Duncan Millar, J. , van Wijck, F., Pollock, A. and Ali, M. (2021)
International consensus recommendations for outcome measurement in
poststroke arm rehabilitation trials. *European Journal of Physical and
Rehabilitation Medicine*, 57(1), pp. 61-68. (doi: 10.23736/S1973-
9087.20.06575-2)

The material cannot be used for any other purpose without further
permission of the publisher and is for private use only.

There may be differences between this version and the published version.
You are advised to consult the publisher's version if you wish to cite from
it.

<http://eprints.gla.ac.uk/230312/>

Deposited on 15 April 2021

Enlighten – Research publications by members of the University of
Glasgow
<http://eprints.gla.ac.uk>

Title Page

Title

International consensus recommendations for outcome measurement in post-stroke arm rehabilitation trials

Authors

Julie Duncan Millar¹, Frederike van Wijck², Alex Pollock¹, and Myzoon Ali¹

Affiliations and addresses

¹Nursing, Midwifery and Allied Health Professions (NMAHP) Research Unit,
Glasgow Caledonian University, Glasgow, UK

² School of Health and Life Sciences, Glasgow Caledonian University,
Glasgow, UK

Corresponding author

Julie Duncan Millar, School of Nursing and Healthcare, University of Glasgow,
57-61 Oakfield Avenue, Glasgow, G12 8LL. United Kingdom.

Julie.duncanmillar@glasgow.ac.uk

Itemised list of three figures in article

Figure 1: Overview of study

Figure 2: Outcome measures flow diagram

Figure 3: Consensus recommendations arranged within ICF categories

Abstract

Background

Existing randomised controlled trials (RCTs) of arm rehabilitation interventions after stroke use a wide range of outcome measures, limiting ability to pool data to determine efficacy. Published recommendations also lack stroke survivor, carer and clinician involvement specifically about perceived relevance and importance of outcomes and measures.

Aim

To generate international consensus recommendations for selection of outcome measures for use in future stroke RCTs in arm rehabilitation considering outcomes important to stroke survivors, carers and clinicians. The recommendations are the Standardising Measurement in Arm Rehabilitation Trials [SMART] toolbox.

Design

Two-round international e-Delphi survey and consensus meeting

Setting

Online and University

Population

Fifty-five researchers and clinicians with expertise in stroke upper limb rehabilitation from 18 countries (e-Delphi); n=13 researchers and clinicians, n=2 stroke survivors, n=1 carer (consensus meeting).

Method

Using systematically identified outcome measures from published RCTs, we conducted a two-round international e-Delphi survey with researchers and clinicians to identify the most important measures for inclusion in the toolbox.

Measures that achieved $\geq 60\%$ consensus were categorised using the International Classification of Functioning, Disability and Health framework (ICF); psychometric properties were ascertained from literature and research resources. At a final consensus meeting, expert stakeholders selected measures for inclusion in the toolbox.

Results

E-Delphi participants recommended 28/170 measures for discussion at the final consensus meeting. Expert stakeholders (n=16) selected the Visual Analogue Scale for pain/ 0-10 Numeric Pain Rating Scale, Dynamometry, Action Research Arm Test, Wolf Motor Function Test, Barthel Index, Motricity Index and Fugl-Meyer Assessment (upper limb section of each), Box and Block Test, Motor Activity Log 14, Nine Hole Peg Test, Functional Independence Measure, EQ-5D, modified Rankin Scale and Canadian Occupational Performance Measure for inclusion in the toolbox.

Conclusion

The SMART Toolbox provides a refined selection of measures that capture outcomes considered important by stakeholders for each ICF domain.

Clinical Rehabilitation Impact

The toolbox will facilitate data aggregation for efficacy analyses thereby strengthening evidence to inform clinical practice. Clinicians can also use the toolbox to guide selection of measures ensuring a patient-centred focus.

Key words: Stroke, outcome measure, upper limb, rehabilitation, outcome, consensus.

