring are that clinic visits can be time consuming and burdensome for patients with MND, and that patients have to visit the clinic irrespective of whether there is a decrease in respiratory function (9,10). This suggests that current respiratory monitoring may be insufficiently tailored to the needs of patients.

A potential solution is the home-monitoring of respiratory function through the use of telehealth. This approach allows for more frequent assessments, higher continuity of monitoring, especially when patients are not able to visit the clinic, and easy communication between patients and the multidisciplinary care team (11–16). The use of telehealth may help to detect respiratory function decline early, and schedule clinic visits and initiate clinical interventions on time. One method of home-monitoring is the assessment of patient-reported symptoms of dyspnea, which was found to be useful for screening whether patients with MND had reduced vital capacity (VC)(17). However, a drawback of dependence on self-reported dyspnea/orthopnea is that patients with low VC but without symptoms will not be identified (false negative rate = 14%). For this reason, combining patient-reported symptoms of dyspnea with home-based VC testing may reduce false negative findings and improve the home-monitoring of respiratory function.

The VC test has prognostic value in patients with MND (18,19), and is easy to perform with a handheld spirometer, which is affordable and widely available. For these reasons, the VC test is suitable and relevant for home-monitoring; however, in MND care, its application for home-monitoring is still lacking (20). Recently, a study showed that during COVID-19, it was feasible to perform home-based VC testing with supervision via video and that it was well-received by patients with MND in a healthcare setting (21). However, one trial showed that when patients performed VC tests at home without supervision, the remote VC measurement was significantly higher than the usual in-clinic VC measurement and compliance was suboptimal

(22). These findings show the potential of home-monitoring of VC, but also indicate that more evidence is needed to support its implementation.

The aim of the present study is, therefore, to evaluate the validity and feasibility of unsupervised home-monitoring of VC in patients with MND.

Methods

Study design and population

This prospective cohort study aimed to include patients with MND, who were 18 years old or over and had access to a smartphone or tablet. Different diagnoses of MND were involved: amyotrophic lateral sclerosis (ALS), progressive muscular atrophy (PMA) and primary lateral sclerosis (PLS). The exclusion criteria were the use of non-invasive ventilation during the daytime, tracheostomy, or the inability to perform a VC test with or without caregiver assistance. Ethics approval from the Medical Ethics Committee of the University Medical Center Utrecht was obtained prior to the start of the study and patients gave their informed consent before participating.

Setting and procedure

This study was conducted by the University Medical Center in Utrecht, the Netherlands, in collaboration with the Revant Center for Rehabilitation in Breda. Both centers have a multidisciplinary care team, coordinated by a physician. All study activities were performed at the patients' homes, meaning that patients could participate in the present study without having to visit a multidisciplinary clinic. Patients who, between August 2020 and February 2021, received MND care from the multidisciplinary care teams were invited by the treating physician to participate. Most patients had access to the telehealth service *ALS Home-monitoring and Coaching* as part of their usual care. This telehealth service included the mobile ALS app for self-monitoring and messaging, which facilitated remote monitoring and communication between the patient and the multidisciplinary care team. A full description of *ALS Home-monitoring and Coaching* is available in a previous publication (12).

Study assessments

Respiratory function was assessed by making three attempts to perform the vital capacity (VC) test in upright position, using a low-cost (ca. €100,-) handheld spirometer with Bluetooth connection to a mobile app (Spirobank Smart®, Medical International Research, Italy). The VC tests were performed with a full-face mask

(Figure 1) to enable testing in patients with bulbar impairment (23). Patients recorded the time, date and VC test scores on a paper form, and also sent the VC test scores digitally to their multidisciplinary care team via the ALS app or by email, which allowed members of the multidisciplinary ALS care team to monitor respiratory function. The revised ALS functional rating scale (ALSFRS-R) was used to assess functional impairment (17,24), and was self-monitored monthly as part of *ALS Home-monitoring and Coaching*. Patients who did not use telehealth completed the ALSFRS-R on paper at every follow-up. We created a survey to evaluate user-experiences of patients and caregivers who assisted with VC testing; see tables 2 and 3 for the items of the survey. Items were scored on a 5-point Likert scale: the extent to which patients/caregivers considered aspects of VC testing to be difficult (answer options: Very easy – Very difficult), or the extent to which they agreed with a statement on VC testing (answer options: Totally agree – Totally disagree).

Figure 1 Performing a vital capacity test with a full-face mask. Left: A hammer grip around the tube. Right: Holding the mask with the tube placed between the fingers

Baseline protocol

At baseline (T0), the supervised VC test and ALSFRS-R were completed during a home visit. The investigator helped patients to install the mobile app on their smartphone, after which the supervised VC was performed. VC tests were either performed forcefully (FVC) or slowly (SVC), depending on which method was most effective/suitable for the patient (19). Patients (and their caregivers) were instructed on how to perform the VC test independently, and practiced VC testing. If required, the investigator gave tips on how to improve the way the VC test was performed. When proper technique was observed (e.g. correct placement of mask, maximal

in and exhalation, upright body position), the investigator left the room, and the patient performed an unsupervised VC test, to ensure that patients were able to do this without supervision. Unsupervised VC tests that were performed during the baseline home visit, were not included in the analysis.

Follow-up protocol

The total follow-up period was 12 weeks, with 4-weekly unsupervised assessments. One day after the home visit (T1), patients completed their baseline unsupervised VC tests. At 4 weeks (T2), 8 weeks (T3) and 12 weeks (T4) after baseline, patients completed unsupervised VC tests and the ALSFRS-R. The investigator sent a reminder on the days of follow-up either via text-message or e-mail, depending on patient preference. At T4 the investigator visited the patient's home at least 1 hour after patients had completed their unsupervised VC tests. During this final home visit, a supervised VC test was performed and patients (and their caregivers) were asked to fill in the survey on user-experiences and to indicate the average duration of their VC testing sessions.

Analyses

The highest VC test score, out of three attempts, at each time-point was converted to a percentage of the predicted (%predicted) VC, using height, age, and ethnicity (reference values used from Global Lung Function Initiative 2012)(25,26). We used the unsupervised test at T1 as baseline, since the unsupervised VC tests performed at T0 may have been affected by the prior supervised VC tests. Validity of unsupervised VC testing was assessed through the Bland-Altman 95% limits of agreement and Lin's concordance correlation coefficient (CCC) between the supervised VC at T0 and the unsupervised VC at T1, and between the supervised and unsupervised VC at T4. Based on clinical experience, we considered a maximal difference of 10 % predicted between supervised and unsupervised VC as an acceptable limit of agreement, since this will allow healthcare professionals to determine a trend of VC over time when the VC is monitored at 4-weekly intervals. Additionally, the coefficient of variation of supervised VC testing in patients with MND was already 6.3 %predicted in a previous study (27). A paired t-test was conducted to assess the change in supervised and unsupervised VC between T0 and T4, and whether there was a systematic difference between supervised and unsupervised VC. Furthermore, we evaluated whether the agreement between supervised and unsupervised VC was different after 12 weeks of home-monitoring compared to baseline. In order to obtain insight into the variation in unsupervised VC testing over time, we used linear regression to determine the

average slope over the 12 week period for each individual patient, and we calculated the standard error (SE) which indicates to what extent the VC values deviate from the linear regression line. We then ranked patients from lowest to highest SE and created a subgroup for each quartile (25%) of patients. These subgroups were used to create 4 separate plots for the longitudinal unsupervised VC data of individual patients to facilitate interpretation of the data. Furthermore, in the Bland-Altman plots, the subgroups are indicated for each data point (i.e. patient), in order indicate whether greater variability showed larger differences between unsupervised and supervised VC. An alpha of <0.05 was considered to be statistically significant. Feasibility of unsupervised home-based VC testing was determined through the adherence to the 4-weekly VC protocol, time cost of VC testing and user-experiences. Unsupervised VC testing was considered feasible when ≥75% of all unsupervised VC tests had been carried out, and each testing session completed within 20 minutes. An item of the user-experience survey was considered feasible when ≥75% of patients answered '(totally) agree' on positive statements, '(totally) disagree' on negative statements, and '(very) easy' on difficulty statements.

Results

We included 33 patients with MND, with an average age of 60.5 years, 79% of whom were male. 76% were diagnosed with ALS, 15% with PMA and 9% with PLS, and 78.8% had spinal onset. At baseline, 3 patients were on nightly NIV, and one patient started with nightly NIV during the study period. Most patients (88%) used telehealth as part of their usual care. All baseline patient characteristics are presented in Table 1. Nine patients were assisted with VC testing by a partner (N=4), family member (N=3) or a home nurse (N=2). The mean change over the 12-week period for the ALSFRS-R total score was −2.1 points.

Table 1 Baseline patient characteristics

Characteristic	**Patients** (N=33)
Gender (male), *n(%)*	26 (78.8)
Age (years), *mean(SD)*	60.5 (13.2)
Diagnosis, *n(%)*	
ALS	25 (75.8)
PMA	5 (15.2)
PLS	3 (9.1)
Site of onset, *n(%)*	
Bulbar	7 (21.2)
Spinal	26 (78.8)
Nightly NIV, *n(%)*	3 (12.1)
Gastrostomy, *n(%)*	2 (6.1)
Telehealth use, *n(%)*	29 (87.8)
Respiratory function (% of predicted VC), *mean (SD)*	78.4 (25.6)
Disease duration from first symptoms (months), *median (IQR)*	35.6 (17.2-52.2)
ALSFRS-R, *mean (SD)*	35.9 (7.3)
ALSFRS-R (respiratory domain), *mean (SD)*	11.0 (1.3)

ALS amyotrophic lateral sclerosis, PMA progressive muscular atrophy, PLS primary lateral sclerosis, NIV non-invasive ventilation, VC vital capacity, SD standard deviation, IQR interquartile range, MND motor neuron disease, ALSFRS-R revised ALS functional rating scale

Validity of unsupervised VC testing

The 95% limits of agreement and the mean difference were [−15.1; 15.4] and 0.12 %predicted (p=0.928) at baseline, respectively, and [−14.3; 11.7] and –1.33 %predicted (p=0.259) at final follow-up, respectively (Figure 2). The difference between unsupervised and supervised VC was smaller than 10 % predicted in 28 of 33 (85%) patients at baseline and in 29 of 33 (88%) patients at final follow-up. The median absolute difference between supervised and unsupervised VC at baseline and final follow-up were 2.6 (IQR=1.3-7.8) and 4.1 (IQR=1.6-5.8) %predicted, respectively. Lin's CCC was excellent at baseline (0.953), as well as at final follow-up (0.971)(Figure 3). Between baseline and final follow-up both the supervised VC (Mean = -3.31, p=0.045) and unsupervised VC (-4.77, p=0.036) decreased significantly. We also compared the change in supervised and unsupervised VC between baseline and final follow-up, which showed a good correlation (ρ=0.74, p<0.001). The plots of individual unsupervised VC data can be found in Figure 4,where the range of SE was 0.36-0.96 %predicted for the first quartile of patients, 1.02-2.16 %predicted for the second

quartile of patients, 2.28-3.98 %predicted for the third quartile of patients, and 4.56-10.47 %predicted for the fourth quartile of patients.