Main text

Introduction

Stroke survivors, carers and clinicians have highlighted upper limb (arm) rehabilitation after stroke as a research priority.^{1,2} However, identifying effective arm rehabilitation interventions has been hampered by heterogeneous assessment of outcomes across research studies; there are over 144 different³ and 48 arm-specific⁴ outcome measures currently used in randomised controlled trials (RCTs) on arm rehabilitation after stroke. To strengthen stroke rehabilitation research, various initiatives have recommended improving data comparability in stroke RCTs to facilitate data pooling within meta-analyses to inform practice.^{5,6} The Stroke Recovery and Rehabilitation Roundtable (SRRR) have generated consensus recommendations on measures for use across all stroke sensorimotor recovery and rehabilitation trials. The SRRR recommendations span the International Classification of Functioning, Disability and Health (ICF) framework.⁷ The ICF provides a common language for outcome measurement and is commonly used for classification of outcomes and measures within core outcome projects. For the upper limb specifically, the SRRR provided consensus on measurement of quality of upper limb movement to explore behaviour restitution at the body function and activity level of ICF.⁸ Alt Murphy et al.,⁹ generated recommendations for robot-mediated arm rehabilitation RCTs based on their systematic review of psychometric properties for arm specific measures. Santisteban et al.,⁴ created an inventory of arm-specific measures used in arm rehabilitation RCTs. However these studies did not take into account the outcomes of importance to stroke survivors, carers and

clinicians, which is key to the applicability of recommendations.¹⁰ Recent work has shown that outcomes that are important to stroke survivors and carers, such as 'Independence, freedom and autonomy', 'Ability to work' and 'Coping and self-management', are rarely captured by measures commonly used in stroke arm rehabilitation RCTs.³ This under-representation of stroke survivor and carer priorities in outcome assessment is a matter of concern and represents a gap in upper limb outcome measure recommendations. Integrating the views of these stakeholders into outcome measure selection would ensure that researchers use measures that are not only valid and feasible, but - importantly - target outcomes that matter to people affected by stroke.

Aims

To generate international consensus recommendations for selection of outcome measures for use in future stroke RCTs in arm rehabilitation considering outcomes important to stroke survivors, carers and clinicians (the Standardising Measurement in Arm Rehabilitation Trials [SMART] toolbox). This toolbox was not designed as a prescriptive core outcome set, but rather as a resource to inform the selection of measures of importance to stroke survivors, carers, clinicians and researchers in future RCTs, whilst considering psychometric properties, the ICF classification and outcomes important to stroke survivors, carers and clinicians.

Method

Ethics

Ethical approval was granted by Glasgow Caledonian University School of Health and Life Sciences Ethics Committee for the e-Delphi study (HLS/NCH/16/06) and consensus meeting (HLS/NCH/16/017).

Design

Following preparatory work that included generating a comprehensive inventory of the outcome measures that are currently used in stroke upper limb RCTs and the outcomes of importance to stroke survivors, carers and clinicians,³ we used a two round e-Delphi survey followed by a face to face meeting to achieve consensus on measures to include in the SMART toolbox (See Figure 1) [insert figure 1 here]. E-Delphi is an anonymous, iterative consensus process using multiple rounds of online questionnaires to reach consensus amongst participants on relevant items.¹² A final consensus meeting permitted a range of participants with suitable experience and expertise to gather to discuss the measures for inclusion in the SMART toolbox before voting on the most important measures for inclusion.

Participants

E-Delphi: Eligible participants were researchers and clinicians with experience in selecting and/or administering measures in stroke arm rehabilitation RCTs or amalgamating data in systematic reviews of clinical trials. Owing to the potential attrition rate,¹² we aimed to retain a minimum of 25 participants who completed round 2.

International consensus meeting: Eligible participants were clinicians and researchers from any country, and stroke survivors with arm impairment

and their carers from Scotland. We aimed to recruit 10-25 participants based on previous consensus work.¹³

Case ascertainment, recruitment and consent

E-Delphi: We purposively sampled participants to represent a broad range of backgrounds and experiences. Potential participants were identified by contacting authors of published stroke arm rehabilitation RCTs (identified from a Cochrane overview,¹⁴ and the Cochrane Stroke trials register). We also used snowball sampling to reach local, national and international research networks, and advertised on Twitter and ResearchGate. Potential participants contacted the researcher (JDM), provided written informed consent to participate and returned a demographics questionnaire.

Consensus Meeting: Researchers and clinicians were purposively sampled to give a wide range of stakeholders based on their background, geographical area and experience of selecting and/or administering measures in stroke upper limb rehabilitation trials and in clinical practice, or their experience conducting systematic reviews. Stroke survivor and carer participants comprised the project's advisory group members, who had experience of arm rehabilitation after stroke, and had previously been involved in ascertaining the outcomes that were most important to stakeholders.³ Participants expressed interest and gave written informed consent to attend the meeting.