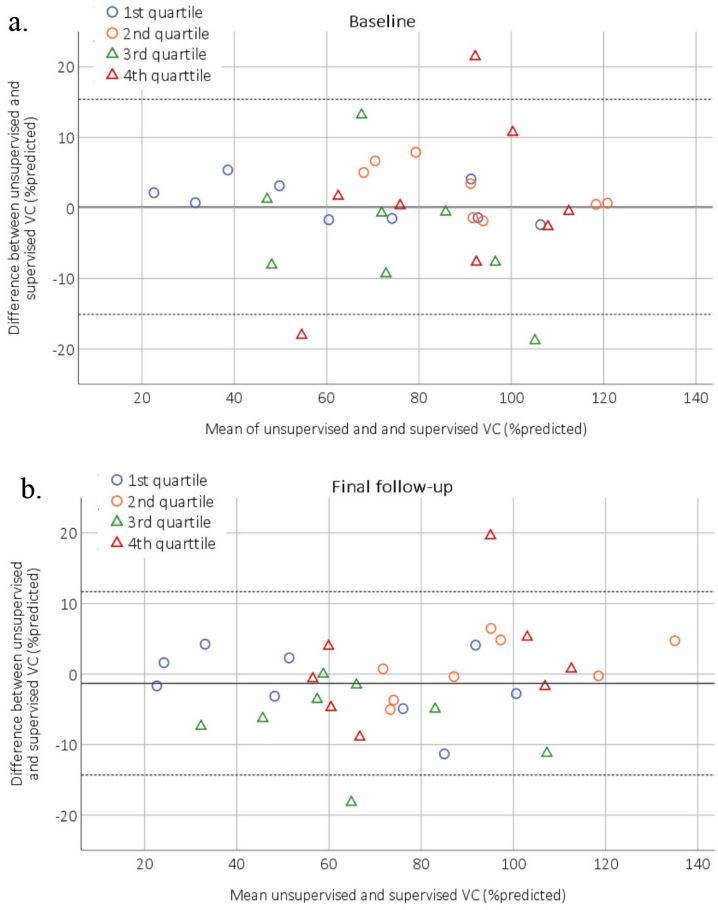


Figure 2 Bland–Altman plots. VC = vital capacity, Dashed line = 95% limits of agreement. The 4 quartile groups are based on thevariability of the unsupervised VC scores over time, where 1st quartile = lowest variability and 4th quartile = highest variability. **a.** At baseline, **b.** at final follow-up

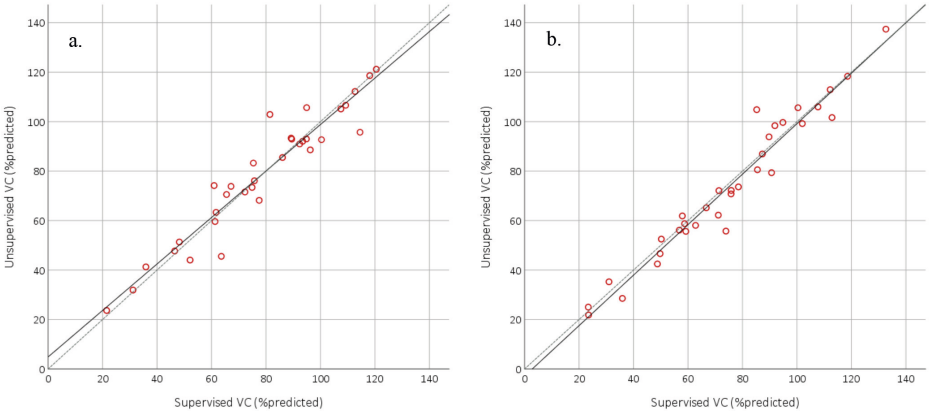


Figure 3 Scatterplot of unsupervised vs supervised vital capacity. VC vital capacity. Dashed line= line of identity. **a** At baseline, Lin's CCC =0.953, **b** at final follow-up, Lin's CCC=0.971

Feasibility of home-monitoring

All 33 participants completed 100% of their VC assessments, 32 (97%) within 20 min, and 29 (88%) within 15 min. Patients reported that the spirometer and spirometry app were user-friendly, and that unsupervised VC testing was considered to be easy and not burdensome (Table 2). Most patients (30, 93.8%) would like their respiratory function from home for care purposes. Even patients with limited hand function were able to handle the spirometer and independently perform a VC test, as 29% (7/24) of patients who were not assisted by a caregiver had an ALSFRS-R fine motor score of ≤6. This was due to the fact that the face

mask, that was attached to the mouthpiece of the spirometer, made it easier to hold the spirometer. 3 patients experienced difficulties with determining whether a VC test was performed correctly and 2 patients felt insecure about their VC test performance in absence of a healthcare professional. Furthermore, 5 patients did not think that the unsupervised VC tests were performed as well as supervised tests in the clinic.

Based on the comments reported by patients during unsupervised VC testing, there were some difficulties that affected VC test performance: excessive mucus in throat (patient 4, Figure 4c at T2), physical fatigue (patient 17, Figure 4d at T1), pain in stomach caused by a gastrostomy tube (patient 24, Figure 4d at T4), not being able to concentrate during testing (patient 26, Figure 4b at T3), or physical discomfort due to an uncomfortable body position in wheelchair (patient 33, Figure 4c at T4).

Table 2 User-experiences of patients

Item	**(Very) easy** *n(%)*	**Neutral** *n(%)*	**(Very) Difficult** *n(%)*	**N***
Placing the mask on my face was	23 (82.1)	4 (14.3)	1 (3.6)	28
Handling the spirometer was	26 (92.8)	1 (3.6)	1 (3.6)	28
Starting a VC test in the app was	30 (96.8)	1 (3.2)	0 (0)	31
Performing a VC test was	29 (87.9)	3 (9.1)	1 (3)	33
Judging whether the test was performed correctly was	26 (78.8)	3 (9.7)	3 (9.7)	32
Item	**(Totally) agree** *n(%)*	**Neutral** *n(%)*	**(Totally) disagree** *n(%)*	**N***
The spirometer is user friendly	31 (93.9)	1 (3)	1 (3)	33
The spirometry app was user-friendly	30 (90.9)	3 (9.1)	0 (0)	33
The spirometer is appropriate for home-monitoring of respiratory function	30 (90.9)	3 (9.1)	0 (0)	33
I would like to monitor my respiratory function from home for care purposes	30 (93.8)	2 (6.3)	0 (0)	32
I know how to perform a VC test	33 (100)	0 (0)	0 (0)	33
I believe that my VC test at home is performed just as well as a usual VC test in the clinic	24 (72.8)	4 (12.1)	5 (15.1)	33
I am unsure about performing the VC test correctly in the absence of a healthcare professional	2 (6.5)	4 (12.9)	25 (80.6)	31
Performing VC tests at home is burdensome	2 (6.3)	2 (6.3)	28 (87.5)	32

*Missing data is due to patients answering “not applicable/ no opinion”, VC = vital capacity.

Most caregivers who assisted with VC testing reported that the spirometer (n=7) and mobile app (n=8) were user-friendly, and that helping with VC testing was easy (n=7) and not burdensome (n=7)(Table 3). The majority of caregivers believed they were able to (help) perform a VC test correctly (n=8), and judge whether a VC test had been performed correctly (n=7). Some of the caregivers (n=3) did not think that they performed the unsupervised VC as well as a healthcare professional in a clinic.

Table 3 User-experiences of caregivers

Item	**(Very) easy**	**Neutral**	**(Very) Difficult**
Placing the mask on his/her face was	8/9	1/9	1/9
Handling the spirometer was	8/9	0/9	1/9
Starting a VC test in the app was	8/8	0/8	0/8
Performing a VC test was	7/8	1/8	0/8
Judging whether the test was performed correctly was	7/9	2/9	0/9

Item	(Totally) agree	Neutral	(Totally) disagree
The spirometer is user friendly	8/9	1/9	0/9
The spirometry app is user-friendly	7/8	1/8	0/8
The spirometer is appropriate for home-monitoring of respiratory function	8/9	1/9	0/9
I would like to monitor my respiratory function from home for care purposes	8/9	0/9	1/9
I know how to (help) perform a VC test	8/9	1/9	0/9
I believe that my VC test at home is performed just as well as a usual VC test in the clinic	5/9	1/9	3/9
I am unsure about performing the VC test correctly in the absence of a healthcare professional	1/9	1/9	7/9
Helping to perform VC tests at home is burdensome	0/9	2/9	7/9

Missing data is due to caregivers answering “not applicable/ no opinion”, VC = vital capacity

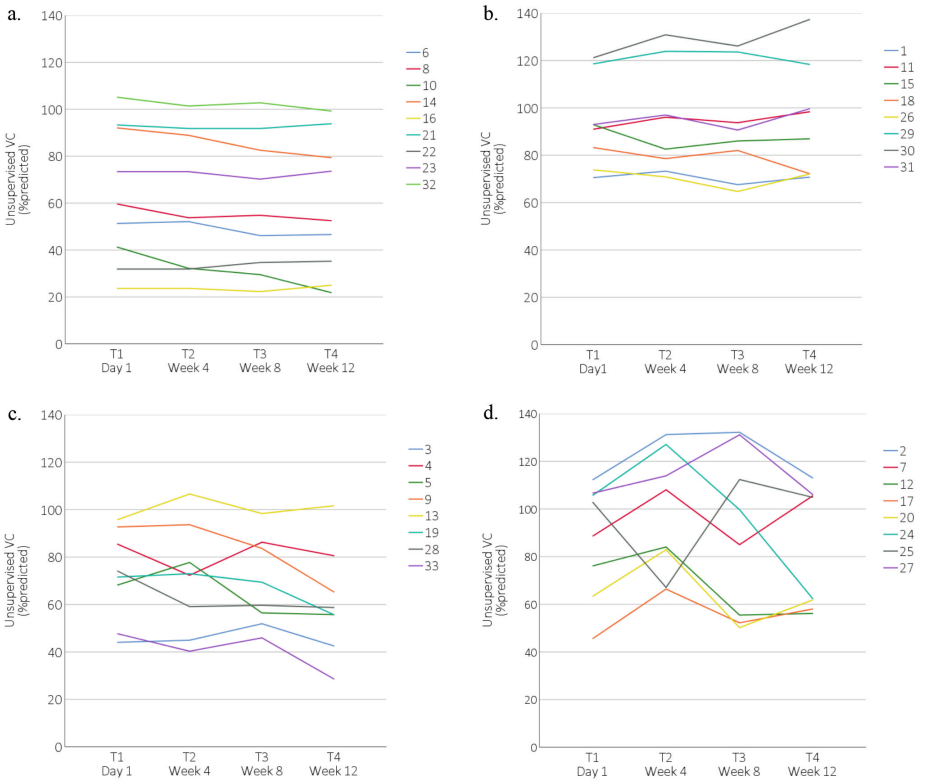


Figure 4 Unsupervised vital capacity over time per individual patient. VC = vital capacity. Patients were ranked from low to high variability, based on the standard error (SE) of the unsupervised VC scores over time and split into four quartiles (i.e. 25% of patients in each group). a) patients in the first quartile (SE range = 0.36–0.96%predicted), b) patients in the second quartile (SE range = 1.02–2.16 %predicted), c) patients in the third quartile(SE range = 2.28–3.98 %predicted), and d) patients in the fourth quartile (SE range = 4.56–10.47 %predicted)

Discussion

The present study showed that unsupervised home-monitoring of VC, after one face-to-face training, was a valid method for the remote monitoring of respiratory function in patients with MND. Furthermore, the 4-weekly home-monitoring of VC without supervision was feasible, since adherence was excellent, and most patients and caregivers experienced VC testing as easy and not burdensome. Lastly, patients and caregivers were motivated to continue with home-monitoring of VC in MND healthcare.

Our results on the validity and feasibility of unsupervised VC testing at home are promising and show that this can be a time-efficient method in MND care for both patients and healthcare professionals for remotely monitoring respiratory function. We provided insight into the variation in unsupervised VC testing over time, which showed that most patients had a stable trend of VC during the 12-week period. However, the course of the unsupervised VC of some patients were highly variable over time, and generally showed larger differences with the supervised VC, indicating that these patients may require additional supervision during home-monitoring, e.g. through video.

We found that there was no systematic difference between unsupervised and supervised VC, but at final follow-up we observed that supervised VC test scores were more likely to be higher than the unsupervised VC test scores, when compared to baseline. This may indicate that the performance of the unsupervised VC test decreases over time in some patients. This finding is in contrast to previous studies, which reported that remote VC assessments were systematically higher than usual in-clinic VC assessments (22,28). An explanation for this finding, is that in the present study all VC tests were performed at patients' homes, including the supervised tests. This limited the factors that may have negatively affected VC test performance, such as the burden of travelling and visiting a clinic.

We found that all patients adhered to the 4-weekly monitoring protocol, and that this frequency was acceptable. This corresponds to findings of a recent study, in which most patients reported that the highest acceptable frequency for remote respiratory assessments was monthly (21). In the present study, facilitating factors for adherence to VC testing at home were that the spirometer and app were user-friendly, and VC testing was easy, not burdensome and not time consuming. A previous study reported suboptimal adherence with a weekly VC protocol, mainly

due to connection problems and patients forgetting to complete measurements (22). During our study we were fortunate that the spirometer and app only rarely malfunctioned, which resulted in re-doing a VC test, but never prevented patients from testing. Furthermore, the problem of forgetting a VC test was tackled by sending a reminder at each follow-up. Another facilitator for adherence was the fact that home-monitoring of VC was part of an existing telehealth service and that VC test results were monitored by the multidisciplinary care team. Patients are likely to be more motivated to complete assessments at home, when they know healthcare professionals are monitoring their data closely and will provide feedback when necessary (29).

During unsupervised home-monitoring there were several factors, unrelated to respiratory muscle weakness, that hindered optimal VC test performance, such as pain, physical fatigue or loss of concentration. This suggests that it is important that patients provide comments on their physical and psychological well-being at time of VC testing, in order to help healthcare professionals interpret VC scores remotely. Moreover, some patients and caregivers experienced difficulties with determining whether a VC test was performed correctly, and felt insecure about proper VC test performance without supervision. These patients may prefer access to online instruction-videos (30) or require video-supervision during home-monitoring, which has been shown to be well-accepted by patients with MND (21,28). A disadvantage of video-supervised monitoring, is that it takes healthcare

professionals considerably more time, compared to unsupervised monitoring. Interestingly, one study reported that only a few patients felt they were able to perform a VC test at home without video supervision, which contrasts with our study sample, where the majority believed they were able to perform a VC test at home without supervision. A reason for this discrepancy may be that patients in the present study were trained in unsupervised VC, and that most patients already had experience with telehealth and remote monitoring.

Clinical implications

Our findings indicate that a single face-to-face training session prior to VC testing at home was sufficient for most patients to learn how to perform a VC test independently. In clinical practice, patients could be trained in VC testing during a visit to a multidisciplinary clinic or at home. Starting home-monitoring of VC shortly after diagnosis is most beneficial, as insight into the rate of disease progression can guide the timing of clinical interventions. When patients show noticeable or

unexpected changes in their unsupervised VC during home-monitoring, a face-to-face or video consultation may be scheduled in order to determine whether a change in VC was caused by respiratory muscle weakness, or other factors, such as pain/discomfort, illness, fatigue or performing the VC test incorrectly. Support during VC testing at home could be improved by including MND-specific prompts, and written and visual feedback (e.g. flow-volume curve) in the mobile spirometry app.

Home-monitoring of VC could be combined with patient-reported symptoms of dyspnea, in order to provide healthcare professionals with more insight into the patient's respiratory function and reduce false negative findings. When home-monitoring data indicates the presence of respiratory dysfunction, based on VC, symptoms or both, patients should be referred to a multidisciplinary clinic for further examination. An advantage of this approach is that the frequency and timing of clinic visits will be tailored to the rate of disease progression and needs of individual patients. In turn, this may result in earlier detection of a respiratory function decline, and more timely referral to a pulmonologist or initiation of NIV, compared to the usual 3 monthly in-clinic care. This study contributes to the recently published Road Map, that was created to facilitate the wide-scale adoption of digital technology and remote monitoring in MND, as it provides evidence on how to measure respiratory function in patients with MND (31).

Strengths and limitations

A strength of the present study is that home-monitoring of VC was part of an existing telehealth service, which facilitated home-monitoring and communication, and optimized adherence. A limitation is the fact that the majority of patients in our cohort were male and relatively young, which reduces the generalizability of our results. Future studies could assess long-term home-monitoring of VC, and determine to what extent the course of unsupervised VC over time corresponds to disease progression, and how it relates to decision-making in MND care. We assessed the upright VC in the present study, despite studies showing that in some patients the upright FVC may remain stable even when respiratory insufficiency is already present (32–35). Based on existing literature, the maximal inspiratory pressure (MIP), sniff nasal inspiratory pressure (SNIP) or supine VC may be more sensitive in detecting respiratory muscle weakness [9,18,27–29]. However, due to the lack of low-cost respiratory pressure meters, home-monitoring of MIP and SNIP will be much more costly. Furthermore, the supine VC test can be challenging and burdensome to perform for patients with gross motor disability, as it requires transfer to a flat

surface. As a result, more patients may require assistance from a caregiver, which increases caregiver burden and may reduce adherence. However, future studies could evaluate whether other pulmonary function tests, besides the upright VC, are valid and feasible for home-monitoring in patients with MND.

Conclusion

Unsupervised VC testing at home, with prior face-to-face training and reminders during follow-up, is a valid and feasible method for the remote monitoring of respiratory function in MND care, and well-received by patients and their caregivers.

References

1. Burkhardt C, Neuwirth C, Sommacal A, Andersen PM, Weber M. Is survival improved by the use of NIV and PEG in amyotrophic lateral sclerosis (ALS)? A post-mortem study of 80 ALS patients. PLoS One. 2017;12(5):1–12.
2. Corcia P, Pradat PF, Salachas F, Bruneteau G, le Forestier N, Seilhean D, et al. Causes of death in a post-mortem series of ALS patients. Amyotroph Lateral Scler. 2008;9(1):59–62.
3. Andersen PM, Abrahams S, Borasio GD, de Carvalho M, Chio A, Van Damme P, et al. EFNS guidelines on the Clinical Management of Amyotrophic Lateral Sclerosis (MALS) - revised report of an EFNS task force. Eur J Neurol. 2012;19:360–75.
4. NICE. Motor neurone disease: assessment and Motor neurone disease: assessment and management management NICE guideline [Internet]. 2016 [cited 2019 Aug 30]. Available from: www.nice.org.uk/guidance/ng42
5. Miller RG, Jackson CE, Kasarskis EJ, England JD, Forshew FD, Johnston W, et al. Practice Parameter update: The care of the patient with amyotrophic lateral sclerosis: Drug, nutritional, and respiratory therapies (an evidence-based review). Neurology. 2009;73:1218–26.
6. Bourke SC, Tomlinson M, Williams TL, Bullock RE, Shaw PJ, Gibson GJ. Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: A randomised controlled trial. Lancet Neurol. 2006;5(2):140–7.
7. Vrijsen B, Buyse B, Belge C, Robberecht W, Van Damme P, Decramer M, et al. Noninvasive ventilation improves sleep in amyotrophic lateral sclerosis: A prospective polysomnographic study. J Clin Sleep Med. 2015;11(5):559–66.
8. Mustfa N, Walsh E, Bryant V, Lyall RA, Addington-Hall J, Goldstein LH, et al. The effect of noninvasive ventilation on ALS patients and their caregivers. Neurology. 2006;66(8):1211–7.
9. Stephens HE, Young J, Felgoise SH, Simmons Z. A qualitative study of multidisciplinary ALS clinic use in the United States. Amyotroph Lateral Scler Front Degener. 2016 Feb 17;17(1–2):55–61.
10. Schellenberg KL, Hansen G. Patient perspectives on transitioning to amyotrophic lateral sclerosis multidisciplinary clinics. J Multidiscip Healthc. 2018;11:519–24.
11. Hobson E, Baird W, Bradburn M, Cooper C, Mawson S, Quinn A, et al. Process evaluation and exploration of telehealth in motor neuron disease in a UK specialist centre. BMJ Open. 2019 Oct 22;9(10):e028526.
12. Helleman J, Van Eenennaam R, Kruitwagen ET, Kruithof WJ, Slappendel MJ, Van Den Berg LH, et al. Telehealth as part of specialized ALS care: feasibility and user experiences with "ALS home-monitoring and coaching." Amyotroph Lateral Scler Front Degener. 2020;21(3–4):183–92.
13. De Marchi F, Cantello R, Ambrosini S, Mazzini L, Sarnelli MF, De Marchi I, et al. Telemedicine and technological devices for amyotrophic lateral sclerosis in the era of COVID-19. Neurol Sci. 2020;41(6):1365–7.
14. Capozzo R, Zoccolella S, Musio M, Barone R, Accogli M, Logroscino G. Telemedicine is a useful tool to deliver care to patients with Amyotrophic Lateral Sclerosis during COVID-19 pandemic: results from Southern Italy. Amyotroph Lateral Scler Front Degener [Internet]. 2020;21(7–8):542–8. Available from: https://doi.org/10.1080/21678421.2020.1773502
15. Selkirk SM, Washington MO, McClellan F, Flynn B, Seton JM, Strozewski R. Delivering tertiary centre specialty care to ALS patients via telemedicine: a retrospective cohort analysis. Amyotroph Lateral Scler Front Degener. 2017;18(5–6):324–32.
16. Van De Rijn M, Paganoni S, Levine-Weinberg M, Campbell K, Swartz Ellrodt A, Estrada J, et al. Experience with telemedicine in a multi-disciplinary ALS clinic. Amyotroph Lateral Scler Front Degener [Internet]. 2018;19(1–2):143–8. Available from: https://doi.org/10.1080/21678421.2017.1392577

17. Helleman J, Kruitwagen-van Reenen ET, Bakers J, Kruithof WJ, van Groenestijn AC, Jaspers Focks RJH, et al. Using patient-reported symptoms of dyspnea for screening reduced respiratory function in patients with motor neuron diseases. J Neurol [Internet]. 2020; Available from: https://doi.org/10.1007/s00415-020-10003-5

18. Baumann F, Henderson RD, Morrison SC, Brown M, Hutchinson N, Douglas JA, et al. Use of respiratory function tests to predict survival in amyotrophic lateral sclerosis. Amyotroph Lateral Scler. 2010;11(1–2):194–202.

19. Calvo A, Vasta R, Moglia C, Matteoni E, Canosa A, Mattei A, et al. Prognostic role of slow vital capacity in amyotrophic lateral sclerosis. J Neurol. 2020;267(6):1615–21.

20. Helleman J, Kruitwagen ET, van den Berg LH, Visser-Meily JMA, Beelen A. The current use of telehealth in ALS care and the barriers to and facilitators of implementation: a systematic review. Amyotroph Lateral Scler Front Degener. 2020;21(3–4):167–82.

21. Tattersall R, Carty S, Meldrum D, Hardiman O, Murray D. The patient's perspective of remote respiratory assessments during the COVID-19 pandemic. Amyotroph Lateral Scler Front Degener [Internet]. 2021;0(0):1–5. Available from: https://doi.org/10.1080/21678421.2021.1920982

22. Meng L, Rudnicki SA, Lechtzin N, Cockroft B, Malik FI, Wolff AA, et al. Correlation between slow vital capacity measured in the home and in the clinic for patients with amyotrophic lateral sclerosis. Neurology. 2019;92(15 supplement):P4.4-002.

23. Pellegrino GM, Sferrazza Papa GF, Centanni S, Corbo M, Kvarnberg D, Tobin MJ, et al. Measuring vital capacity in amyotrophic lateral sclerosis: Effects of interfaces and reproducibility. Respir Med [Internet]. 2021;176(November 2020):106277. Available from: https://doi.org/10.1016/j.rmed.2020.106277

24. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. J Neurol Sci. 1999;169(1–2):13–21.

25. van Eijk RPA, Bakers JNE, van Es MA, Eijkemans MJC, van den Berg LH. Implications of spirometric reference values for amyotrophic lateral sclerosis. Amyotroph Lateral Scler Front Degener [Internet]. 2019;20(7–8):473–80. Available from: https://doi.org/10.1080/21678421.2019.1634736

26. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: The global lung function 2012 equations. Eur Respir J. 2012;40(6):1324–43.

27. Mendoza M, Gelinas DF, Moore DH, Miller RG. A comparison of maximal inspiratory pressure and forced vital capacity as potential criteria for initiating non-invasive ventilation in amyotrophic lateral sclerosis. Amyotroph Lateral Scler. 2007;8(2):106–11.

28. Geronimo A, Simmons Z. Evaluation of remote pulmonary function testing in motor neuron disease. Amyotroph Lateral Scler Front Degener. 2019;20(5–6):348–55.

29. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: Systematic review and content analysis of findings. Vol. 20, Journal of Medical Internet Research. Journal of Medical Internet Research; 2018.

30. Edwards C, Costello E, Cassidy N, Vick B, Russell AM. Use of the patientMpower app with home-based spirometry to monitor the symptoms and impact of fibrotic lung conditions: Longitudinal observational study. JMIR mHealth uHealth. 2020;8(11):1–9.

31. van Eijk RPA, Beelen A, Kruitwagen ET, Murray D, Radakovic R, Hobson E, et al. A Road Map for Remote Digital Health Technology for Motor Neuron Disease. J Med Internet Res. 2021;23(9):e28766.

32. Jackson CE, Rosenfeld J, Moore DH, Bryan WW, Barohn RJ, Wrench M, et al. A preliminary evaluation of a prospective study of pulmonary function studies and symptoms of hypoventilation in ALS/MND patients. J Neurol Sci. 2001;191:75–8.