E-Delphi procedure

We previously identified a list of measures from RCTs contained within a Cochrane Overview of systematic reviews for arm rehabilitation after stroke¹⁴ as detailed in Duncan Millar et al.,³ and from key systematic reviews of arm measures.^{4,9,15} We defined an outcome measure as a reproducible ‘...scale, scoring system, questionnaire or other tool used for measuring an outcome’^{16,p.214}, that must be undertaken at baseline and repeated at least one other time point. Participants were provided with the list of outcome measures prior to starting the e-Delphi and invited to identify any additional measures that they deemed essential for discussion at the consensus meeting. Material for the e-Delphi was compiled in the form of a comprehensive list of all identified measures plus any additional measures that were recommended by participants. Participants were provided with a summary of each of the measures in an outcome measure card. Each card detailed, per measure, relevant information about practical application, including importance of the outcome to stroke survivors, carers and clinicians as described in preparatory work for this study³, administration, equipment requirements and indicative costs (see supplementary information 1 for an example Supplementary Digital Material 1: Supplementary table 1).

In round 1 of the e-Delphi hosted on DelphiManager, participants scored whether each measure on the list was essential for consideration at the consensus meeting using a Likert scale of 1-10 (1 to 3 = limited importance; 4 to 6 = important but not essential; 7 to 9 = essential; 10 unable to score), whilst considering the following questions:

- “Is this measure important to you?”
- Is it important to stroke survivors, carers and clinicians?

- Is it feasible in a research context?”

Feedback on group scores from round 1 was compiled and fed into round 2, which involved re-scoring the same measures as in round 1, with consideration of this feedback as well as the participant's original round 1 score.

There is currently no agreement on the criterion that defines 'consensus'. Therefore, in line with other consensus activities,¹² we defined, a priori, consensus as agreement in $\geq 70\%$ of participants on a rating of 7 to 9 (essential) for the measure; measures that achieved this level of consensus were all taken forward to the final consensus meeting. Those that achieved 60-69% agreement were identified as a second tier of consensus and would only be taken forward to the final consensus meeting to enhance discussions if time allowed.

Consensus meeting materials

In order to provide sufficient information to consensus meeting participants, key psychometric properties and ICF classification of measures with $\geq 60\%$ agreement in the e-Delphi were described.

Description of key psychometric properties

We sought to describe content validity and internal consistency (which included structural validity), and/or reliability (inter-rater, intra-rater and/or test-retest) based on recommendations by Prinsen et al.¹¹

One author (JDM) extracted data, rated and reported key psychometric properties (see supplementary information 2 – Supplementary Digital Material

2: Supplementary table 2). Data were then categorised as either: 'sufficient information on property'; 'unclear what information is available'; or 'information available on property but did not meet requirements stipulated in Prinsen et al.¹¹'.

ICF classification

One author (JDM) classified each measure in terms of the ICF by using linking rules¹⁷ for all measures that reached $\geq 60\%$ consensus to the second level of the ICF; as the linking rules do not differentiate between activity and participation, this is described as one domain, and those that were linked to multiple domains were also described. All linked measures were second reviewed (AP, FvW, MA) and discrepancies resolved through discussion. A third reviewer was available but was not required.

Consensus Meeting Procedure

Our international consensus meeting used a modified Nominal Group Technique approach.¹⁸ The modifications to the approach were that consensus meeting participants were presented with the list of measures for discussion rather than silently generating the list and that measures were voted as yes/no rather than ranked priorities. However, the remaining Nominal Group Technique approach of each person being given the opportunity to share their ideas in turn and discussion of ideas before private voting on priorities was intact. Stroke survivor and carer participants met with one of the authors (JDM) prior to the meeting to ensure that they were adequately prepared and informed.

The meeting was held in Glasgow, Scotland May 2017 with the option to join by videolink. At the start of the meeting, participants introduced themselves, and meeting ground rules were agreed to ensure equitable involvement in discussion. Participants had received written information on the measures that achieved $\geq 60\%$ consensus on the e-Delphi (see supplementary information 1). One author chaired the meeting (JDM) and a second author co-facilitated (MA). To enable discussion from all participants, a speech and language therapist assisted stroke survivors, one of whom had aphasia. A note taker annotated the session, and with participants' consent the session was audio recorded to aid accuracy in reporting.