33. Elman LB, Siderowf AD, McCluskey LF. Nocturnal Oximetry: Utility in the Respiratory Management of Amyotrophic Lateral Sclerosis. Am J Phys Med Rehabil. 2003;82(11):866–70.

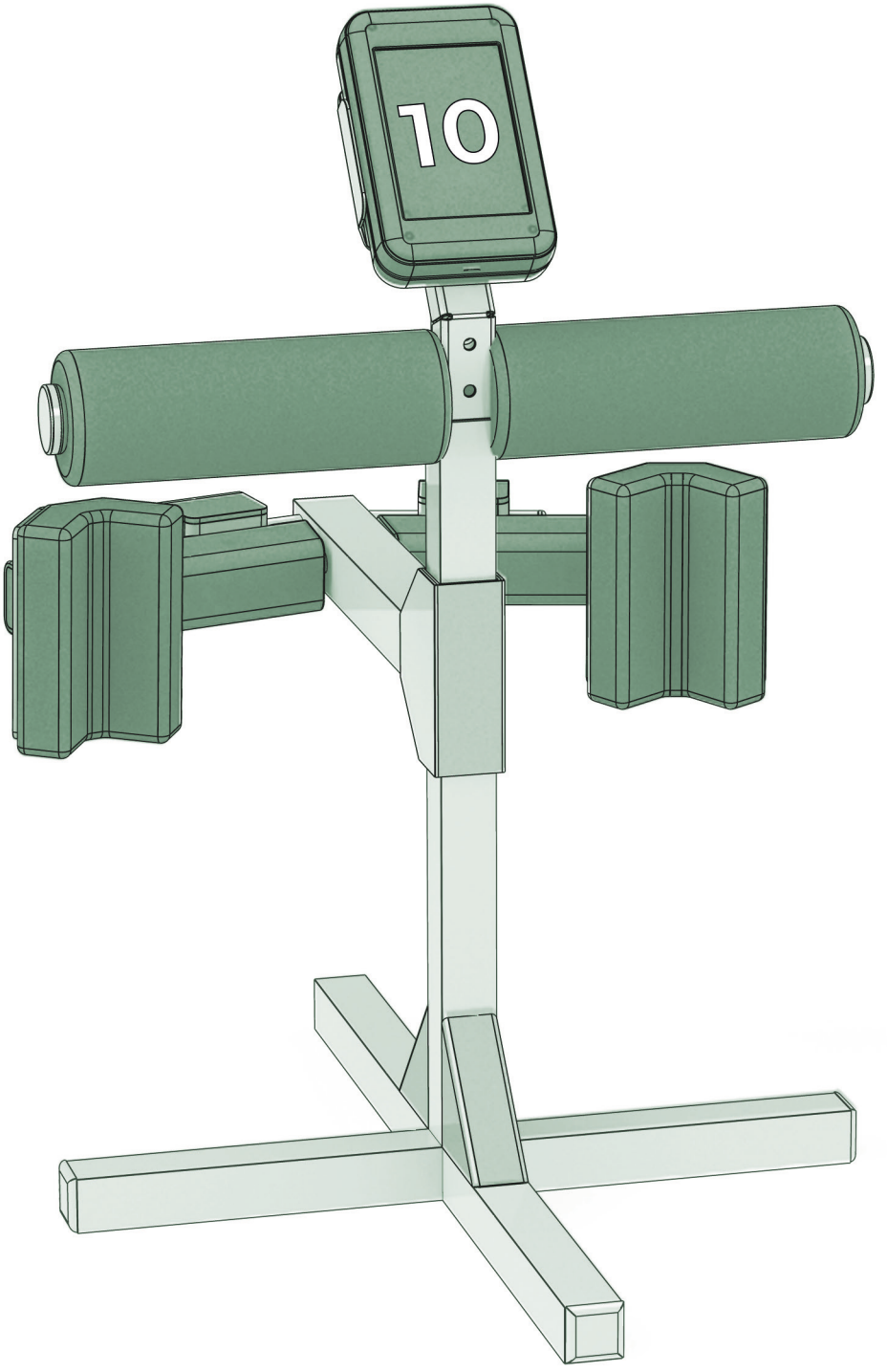
34. Panchabhai TS, Mireles Cabodevila E, Pioro EP, Wang X, Han X, Aboussouan LS. Pattern of lung function decline in patients with amyotrophic lateral sclerosis: implications for timing of noninvasive ventilation. ERJ Open Res [Internet]. 2019;5(3):00044–2019. Available from: http://dx.doi.org/10.1183/23120541.00044-2019

35. Prell T, Ringer TM, Wullenkord K, Garrison P, Gunkel A, Stubendorff B, et al. Assessment of pulmonary function in amyotrophic lateral sclerosis: When can polygraphy help evaluate the need for non-invasive ventilation? J Neurol Neurosurg Psychiatry. 2016;87(9):1022–6.

36. Schmidt EP, Drachman DB, Wiener CM, Clawson L, Kimball R, Lechtzin N. Pulmonary predictors of survival in amyotrophic lateral sclerosis: Use in clinical trial design. Muscle and Nerve. 2006;33(1):127–32.

37. Lechtzin N, Wiener CM, Shade DM, Clawson L, Diette GB. Spirometry in the supine position improves the detection of diaphragmatic weakness in patients with amyotrophic lateral sclerosis. Chest. 2002;121:436–42.

38. Carratú P, Cassano A, Gadaleta F, Tedone M, Dongiovanni S, Fanfulla F, et al. Association between low sniff nasal-inspiratory pressure (SNIP) and sleep disordered breathing in amyotrophic lateral sclerosis: Preliminary results. Amyotroph Lateral Scler. 2011;12(6):458–63.


10

Chapter 10

GENERAL DISCUSSION

Main conclusions

Current use of clinical outcome measures

Respiratory function tests are currently used in amyotrophic lateral sclerosis (ALS) care to monitor disease progression and to support timely initiation of symptomatic interventions, such as non-invasive ventilation (NIV) and air-stacking. However, the most commonly used test, the forced vital capacity in sitting position (FVC), does not seem to be sufficiently sensitive to detect early respiratory dysfunction. As an alternative, we recommend combining evaluation of signs and symptoms with more direct measurements of respiratory strength (**chapter 2**). In multifocal motor neuropathy (MMN), muscle strength measurements of distal muscle groups are used to evaluate treatment effects. However, we found equally large treatment responses in proximal muscle groups (**chapter 3**). Therefore, adding proximal muscle groups to muscle strength evaluations contributes to optimisation of administration, dosing and treatment interval of immunoglobulins (**chapter 4**).

Shortcomings of clinical outcome measures

Although assessment of clinical outcome measurements is relevant in care and clinical trials, such measurements can be prone to limitations. Our longitudinal analysis of ALS functional rating scale revised (ALSFRS-R) scores from two multi-centre drug trials showed that implausible improvements in the ALSFRS-R score occurred frequently (**chapter 5**). Another limitation is the use of different reference values in respiratory function testing, as this has a major impact on clinical decision-making and sample size calculations for clinical trials (**chapter 6**). A final limitation addressed in this thesis, are the strong ceiling effects in manual muscle testing, caused by the inability of the examiner to provide sufficient counter-pressure as the strength of the participant increases. These ceiling effects were significantly reduced by using a fixed dynamometer (**chapter 7**).

The potential of remote clinical outcome measurements.

Currently, clinical outcome measurements are collected during supervised, in-clinic assessments. However, remote measurements (e.g. independent measurements at patients' homes) promise several benefits as travel burden decreases and measurements can be assessed more frequently. This thesis includes multiple studies in which ALS patients performed remote clinical outcome measurements. We found that remote measurements of accelerometry (**chapter 8**), respiratory function (**chapter 9**) and muscle strength (**in preparation**) were generally feasible and reliable and allowed frequent assessment. These results contribute significantly

to the transition from in-clinic to remote measurements, allowing faster and better clinical decision-making in care, and the collection of high quality data in the search for a cure for ALS.

General discussion

Optimal use of clinical outcome measurements in care and clinical trials

Due to heterogeneity in disease pattern and course of motor neuron diseases (e.g. ALS and MMN), measurement of relevant indicators of disease progression is critical for the delivery of personalised care and evaluation of effectiveness during clinical trials. Although there are common interests, the interpretation and use of clinical outcome measures can differ in care and clinical trials.

To understand the impact of a disease on the limitations on daily activities in individual patients, it is important to find out what these limitations are based on. The scope of this search is not only focussed on body functions, such as muscle strength (including respiratory muscles), but also on personal and environmental factors related to the individual patient.[1] As limitations on activities and participation strongly depend on personal and environmental factors, these factors also influence the choice and interpretation of clinical outcome measurements in care.[2] In **chapter 2**, we provide clinical recommendations for early recognition of respiratory dysfunction. We found that not only respiratory function tests, which reflect respiratory muscle strength, but also symptoms (i.e. orthopnoea and inefficient cough) play an important role in the detection of early respiratory dysfunction. These symptoms are often not confirmed by respiratory tests, as they are also related to personal and environmental factors, such as exercise and lifestyle behaviour. Detection of early respiratory dysfunction facilitates timely initiation of interventions that can alleviate these symptoms, and contributes to optimal timing of decision-making by patients and their caregivers with regard to respiratory management.

Unlike in care, in early experimental drug trials (e.g. phase 2) the primary aim is to find treatment options that modify physiological factors directly related to the disease (e.g. body functions such as muscle strength). As measurements of daily activities are also determined by multiple factors other than physiological disease-related factors, these measurements (including questionnaires) are not necessarily suitable for clinical trials of experimental drugs.[3]

For example, a patient, whose walking ability is decreasing due to the natural course of the disease, is participating in an experimental drug trial. At a certain point after inclusion, he begins to stumble because his dropping foot gets stuck behind thresholds in the house, especially when he is tired after his daily one-hour walk. This leads to him taking several measures. His ALS care team provides him with an ankle foot orthosis and he has the thresholds removed. As a result, he is better able to complete his daily walk without exhaustion and no longer stumbles indoors. However, this improvement in walking ability is not caused by an improvement in his muscle strength or by a treatment effect of the experimental drug, but by external factors.

In the current absence of definitive diagnostic and prognostic biomarkers,[4, 5] the most frequently used clinical outcome measurement as primary endpoint in ALS clinical trials is the ALSFRS-R questionnaire.[6] However, as this questionnaire uses daily activity levels as a reference, the outcome may not provide a way of distinguishing between treatment effects of symptomatic interventions and experimental treatment effects. Just as in the previous example, where improvements in walking ability were not caused by improvements directly related to the disease course, this could also occur during clinical trials; for example, after initiation of effective symptomatic interventions, such as non-invasive ventilation in case of respiratory dysfunction.[7] Especially when (expected) treatment effects during clinical trials are modest, when using clinical outcome measurements further removed from body functions (i.e. towards activity and participation level), it may become increasingly difficult to explain effects by the experimental treatment.

Another factor that may hamper optimal use of the ALSFRS-R for clinical trials is multidimensionality. Although the ALSFRS-R covers the three main domains of ALS symptoms (bulbar, motor and respiratory function), due to the summation of the domains, patients with identical ALSFRS-R total scores may not be comparable as far as disease stage or prognosis are concerned.[8] As a result, a potential treatment effect within one domain may remain unnoticed, as is it masked by the other domains. Finally, in this thesis we show that the ALSFRS-R questionnaire is prone to unexpected variation related to several factors, such as ambiguous standard operating procedures (**chapter 5**).

Rather than using one composite outcome measurement of activity level, it might be of value to assess each domain using its own powerful, innovative outcome measure. This approach would allow targeted comparison of patients with the same

disease stage per domain. In addition, it would be useful to measure both body function and activity/participation level within each domain, as this would allow one to determine whether small effects at the level of body functions (including future definite biomarkers) are noticeable in daily life.

Implementation of innovative outcome measurements of motor function: do the advantages outweigh the disadvantages?