Participants first discussed all measures that had achieved $\geq 70\%$ agreement from e-Delphi participants and then used the 'Responseware' system to anonymously vote to include, exclude or abstain for each item. When discussing and voting on the measures, participants were asked:

- "Should this outcome measure be part of the SMART toolbox?
- Is it important to you?
- Is it important to other stakeholders?
- Is it feasible?
- Are data available on the key psychometric properties?"

A score of $>50\%$ of participants voting to include a measure had been set a priori as a cut-off for inclusion. The results of each vote were immediately shown to participants. After completion of voting on all measures with $\geq 70\%$ agreement from the e-Delphi, participants then screened the items that reached consensus at the level of 60% to 69%, selecting measures from this second tier group to discuss further and vote on, giving consideration to the

items that had already been discussed. Final results were then presented to the group.

Results

E-Delphi

There were 170 measures included in round 1 (n=150 identified by reviewing published trials and reviews; n=20 additional measures identified by e-Delphi participants; see supplementary table 2).

We recruited 55 participants (n=8 clinicians; n= 12 clinician/researchers; n=35 researchers) from 18 countries to participate in the e-Delphi. Their median number of years' experience in trials was 11 years (IQR 8-20). A total of 53 participants completed round 1 (96%) and 43 (78%) completed round 2. Of the 170 measures reviewed, 14 achieved consensus at $\geq 70\%$; 12 additional measures met consensus at 60-69% (see Figure 2). Of these 26 measures, two were expanded to include both their upper limb component and the full measure for discussion in the consensus meeting (i.e. the Fugl Meyer Assessment and Motricity Index) to yield a total of 28 measures for the consensus meeting. [insert figure 2 here]

Psychometrics

We reported on the availability and level of key psychometric properties of the 28 measures that achieved $\geq 60\%$ consensus. The Functional Independence Measure was the only measure for which all relevant constructs were present (see supplementary information 3 – Supplementary Digital Material 3: Supplementary table 3).

ICF Classification

Of the 28 measures that achieved $\geq 60\%$ consensus, we categorised 10 measures as 'Body Function'; three as 'Activity/Participation'; 10 as 'Body Function and Activity/Participation'; three as 'Body function, Activity/Participation and other' ('other' included personal and environmental factors, not classified and not defined); and two could not be classified within the ICF.

Consensus meeting

We recruited 18 participants, of whom 16 attended (n=11 in person; n=5 via videolink). Participants were from the UK (n=12), Turkey (n=1), Sweden (n=1) and Australia (n=2). Participants included: 2 stroke survivors, 1 carer, 13 clinicians and researchers with backgrounds in physiotherapy (n=5), occupational therapy (n=5), medicine (n=2), and bioengineering (n=1).

In total, 28 measures were put forward for discussion at the consensus meeting. Four measures were not discussed due to overlap with other measures: Digital pinch/grip analyser; Shoulder Abduction Finger Extension; Tardieu Scale; and Activity of Daily Living Observation. Therefore 24 measures were fully debated; 15 measures met consensus of $>50\%$ of participants voting to include a measure in the SMART toolbox: Visual Analogue Scale for pain or 0-10 Numeric Pain Rating Scale, Dynamometry, Action Research Arm Test (ARAT), Fugl-Meyer Assessment (upper limb section; FMA-UE), Wolf Motor Function Test, Barthel Index (BI), Motricity Index (upper limb section), Box and Block test, Motor Activity Log 14 (MAL),

Nine Hole Peg Test, Functional Independence Measure, EQ-5D, modified Rankin Scale (mRS) and Canadian Occupational Performance Measure (COPM; see figure 3 and supplementary information 3). Of these, six were patient-reported measures (where BI is used as self-report). Additionally, participants indicated that the National Institute for Health Stroke Scale (NIHSS) should be recommended as a participant descriptor and not an outcome measure. Eight measures did not reach the threshold of >50% consensus and were therefore not included: Accelerometer to measure activity; Fugl Meyer Assessment (full); Nottingham Sensory Test (shortened and revised version); Goal Attainment Scaling; Stroke Impact Scale v3.0; Motricity Index (full); Manual Muscle Test/Medical Research Council; and Modified Ashworth Scale.