A wide array of instruments is available for the measurement of respiratory function and isometric muscle strength, as these instruments are also relevant in other patient groups with neuromuscular diseases.[9] The decision to use traditional instruments in the context of ALS trials and care has often been arbitrary or retrieved from other diseases, but not based on broad consensus.[10] In **chapter 2**, we discuss the fact that the Forced Vital Capacity (FVC), traditionally used in ALS research and care, is not on its own sufficiently sensitive to detect early respiratory dysfunction. Amongst others, the Maximal Inspiratory Pressure (MIP) and Peak Cough Flow (PCF) are more suitable alternatives for sensitive detection of early respiratory dysfunction. However, a disadvantage of these innovative respiratory function tests is a lack of high-quality reference values.[11, 12] In **chapter 6,** however, we show that the use of high quality reference values (e.g. GLI-2012) is indispensable, both for clinical decision-making and as a powerful outcome measure in trials. The absence of high quality reference values in innovative outcome measures results in the use of absolute, and not relative, cut-off values to determine whether (early) respiratory dysfunction is present.[13, 14] As a consequence, for innovative respiratory function tests (i.e. MIP and PCF), the same cut-off value is used in a person of large stature in their twenties, as in a person of short stature in their seventies. Clearly it is difficult to interpret absolute values in individual cases, thus greatly reducing their clinical value.

The measurement of muscle strength presents a comparable situation: hand-held dynamometry (HHD) is less reliable than portable fixed dynamometry (PFD) due to an examiner-induced ceiling effect **(chapter 7)**. However, there are widely used reference values for HHD,[15, 16] making it, despite its demonstrated methodological weakness, the most commonly used measurement method in motor neuron disorders research and care.[17, 18] This lack of high-quality reference values for respiratory function and muscle strength causes researchers and practitioners to drift into the arms of traditional but inferior clinical outcome measures, with the risk of suboptimal personalised care and of missing potential treatment effects.

Although the search to find a cure in ALS is in full swing, we can learn from other MNDs (e.g. MMN and Spinal Muscular Atrophy), for which disease-modifying treatments with substantial effects are already available, that functional limitations due to muscle weakness will unavoidably remain linked to the disease (**chapter 3**).[19, 20] Although finding a cure/prolonging survival is the highest aspiration, it is likely that the treatments available in the near future will not be a cure, but initially will slow down the rate of progression and reduce the severity of motor function loss. Therefore, in the search for a treatment for ALS, not only survival time, but also treatment effects on clinically relevant motor function outcome measurements are indispensable. In addition, subgroups may arise, for example among patients with genetic variants, in which a treatment could be more effective.[21] As a result, heterogeneity in disease course within the ALS population will increase even further, with a parallel increase in the need for personalised care. Relevant and powerful measurements of motor function will be crucial for optimal dosage and treatment frequency of medication, notwithstanding the adequate timing of treatments for residual symptoms. Therefore, the measurement of motor function in ALS will remain relevant in the foreseeable future, and warrants innovation.

Remote assessment of motor function improves trial participation rate, data quality and personalised care

Currently, the most commonly used efficacy endpoints in ALS trials are the ALSFRS-R questionnaire, survival time, respiratory function, and muscle strength.[6] With the exception of survival time, these outcome measures are typically administered longitudinally, in-clinic, by trained examiners, such as research nurses. Although uniform scoring strategies and adequate skills are conditional for reliable and comparable results, we show that such standardisation is problematic, causing unwanted variation (**chapter 5**) and ceiling effects (**chapter 7**). The hope is that skill-training will prevent measurement errors as much as possible. However, ambiguous operating procedures and complex measurement methods are likely to be an underestimated source of unwanted variation.

Other disadvantages are inherent in the current practise of assessing motor function at a centralised location (e.g. in-clinic). Besides the increasing pressure on health systems, due to demographic shifts, that advocates the implementation of new directions in healthcare, the travel burden is an important reason for patients to reject participation in clinical trials thus contributing to selection bias.[22, 23] But the

journey to the outpatient clinic could in itself also reduces data quality of motor function tests (see textbox for an example).

> Imagine an ALS patient with motor disorders. Today she will undergo measurements as part of a clinical trial. With the help of her partner who has taken time off from work to accompany her, they drive to the hospital where they have to walk from the parking garage towards the main entrance of the hospital. From there they go to the ward and, tired from the journey, she takes a seat in the waiting room. She will then be picked up by a research nurse and will walk to the examination room, where a blood sample will be taken and various questionnaires administered. Finally, she will undergo a respiratory function test, the score is depending on how deeply one breathes in and out, and different muscle strength tests for the limbs, where it is important that one produces as much strength as possible.

In this example, one has to ask how representative the scores of these in-clinic measurements were for the actual motor function of this patient. Were any changes in the score caused entirely by disease progression or treatment effect? Wouldn't it make more sense if one could perform the motor function tests and fill in the questionnaires at home, resulting in a reduction of the burden of study participation and an improvement in the quality of measurements?

In the context of ALS care, clinical guidelines recommend standard visits to the multidisciplinary care team every two to three months, including the assessment of relevant clinical outcome measurements (i.e. respiratory function).[24, 25] In trials, the follow-up interval varies, but is also typically two to three months. However, remote measurement allows more frequent (**chapter 7 and 9**), or even continuous monitoring (**chapter 8**). This higher frequency makes the scores less vulnerable for noise, and could provide more insight into changes over time, potentially reducing required sample size or improving precision in clinical trials.[26] In care, early recognition of motor dysfunction allows prompt initiation of interventions that can alleviate symptoms (**chapter 2**). In addition, frequent monitoring of symptoms allows personalisation of timing and frequency of clinical visiting schemes, and optimizes the detection (or prediction) of events, such as respiratory failure or wheelchair dependency. Importantly, the preference for remote measurements has also recently been expressed from a patient perspective. In an international survey amongst 332

ALS patients, 69% responded that they would be willing to self-monitor their health from home, and indicated respiratory function and muscle strength to be the most valuable remote outcome measurements.[27]

The potential of remote independent measurements of motor function

To reduce the burden of hospital visits and to optimise the value of outcome measures, a shift from the current practise of in-clinic supervised to remote independent assessment of motor function is currently taking place. Within and outside the field of ALS, numerous initiatives of remote monitoring have been undertaken recently and published, partly accelerated due to the covid-19 pandemic.[28] From these initiatives, we have learned that implementation of remote measurements brings several challenges. Conditional factors are that the instruments and their methods of assessment are easy to use. Also independent (unsupervised) measurements should be of sufficient psychometric quality, clinically relevant, and sensitive to capture disease progression. This transition from in-clinic to remote assessment, therefore, involves modifying complex standard operating procedures (SOPs) of supervised measurements to make them suitable for independent home measurement. Modifying the SOPs, and having the measurements performed by relatively inexperienced assessors (e.g. patients and their carers) increases the risk of measurement errors. Exchanging measurement data and communicating remotely between patients and researchers/care teams also presents challenges. Innovation under controlled conditions (e.g. scientific research) is, therefore, necessary to slowly but surely overcome these hurdles and work towards broad implementation where the benefits of home measurement outweigh the disadvantages. A good example of an application of home measurement with excellent implementation is the blood pressure measurement. Whereas previously people had to visit a doctor to have their blood pressure measured with the aid of a stethoscope, a microphone built into the arm cuff now enables people to take the measurement independently at home. Blood pressure meters with this microphone are user-friendly and affordable, and although measuring with a stethoscope is still the gold standard, FDA-approved blood pressure monitors for home use are considered to be sufficiently reliable for wide use in healthcare and scientific research.[29]

This thesis has made the first moves towards remote, independent motor function assessments of muscle strength **(chapter 7)**, accelerometry **(chapter 8)**, and respiratory function (**chapter 9**). By using a newly developed portable fixed dynamometer (PFD), instead of a hand-held dynamometer (HHD), isometric

quadriceps muscle strength was tested without the help of an examiner (**chapter 7**). In addition to a slightly superior reliability of the PFD, we found that the PFD was able to reduce the examiner-induced ceiling effect of the HHD to a significant extent, as the PFD increased the range of muscle strength measurements across individuals nearly two-fold from 414 to 783 N. Currently, a new research report is in preparation about a promising follow-up study, in which 17 ALS patients measured their muscle strength independently at home once every two weeks for six months, using a further developed version of the PFD (Iso-Quad). We found that adherence rates were high; of the 221 planned muscle strength measurements, 194 (88%) were completed. Also, participants indicated that the device was generally easy to use and suitable for independent use at home. Please note that images of the Iso-Quad have been used on the cover and the chapter pages of this thesis.

A striking similarity in the remote independent assessment of motor function studies, included in the thesis, is that the majority of the participants perceive the measurements as not burdensome and prefer home measurement to in-clinic measurement. In addition, adherence rates (i.e., the percentage of planned measurements that are actually completed) were generally excellent. One possible explanation for the high satisfaction and adherence rates in these studies is that a large proportion of the included participants was already using a digital ALS care platform in usual care. Therefore, they were used to performing simple remote measurements (i.e. body weight) and to communicating remotely with practitioners from their multidisciplinary ALS teams.[30] The motor function measurements data were shared with their practitioners, and could subsequently be taken into account to enhance shared decision-making. A possible explanation for the mostly high psychometric quality of the independent remote measurements is the effort to simplify instruments and their methods of assessment.

Recommendations for future research

Optimal exchange of outcome measurements between and within care and research

Conditional for facilitating wide participation of the ALS population in clinical trials are easily assessable, user-friendly outcome measures. In addition, the overlapping use of relevant and powerful outcome measures in care and trials offers mutual benefits. For example, high quality databases collected in care can serve as a replacement for traditional control groups in randomised trials. Innovative trial designs then allow more participants to be assigned to the intervention group, increasing convenience for participants and reducing required sample rates. Also pre-slopes (rates of disease

progression) of motor function collected in care allow evaluation of treatment effects within participants. Vice-versa, ALS care teams could be informed (real time) by outcome measurements that are assessed during clinical trials. Several national and international initiatives have recently been launched that will facilitate electronic data exchange through platforms that merge databases.[31, 32] These initiatives pave the way for all kinds of innovations in the field of data-driven care and reuse of health data in clinical trials.

Prognostic model for functional (dis)ability in ALS
Due to variability in disease pattern and course, timely and accurate provision of assistive devices like wheelchairs is challenging. A prognostic model for survival in ALS already exists;[33] however, for patients with progressive muscle strength loss, it is of great importance to know and anticipate when disability will set in. As muscle strength is associated with functional ability, the course of quadriceps strength could, for example, provide insight into how long walking ability will be preserved or provide an indication for the risk of falling.

Implementation of the use of clinically relevant and user-friendly but sensitive and reliable outcome measurements of motor function.
The transition to remote measurement provides the opportunity to implement user-friendly, but sensitive outcome measurements. Ambiguous SOPSs should be adjusted, allowing comparison and exchange of data worldwide. An important prerequisite for implementation of innovative outcome measurements is the collection of high quality reference values. An important opportunity for acceleration of implementation of remote measurements, therefore, lies with international ALS trial networks, such as The Trial Research Initiative to Cure ALS (TRICALS), and the North East ALS Consortium (NEALS). To harness worldwide data exchange, registries of expert centres in ALS trial networks are necessary.

In conclusion, this thesis shows that clinical outcome measurements on activity and participation level are particularly suitable for personalised care, but not necessarily for clinical trials. Especially during the search for an effective treatment (e.g. during early phase trials), clinical outcome measures at the level of motor function have the potential to powerfully demonstrate disease-modifying treatment effects. However, the current practise of supervised in-clinic measurements of motor function has several disadvantages. These can largely be overcome by using remote independent measurements. We found that measurements of muscle strength, accelerometry and respiratory function can be reliably performed at home. Importantly, participants

experienced independent home measurement to be remarkably user-friendly and not burdensome. These studies show the potential of independent measurements of motor function at home for the delivery of personalised care and patient-centric clinical trials.

References

1. Van Groenestijn AC, Schroder CD, Kruitwagen-Van Reenen ET, Van Den Berg LH, Visser-Meily JMA. Participation restrictions in ambulatory amyotrophic lateral sclerosis patients: Physical and psychological factors. Muscle Nerve. 2017;56:912-918.
2. World Health Organization. International classification of functioning, disability and health : ICF. Geneva: World Health Organization. 2001.
3. Coster WJ. Making the best match: selecting outcome measures for clinical trials and outcome studies. Am J Occup Ther. 2013;67:162-170.
4. Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, et al. Amyotrophic lateral sclerosis. Nat Rev Dis Primers. 2017;3:17085.
5. Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, et al. Amyotrophic lateral sclerosis. Lancet. 2011;377:942-955.
6. van Eijk RPA, Kliest T, van den Berg LH. Current trends in the clinical trial landscape for amyotrophic lateral sclerosis. Curr Opin Neurol. 2020;33:655-661.
7. Piepers S, van den Berg JP, Kalmijn S, van der Pol WL, Wokke JH, Lindeman E, et al. Effect of non-invasive ventilation on survival, quality of life, respiratory function and cognition: a review of the literature. Amyotroph Lateral Scler. 2006;7:195-200.
8. van Eijk RPA, de Jongh AD, Nikolakopoulos S, McDermott CJ, Eijkemans MJC, Roes KCB, et al. An old friend who has overstayed their welcome: the ALSFRS-R total score as primary endpoint for ALS clinical trials. Amyotroph Lateral Scler Frontotemporal Degener. 2021;22:300-307.
9. Gartman EJ. Pulmonary Function Testing in Neuromuscular and Chest Wall Disorders. Clin Chest Med. 2018;39:325-334.
10. Schmidt EP, Drachman DB, Wiener CM, Clawson L, Kimball R, Lechtzin N. Pulmonary predictors of survival in amyotrophic lateral sclerosis: use in clinical trial design. Muscle Nerve. 2006;33:127-132.
11. Petrini MF, Haynes D. In search of maximum inspiratory and expiratory pressure reference equations. Respir Care. 2009;54:1304-1305.
12. Cleary S, Misiaszek JE, Kalra S, Wheeler S, Johnston W. The effects of lung volume recruitment on coughing and pulmonary function in patients with ALS. Amyotroph Lateral Scler Frontotemporal Degener. 2013;14:111-115.
13. Fregonezi G, Araujo PR, Macedo TL, Dourado Junior ME, Resqueti VR, Andrade Ade F. Monitoring respiratory muscle strength assists in early diagnosis of respiratory dysfunction as opposed to the isolated use of pulmonary function evaluation in amyotrophic lateral sclerosis. Arq Neuropsiquiatr. 2013;71:146-152.
14. Tilanus TBM, Groothuis JT, TenBroek-Pastoor JMC, Feuth TB, Heijdra YF, Slenders JPL, et al. The predictive value of respiratory function tests for non-invasive ventilation in amyotrophic lateral sclerosis. Respir Res. 2017;18:144.
15. Andrews AW, Thomas MW, Bohannon RW. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. Phys Ther. 1996;76:248-259.
16. Bohannon RW. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. Arch Phys Med Rehabil. 1997;78:26-32.
17. Paganoni S, Cudkowicz M, Berry JD. Outcome measures in amyotrophic lateral sclerosis clinical trials. Clin Investig (Lond). 2014;4:605-618.
18. Rutkove SB. Clinical Measures of Disease Progression in Amyotrophic Lateral Sclerosis. Neurotherapeutics. 2015;12:384-393.

19. Keddie S, Eftimov F, van den Berg LH, Brassington R, de Haan RJ, van Schaik IN. Immunoglobulin for multifocal motor neuropathy. Cochrane Database Syst Rev. 2022;1:CD004429.

20. Wadman RI, van der Pol WL, Bosboom WM, Asselman FL, van den Berg LH, Iannaccone ST, et al. Drug treatment for spinal muscular atrophy types II and III. Cochrane Database Syst Rev. 2020;1:CD006282.

21. van Eijk RPA, Eijkemans MJC, Nikolakopoulos S, Jansen MD, Westeneng HJ, van Eijk KR, et al. Pharmacogenetic interactions in amyotrophic lateral sclerosis: a step closer to a cure? Pharmacogenomics J. 2020;20:220-226.

22. Chio A, Canosa A, Gallo S, Cammarosano S, Moglia C, Fuda G, et al. ALS clinical trials: do enrolled patients accurately represent the ALS population? Neurology. 2011;77:1432-1437.

23. van Eijk RPA, Westeneng HJ, Nikolakopoulos S, Verhagen IE, van Es MA, Eijkemans MJC, et al. Refining eligibility criteria for amyotrophic lateral sclerosis clinical trials. Neurology. 2019.

24. Diagnosis ETFo, Management of Amyotrophic Lateral S, Andersen PM, Abrahams S, Borasio GD, de Carvalho M, et al. EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS)--revised report of an EFNS task force. Eur J Neurol. 2012;19:360-375.

25. Oliver D, Radunovic A, Allen A, McDermott C. The development of the UK National Institute of Health and Care Excellence evidence-based clinical guidelines on motor neurone disease. Amyotroph Lateral Scler Frontotemporal Degener. 2017;18:313-323.

26. van Eijk RPA, Beelen A, Kruitwagen ET, Murray D, Radakovic R, Hobson E, et al. A Road Map for Remote Digital Health Technology for Motor Neuron Disease. J Med Internet Res. 2021;23:e28766.

27. Helleman J, Johnson B, Holdom C, Hobson E, Murray D, Steyn FJ, et al. Patient perspectives on digital healthcare technology in care and clinical trials for motor neuron disease: an international survey. J Neurol. 2022;269:6003-6013.

28. Helleman J, Kruitwagen ET, van den Berg LH, Visser-Meily JMA, Beelen A. The current use of telehealth in ALS care and the barriers to and facilitators of implementation: a systematic review. Amyotroph Lateral Scler Frontotemporal Degener. 2020;21:167-182.

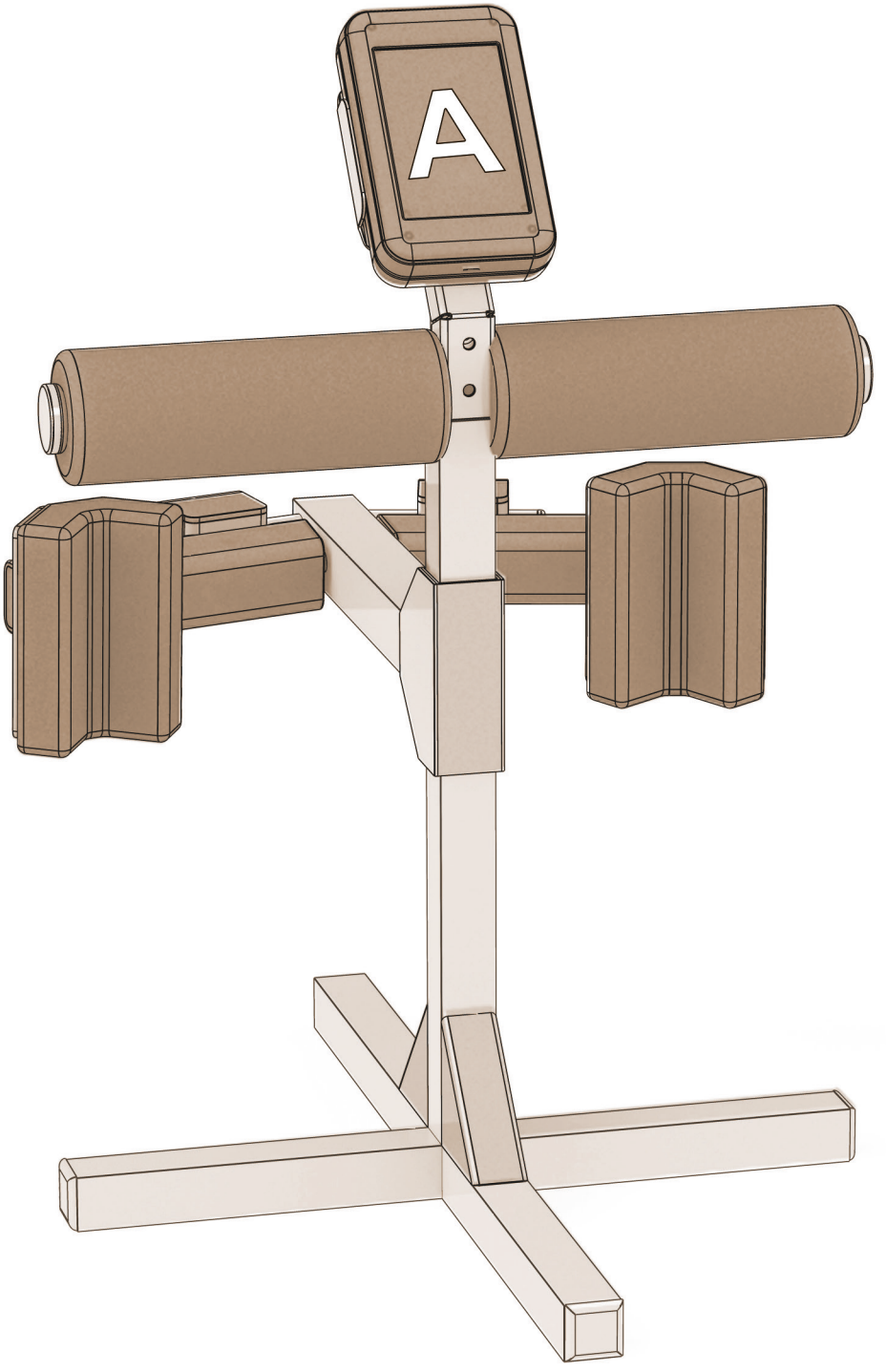
29. US Federal Drug Administration. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/non-invasive-blood-pressure-nibp-monitor-guidance. Accessed 20-10-2022.

30. Helleman J, Van Eenennaam R, Kruitwagen ET, Kruithof WJ, Slappendel MJ, Van Den Berg LH, et al. Telehealth as part of specialized ALS care: feasibility and user experiences with “ALS home-monitoring and coaching”. Amyotroph Lateral Scler Frontotemporal Degener. 2020;21:183-192.

31. European Commision. Available at: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en. Accessed 20-10-2020.

32. Cumuluz. Available at: https://www.cumuluz.org/. Accessed 20-10-2020.

33. Westeneng HJ, Debray TPA, Visser AE, van Eijk RPA, Rooney JPK, Calvo A, et al. Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model. Lancet Neurol. 2018;17:423-433.


A

Appendices

SAMENVATTING IN HET NEDERLANDS

DANKWOORD

CURRICULUM VITAE

LIST OF PUBLICATIONS
NOT INCLUDED IN THIS THESIS

Samenvatting in het Nederlands

Motorische zenuwziekten (motor neuron disorders) is een verzamelnaam voor een groep uiteenlopende neurologische aandoeningen die als overeenkomst hebben dat de zenuwen die de spieren aansturen beschadigd raken. Hierdoor treedt spierkrachtverlies en verlamming op, doorgaans zonder gevoelsstoornissen. Dit proefschrift richt op het meten van spierkracht, ademhalingsfunctie en functionele activiteiten bij twee motorische zenuwziekten waarvoor het Universitair Medisch Centrum Utrecht een landelijke expertisefunctie heeft:

ALS

Amyotrofe laterale sclerose (ALS) is een agressieve en ongeneeslijke ziekte die gekenmerkt wordt door sterfte van motorische zenuwcellen in de hersenen, de hersenstam en het ruggenmerg. Daardoor leidt ALS tot snel progressieve zwakte en verlamming, ook van de ademhalingsspieren. Echter, bij iedere patiënt verloopt het patroon en de snelheid van spierkracht uitval anders. ALS kent mede hierdoor veel verschillende uitingsvormen waardoor de zorg complex is. De overlevingsduur verschilt ook sterk per persoon, maar is gemiddeld 3 tot 4 jaar na het ontstaan van de eerste symptomen, doorgaans ten gevolge van ademhalingsfalen.

MMN

Bij multifocale motorische neuropathie (MMN) raken de zenuwen die ontspringen uit het ruggenmerg en die naar de spieren lopen beschadigd. Ook deze zenuwschade wordt gekenmerkt door spierzwakte. Echter, het beloop van MMN is veel minder agressief en de ademhalingsspieren zijn niet betrokken. Ook bestaat er voor MMN een effectieve behandeling met immunoglobulines die via een infuus worden toegediend (IVIg). Hoewel IVIg de spierkracht van bijna alle patiënten met MMN verbetert, zijn de effecten van beperkte duur en kan blijvend spierkrachtverlies optreden.

Omdat de uitingsvormen tussen en binnen motorische zenuwziekten zeer uiteenlopend zijn is standaard zorg niet mogelijk. Ook binnen wetenschappelijk onderzoek is het lastig om patiënten met elkaar te vergelijken, waardoor de zoektocht naar nieuwe effectieve behandelingen wordt bemoeilijkt. Op dit moment worden metingen van motorische functies, zoals spierkracht en ademhalingsfunctie, standaard door een professional uitgevoerd in het ziekenhuis of behandelcentrum. Echter, voor patiënten met spierkrachtverlies is het zeer belastend om naar het ziekenhuis te komen. Tussen de bezoeken door worden zij niet gemeten waardoor

ziektesymptomen vaak verlaat of niet opgemerkt worden. Als patiënten de mogelijkheid zouden hebben deze metingen zelfstandig, thuis uit te kunnen voeren zou een deel van die nadelen worden weggenomen. De vraag is echter of patiënten dit zelfstandig kunnen en willen uitvoeren en of de metingen van voldoende kwaliteit zijn. Daarom is het van belang om meer kennis te vergaren over goed gebruik van klinische uitkomstmaten (clinical outcome measurements), voor op maat gemaakte zorg en voor het meten van behandeleffecten in wetenschappelijke studies. Het eerste deel van dit proefschrift richt zich op de klinische relevantie van uitkomstmaten bij ALS en MMN. Het tweede deel identificeert belangrijke tekortkomingen van uitkomstmaten die relevant zijn voor ALS zorg en onderzoek. In het derde deel wordt onderzocht of zelfstandig thuismeten de kwaliteit van metingen kan verbeteren en de belasting er van voor patiënten kan verminderen.

Deel 1: De klinische relevantie van uitkomstmaten bij motorische zenuwziekten

Om een richtlijn te ontwikkelen voor het vroegtijdig opsporen van ademhalingsstoornissen en het tijdig initiëren van ondersteunende ademhalingsinterventies bij ALS, combineerden we wetenschappelijk bewijs, de voorkeuren van patiënten en de mening van deskundigen (**hoofdstuk 2**). Om vroegtijdige ademhalingsstoornissen en inefficiënte hoest op te sporen raden we in de richtlijn aan om klachten en symptomen van ademhalingszwakte te evalueren in combinatie met metingen van ademhalingsfunctie zoals de vitale capaciteit in liggende positie en de hoestkracht (peak cough flow). In het geval van inefficiënte hoest zijn verschillende effectieve, maar eenvoudig uit te voeren hoest-ondersteunende technieken beschikbaar. Deze aanbevelingen helpen om in de zorg ademhalingsstoornissen vroeg te kunnen detecteren en daarmee snel eenvoudige interventies op te starten die de klachten kunnen verlichten.

Bij MMN wordt de behandelrespons na start van intraveneuze immunoglobulinen (IVIg) vaak geëvalueerd door middel van spierkracht metingen van spiergroepen aan het uiteinde van de ledematen (distaal). Door spierkrachtevaluaties te analyseren die ook spiergroepen in het midden en aan het begin van de ledematen (proximaal) omvatten, wilden wij meer inzicht krijgen in de respons op IVIg-behandeling in een breder spectrum van spiergroepen (**hoofdstuk 3**). Wij vonden dat verbeterde spierkracht in reactie op IVIg niet alleen in distale, maar in vergelijkbare mate ook in proximale spiergroepen kan worden waargenomen. Evaluatie van een breed spectrum van spiergroepen bij MMN zou daarom kunnen bijdragen aan een betere dosering en behandelingsinterval van IVIg.

In **hoofdstuk 4** onderzochten wij de veiligheid (onder andere door spierkracht evaluaties) en tevredenheid bij verschillende manieren van toediening van immunoglobulinen, namelijk onderhuids (subcutaan: SCIg) of in de ader (intraveneus: IVIg). Op groepsniveau veranderden de spierkracht en de behandelingstevredenheid niet significant na overschakeling op SCIg bij patiënten met MMN die stabiel waren onder IVIg-behandeling. Deze resultaten tonen aan dat spierkrachtmetingen in klinische studies relevant zijn voor vergelijkingen tussen groepen in de context van motorische zenuwziekten.

Deel 2: Huidige beperkingen van uitkomstmaten bij motorische zenuwziekten

Uitkomstmaten zijn dus belangrijk voor zorg en onderzoek. Echter kunnen deze metingen gevoelig zijn voor ongewenste variaties/meetfouten zoals we laten zien in **hoofdstuk 5**. Om inzicht te krijgen in hoe vaak meetfouten voorkomen, analyseerden wij met terugwerkende kracht de scores van een ALS vragenlijst over lichamelijk functioneren (ALSFRS-r) die werden verzameld in twee grote studies, waar meerdere onderzoekscentra aan meededen. Bij ruim 10% van de patiënten bleek dat de scores plotseling sterk toenamen, zonder dat dit het natuurlijke ziekteverloop weerspiegelde. Met name scores met betrekking tot ademhaling, spreken en slikken waren gevoelig voor deze stijgingen, hetgeen kan samenhangen met beschikbare symptomatische behandelingen, zoals niet invasieve beademing, voor deze domeinen. Deze plotse stijgingen kwamen echter niet even vaak voor tussen de onderzoekscentra, vandaar dat het van belang is in multicenter onderzoek om de scoringsstrategieën van de ALSFRS-r te vereenvoudigen en te standaardiseren.

Ook het testen van de ademhalingsfunctie (met behulp van de FVC test) kan worden beïnvloed door een gebrek aan standaardisatie. In **hoofdstuk 6** laten wij zien hoe het gebruik van verschillende referentiestandaarden de klinische besluitvorming en de selectie voor klinisch onderzoek kan beïnvloeden. Omdat de GLI-2012 referentiestandaard de sterkste samenhang vertoonde met overleving en door patiënten gerapporteerde symptomen, bevelen wij het uniforme gebruik ervan aan in zowel in de zorg als in onderzoek.

Deel 3: Innovaties in uitkomstmaten bij motorische zenuwziekten

Bij handmatige spierkrachttesten met behulp van een hand held dynamometer (HHD) hangt de betrouwbaarheid van de score voor een groot deel af van de techniek en de spierkracht van de testafnemer.


Figuur 1: Handmatige spierkracht test met behulp van een hand held dynamometer (HHD)

De betrouwbaarheid kan daarom worden verbeterd door gebruik te maken van een dynamometer (krachtsensor) die in een vaste constructie is gefixeerd en daarmee de testafnemer vervangt. De huidige systemen die gefixeerde dynamometrie toepassen zijn echter niet draagbaar en het gebruik is complex, waardoor deze testen ook door een professional in een ziekenhuis moeten worden afgenomen. Om patiënten met ALS in de toekomst in staat te stellen hun spierkracht zelfstandig thuis te meten hebben we een gebruiksvriendelijke en draagbare vaste dynamometer (portable fixed dynamometer (PFD)) ontwikkeld.


Figuur 2: Dit eerste prototype van de PFD bevat twee houders waar twee losse HHD's (op de foto gemarkeerd met *)

Door herhaalde metingen uit te voeren bij 45 ALS patiënten en 43 controles bepaalden we de test-hertest betrouwbaarheid van de PFD in vergelijking met de HHD (**hoofdstuk 7**). Naast een betere betrouwbaarheid van de PFD te opzichte van de HHD, bleek dat tijdens de HHD metingen de testafnemer niet in staat was om hogere krachten tegen te houden. Hierdoor ontstond een plafondeffect waardoor de werkelijke spierkracht werd onderschat. De PFD was wel in staat om deze krachten tegen te houden, en biedt daarmee een groot voordeel voor het meten van sterkere

spiergroepen. Daarnaast gaf de meerderheid van de deelnemers (ALS patiënten en controles) de voorkeur aan de PFD omdat deze metingen als gebruiksvriendelijker en comfortabeler werden ervaren. Deze bevindingen effenen de weg voor de overgang van metingen in de kliniek naar metingen op afstand (thuis) binnen studies en zorg.

Momenteel wordt een veelbelovende vervolgstudie afgerond waarin 17 ALS patiënten gedurende zes maanden eenmaal per twee weken hun spierkracht zelfstandig, thuis hebben gemeten met een doorontwikkelde versie van de PFD, genaamd Iso-Quad (zie ook de afbeelding op de omslag van dit proefschrift).

Figuur 3 De Iso-Quad met geïntegreerde krachtsensoren en een scherm waarop de instructies worden gegeven zodat de meting om de juiste manier zelf kan worden uitgevoerd.

Uit de eerste resultaten van dit onderzoek komt naar voren dat de meetfout van de metingen zeer laag was. Alle deelnemers bleken de metingen zelfstandig of met hulp van een naaste te kunnen uitvoeren en 86% van de geplande metingen werd voltooid. De meerderheid van de deelnemers vond het apparaat gemakkelijk te gebruiken en alle deelnemers prefereerden het thuismeten van spierkracht boven het meten van spierkracht in het ziekenhuis.

Accelerometrie belooft ook van toegevoegde waarde te zijn bij het op afstand meten van de motorische functie. Een accelerometer is een sensor die versnelling en vertraging in drie assen (3D) kan meten. Door vrijwel continue metingen kan dus de hoeveelheid en intensiteit van lichaamsbewegingen worden geobjectiveerd. Omdat ALS motorisch functieverlies veroorzaakt, werd verondersteld dat dit ook de lichaamsbewegingen zou beïnvloeden. In een groep van 42 ALS-patiënten, die de accelerometer gedurende 12 maanden tijdens het dagelijks leven droegen, werd

gevonden dat accelerometrie gegevens gevoeliger waren voor het meten van ziekteprogressie in vergelijking met de ALSFRS-r (**hoofdstuk 8**).

Naast het thuismeten van spierkracht en de accelerometrie, zou ook het thuismeten van ademhalingsfunctie patiënten in staat kunnen stellen om frequenter te meten en het aantal ziekenhuisbezoeken te verminderen, vooral in de fase waarin ademhalingsstoornissen nog afwezig of licht zijn. Gedurende 12 weken voerde een groep van 33 ALS patiënten eens in de 4 weken thuis een zelfstandige ademhalingsfunctie meting uit (vitale capaciteit (VC) in zittende positie) met een draagbare spirometer. Uit deze haalbaarheidsstudie kwam naar voren dat naast een uitstekende overeenkomst tussen gesuperviseerde en ongesuperviseerde metingen, het merendeel (88%) van de patiënten de metingen als niet belastend ervoer en voerden zij zonder uitzondering alle geplande metingen uit (**hoofdstuk 9**).

Concluderend, dit proefschrift richt zich op goed gebruik van klinische uitkomstmaten binnen spierkracht, ademhalingsfunctie en functionele activiteiten. Deze zijn relevant voor zowel het bevorderen van op maat gemaakte zorg als voor efficiënt en adequaat meten van behandeleffecten in wetenschappelijke studies. Er bestaan momenteel echter verschillende belemmeringen die optimaal gebruik in de weg staan, zoals onduidelijke procedures, het vasthouden aan verouderde uitkomstmaten en gecentraliseerde afname van metingen. Daarom zijn er grofweg 3 grote verbeterslagen te maken:
Het gebruik en de interpretatie van huidige uitkomstmaten moet worden vereenvoudigd en gestandaardiseerd om het risico op meetfouten te verminderen en uniform gebruik ervan te bevorderen.
De implementatie van innovatieve uitkomstmaten die veel (potentiële) voordelen met zich meebrengen voor diagnostiek, prognostiek en effectevaluatie kan worden bevorderd door het vergaren van referentiewaarden van hoge kwaliteit.
Door patiënten in staat te stellen zelfstandig thuis te meten hoeven zijn minder vaak naar het ziekenhuis te komen, kunnen zij metingen vaker uitvoeren waardoor er minder sprake is van een momentopname en ervaren zij grip op hun ziektebeloop.

In dit proefschrift komt herhaaldelijk naar voren dat het thuismeten van motorische uitkomstmaten haalbaar en betrouwbaar is. Overeenkomstig tussen de verschillende thuismeten studies is dat deelnemende patiënten zonder uitzondering enthousiast zijn over thuismeten en hun sterke behoefte hieraan aangeven. Om op maat gemaakte zorg en efficiënt wetenschappelijk onderzoek te bevorderen is het daarom van belang dat het thuismeten met gebruiksvriendelijke en waardevolle uitkomstmaten hoog op de implementatie agenda komt te staan.

Dankwoord

De afgelopen ruim 5 jaar heb ik met veel plezier mijn werk als academisch fysiotherapeut binnen het spierziekten team van het UMC Utrecht kunnen combineren met het uitvoeren van wetenschappelijk onderzoek. Mijn rol als behandelaar heeft dan ook de richting van het onderzoek dat in dit proefschrift is beschreven in belangrijke mate beïnvloed.

Zonder de hulp van velen was het niet mogelijk geweest om deze twee functies te combineren. Ik wil daarom graag van deze gelegenheid gebruik maken om de mensen te bedanken die mij in deze periode hebben geholpen en gesteund.

Deelnemers en leden van de ALS patiënten vereniging (APV)
Dat ALS en aanverwante ziekten een ingrijpend ziektebeloop hebben komt pijnlijk naar voren als ik denk aan alle deelnemers aan de studies van de afgelopen jaren. De meerderheid van hen is in de loop van de tijd toenemend verlamd geraakt of overleden. Dank jullie wel voor jullie bijdrage aan het onderzoek, jullie kritische vragen en jullie interesse in de uitkomsten van studies.

Promotiecommissie
Prof. dr. L.H. van de Berg, beste Leonard. Zeven jaar geleden vroeg je of ik wilde helpen bij een workshop over het meten van spierkracht in Dublin. Dat was het startschot van onze samenwerking -onder andere voor TRICALS en het ALS centrum- en heeft uiteindelijk ook tot deze promotie geleid. Stuk voor stuk projecten waar ik met heel veel plezier aan heb gewerkt, dank je wel daarvoor! Jouw visie op 'the highway towards a cure' werkt verbindend en inspirerend. Ik heb hier veel bewondering voor en hoop bij te kunnen blijven dragen aan de verwezenlijking van deze ambitie.

Prof. dr. J.M.A. Visser-Meily, beste Anne. Naast professor en medisch afdelingshoofd ben jij ook vooral revalidatiearts. Wij vinden elkaar in een mensgerichte visie op de wetenschap. Jouw tomeloze inspanning om een brug te slaan tussen zorg en wetenschap werkt voor mij aanstekelijk en is leerzaam. Jouw empathische en toegankelijke manier van begeleiden heb ik erg gewaardeerd. Ik ben je dan ook zeer dankbaar voor de mogelijkheid die je mij hebt geboden om mij academisch te verbreden.

Dr. J.A.J. Beelen, beste Anita. Vanaf het moment dat jij aan boord kwam ontstond er structuur in mijn promotietraject en hiermee richting en vertrouwen. Ik ben jou

daar zeer dankbaar voor. Ik heb veel geleerd van jouw gave om uit elke zin en analyse het beste te halen. Je hebt mijn publicaties hiermee op een hoger plan getild. Ook buiten mijn promotie hebben we aan veel projecten samengewerkt. We hebben samen de behandelrichtlijn fysiotherapie bij ALS geschreven, met hulp van Taco een IMDD verkregen en ook voor de komende tijd staan er mooie projecten op de rit.

Dr. R.P.A. van Eijk, beste Ruben. Naast jouw talent voor analyses en oog voor de rode draad heb jij ook een ogenschijnlijk onuitputtelijke drive om het onderste uit de kan te halen. Ik was dan ook met trots jouw paranimf toen jij cum laude promoveerde. Met plezier kijk ik terug op het pad dat wij samen hebben bewandeld. Van bouwen aan prototypes tot samen een studie uitvoeren met jou als PI. Ik hoop dat er nog vele studies zullen volgen en met jouw inspanningen om klinische studies efficiënter te maken zal dat vast goed komen!

Ook wil ik graag de leden van de beoordelingscommissie, Prof. dr. Saskia Teunissen, Prof. dr. Peter Wijksta, Prof. dr. Ludo van de Pol, Prof. dr. Frans Nollet en Prof. dr. Cindy Veenhof bedanken voor het lezen en beoordelen van mijn proefschrift.

Onderzoeksteam

Adriaan, **Remko**, **Boudewijn** en **Jordi**, dank jullie wel voor jullie adviezen en gezelligheid tijdens de pauzemomenten. **Jochem**, onze promotietrajecten gingen lange tijd tegelijk op en onze interessegebieden hebben veel overeenkomsten. Naast dat daar een paar mooie publicaties uit zijn voortgekomen heb ik ook veel plezier beleefd aan de gemoedelijke en sportieve momenten met jou. **Steure**, **Toju**, **Maxine**, **Jill** en **Veerle**, jullie frisse energie en onbaatzuchtige hulp tijdens jullie wetenschapsstages was van onmisbare waarde!

TRICALS en ALS centrum

Tommy, het was en is ontzettend leuk om samen met jou het onderwijs voor uitkomstmaten bij TRICALS te geven. **Romy**, dank je wel voor jouw uitstekende projectleiding. **Danique**, **India** en **Nienke** dank jullie wel voor het onder de aandacht brengen van de onderzoeken en jullie ondersteuning bij het onderwijs.

ALS-behandelteam en collega fysiotherapeuten

Esther en **Willeke**, naast jullie inhoudelijke bijdrage aan de onderzoeken en hulp bij het includeren van deelnemers wil ik jullie vooral bedanken voor de prettige samenwerking in de zorg. Alle andere behandelaren van het **ALS-team**: jullie werken niet volgens protocollen, maar met veerkracht en passie voor goede zorg. Het is mij een plezier om met jullie multidisciplinair samen te werken.

Dyta, Hans en andere collega fysiotherapeuten, grote dank voor de fijne samenwerking en jullie waardevolle klinische expertise.

Ondersteuners

Janneke, dank je wel voor de kansen die jij op onze afdeling hebt gecreëerd, ik zie deze promotie als een van de vruchten van jouw beleid. **Marrette** en **Annemarie**, jullie magie met de agenda's was essentieel voor de continuïteit van alle projecten. Mijn dank gaat ook uit naar alle andere ondersteuners, wier hulp niet onopgemerkt mag blijven: Linda, Angela, Alessa, Rifka, Jolanda, Esther D., Karin, Jakub, Diana, Kelly, Mark en Daphne.

De Schaverij en Inventeers

Jesse, je bent waarschijnlijk de beste (en leukste) meubelmaker die er is. Samen prototypes van krachtmeters bouwen in jouw werkplaats, daar kan je mij voor wakker maken. **Jasper en Merel**, dankzij Inventeers zijn ingewikkelde technische verbeter ideeën voor de Iso-Quad krachtmeter werkelijkheid geworden. Jullie toewijding was onmisbaar voor het project en smaakt naar meer!

Familie

Mijn tante en ooms **Maria**, **Joris** en **Jan**. Jullie zijn alle 3 later in jullie carrière gepromoveerd en goed voorbeeld doet goed volgen. Dank voor alle inspirerende discussies.

Paranimfen

Rogier, terugkijkend op onze middelbare schooltijd is het geweldig om te zien hoeveel persoonlijke groei we allebei hebben doorgemaakt. Ik ben heel blij dat onze verwantschap door de jaren heen onveranderd sterk is gebleven. **Maartje**, lieve grote zus. Ontzettend leuk dat we onze band op deze manier kunnen vieren en dat je mij wilt steunen op de feestelijke dag straks.

Als laatste wil ik Jannemieke en mijn ouders bedanken. Pappa en mamma, hoewel ik de 40 ben gepasseerd, is het is ongelofelijk hoe betrokken jullie nog altijd zijn bij mijn leven. Ik prijs mij zeer gelukkig met zulke lieve ouders. Lieve Miepske, jij houdt niet van vleierij, maar hier komtie toch: met jou zijn de pieken hoger en de dalen minder diep.

Curriculum Vitae

Jaap (Japie) Bakers is geboren op 15 juni 1982 in Breda. In 2000 behaalde hij zijn middelbareschooldiploma op het Onze Lieve Vrouw Lyceum in Breda. In 2002 verhuisde hij naar Utrecht om daar fysiotherapie te studeren en studeerde af in 2006. Tussen 2008 en 2010 volgde hij de master bewegingswetenschappen op de Vrije Universiteit van Amsterdam. Tijdens zijn afstudeeronderzoek deed hij onder andere onderzoek naar constraint induced movement therapy, een revalidatie behandelmethode voor halfzijdige verlamming na beroerte, hetgeen leidde tot zijn eerste gepubliceerde wetenschappelijke artikel als coauteur.

Vanwege zijn belangstelling voor neurorevalidatie werkte hij na de universiteit eerst als fysiotherapeut in het Diakonessenhuis in Utrecht maakte in 2011 een 'transfer' naar het UMC Utrecht. In de daaropvolgende jaren heeft hij zijn werk als fysiotherapeut gecombineerd met verschillende andere werkzaamheden. Tussen 2011 en 2014 gaf hij vaardigheidslessen neurorevalidatie op de Hogeschool Utrecht en was hij junior onderzoeker bij het lectoraat gezondheid en bewegen. Vanaf 2016 is hij zich gaan specialiseren in fysiotherapie bij neuromusculaire aandoeningen binnen het spierziekten team van het UMCU. In 2016 begon hij als trainer binnen het treatment and research institute for the cure of ALS (TRICALS) voor het meten van isometrische spierkracht bij medicijn studies. Tussen 2016 en 2019 werkte hij aan de behandelrichtlijn fysiotherapie bij ALS en in 2017 begon hij met zijn promotietraject. Voor het ALS Centrum geeft hij regelmatig lezingen en workshops, waaronder de landelijke cursus 'respiratoire functietesten en behandelmogelijkheden bij ALS', en is hij betrokken bij verschillende innovatieprojecten. Hij woont samen met Jannemieke en samen hebben zij 2 zoons; Simon (8) en Camiel (6).

List of publications not included in this thesis

International

van Unnik JW, Bakers J, Kokx S, van den Berg LH, Visser-Meily JMA, Beelen J, van Eijk RPA. Portable fixed dynamometry enables home-based and reliable assessment of muscle strength in patients with amyotrophic lateral sclerosis: a pilot study. *Submitted*

Helleman J, Kruitwagen-van Reenen ET, Bakers J, Kruithof WJ, van Groenestijn AC, Jaspers Focks RJH, et al. Using patient-reported symptoms of dyspnea for screening reduced respiratory function in patients with motor neuron diseases. J Neurol. 2020;267(11):3310-8.

Goedee HS, Sleutjes B, Bakers JNE, Kruithof WJ, Kruitwagen-van Reenen ET, van der Pol WL. Electrophysiology of fatigue in chronic inflammatory demyelinating polyneuropathy: Can it be useful? Clin Neurophysiol. 2020;131(12):2912-4.

Outermans J, Pool J, van de Port I, Bakers J, Wittink H. What's keeping people after stroke from walking outdoors to become physically active? A qualitative study, using an integrated biomedical and behavioral theory of functioning and disability. BMC Neurol. 2016;16(1):137.

Nijland R, Kwakkel G, Bakers J, van Wegen E. Constraint-induced movement therapy for the upper paretic limb in acute or sub-acute stroke: a systematic review. Int J Stroke. 2011;6(5):425-33.

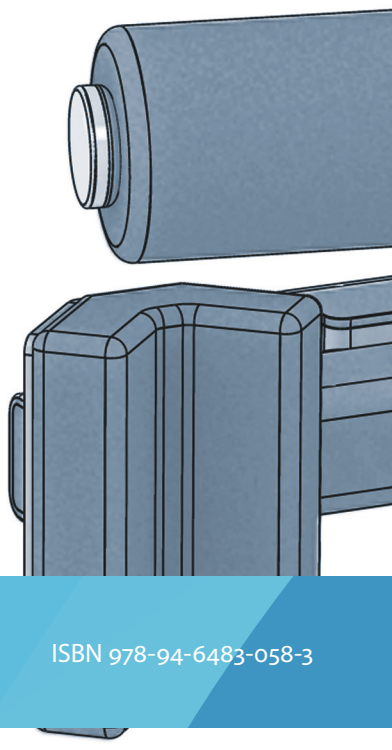
National

Nynke de Mochy, Jaap Bakers. Een bijdrage leveren aan de transitie van zorg in de kliniek naar zorg aan huis. Keypoint. 2022; augustus

Jaap Bakers, Femke Driehuis. Multidisciplinaire richtlijn ALS, PSMA en PLS herzien. Fysiotherapie bij ALS. Fysiopraxis. 2020; februari

Jaap Bakers, Anita Beelen. Behandelrichtlijn fysiotherapie bij ALS. In samenwerking met Koninklijk Nederlands Genootschap voor Fysiotherapie (KNGF) Spierziekten Nederland ALS Patients Connected. ALS Centrum Nederland; 2019

Jaap Bakers, Olaf Verschuren, Mia Willems, Anne Visser-Meily. Elke beweging telt. Een innovatief project in CVA-zorg. Nederlands Tijdschrift voor Revalidatie 2015; 157(3)



UMC Utrecht

Universiteit Utrecht

ISBN 978-94-6483-058-3