Additional recommendations

Consensus meeting participants additionally commented on the practical application of the toolbox, and recommended that arm rehabilitation researchers may wish to consider selecting from a range of different ICF categories, and beyond the ICF since upper limb dysfunction affects a person as a whole and their life after stroke.³ Also, as changes at the body function level do not necessarily relate to changes at activity/participation level,¹⁹ a range of measures may be necessary to capture change at the different ICF levels. It was recommended to consider burden to the stroke survivor/carer when combining measures. For example, if the BI is already undertaken as part of routine care, perhaps it would not be necessary to include the FIM since FIM is based on the BI but takes longer to complete. It was encouraged to consider at least one patient reported measure if this was feasible for the

trial since the use of patient reported measures is encouraged in all clinical trials²⁰ and clinical practice.²¹ Additionally there is evidence that patient reported measures may provide greater insight than clinical measures into mild arm motor deficits.²² It was suggested to consider aphasia friendly measures, or measures that considered hemi-neglect, e.g. vertical VAS for pain, to make the measure inclusive to a range of stroke survivors. Finally, in order to facilitate comparison between other datasets, measures such as EuroQol or mRS were recommended as these were common to other datasets.^{7,23} [insert figure 3 here]

Discussion

We aimed to achieve consensus on a selection of measures of importance to stroke survivors, carers, clinicians and researchers, that are underpinned by key psychometric properties and are considered feasible, for inclusion in a ‘toolbox’ for researchers undertaking stroke arm rehabilitation RCTs. Our findings provide international consensus on 15 recommended measures.

Our results are congruous with the overarching consensus recommendations for trials exploring sensorimotor recovery following stroke;⁷ however we provided further specific recommendations for trials focusing on arm rehabilitation, using robust consensus methods with stakeholders including service users and providers. We identified consistency with SRRR recommendations with respect to the inclusion of the ARAT, FMA-UE section, EQ-5D, mRS and the NIHSS (patient descriptor)⁷ and dynamometry⁸ in future RCTs. Interestingly, quality of movement was considered by researchers as

important to understand behavioural restitution for upper limb recovery,⁸ however, this was not considered important by stroke survivors.³ Instead, stroke survivors were interested in being able to independently undertake movements to help them complete life tasks more easily. Thus, the SMART toolbox complements existing recommendations^{7,8} but adds the important dimension of patient-centredness by recommending outcomes that reflect matters considered important by those affected by stroke. For example, the COPM (a patient-reported outcome measure) describes stroke survivor's perception of their satisfaction and performance in their everyday life. Thus, the COPM affords the opportunity to capture outcomes important to the stroke survivor that are not currently captured in stroke upper limb trials, such as 'independence, freedom and autonomy', 'ability to work' or 'coping and self-management'. Additionally, the EQ-5D provides data that enables comparison of quality adjusted life years and economic analyses between datasets. Furthermore, our stakeholders recommended that the mRS, a descriptor of overall disability, was appropriate for inclusion, since the mRS is also recommended in acute stroke settings²³ and could thereby enable comparison with other datasets. Participants identified the need to compare different stroke populations and datasets, and acknowledged the value of these common measures. Thus, our results are the consequence of participants also taking into account the wider body of stroke rehabilitation research when considering their recommendations. Taken together, the SMART toolbox complements and augments existing recommendations by considering outcomes of importance to stroke survivors, carers and clinicians, and by including a more selective range of measures, specific to the

requirements of arm rehabilitation research, from which researchers can select measures based on the particular research questions within their trial.

The SMART toolbox has several applications, including in clinical practice, where it may help to guide outcome measure selection and use. Use of the measures by clinicians, alongside the work of initiatives such as the International Consortium for Health Outcomes Measurement²¹ and existing clinical audit could produce data on outcomes in clinical practice, to help to better understand areas of excellence, or the need for improvement to meet standards of clinical care.

A strength of our study was that it involved international representation. However, whilst including participants from Europe, North America, South America, Australasia and Asia, views were not gathered from all areas of the world. Furthermore, stroke survivors and carers all came from Scotland, as involving representatives from other countries or other regions within the UK was not logistically feasible for face-to-face meetings. Instead, the consensus meeting involved members of the project's stroke survivor advisory group, who were based in Glasgow. They had a good understanding of the project which was a strength of the design as this enabled these stakeholders to discuss details of the measures and clarify any queries with the author before or during the meeting.

The SMART toolbox contains some measures which are seldom used in stroke arm RCTs, such as EuroQoL EQ-5D, mRS and COPM.³ However, these measures were deemed to capture outcomes that were important to stakeholders, and were therefore identified as important to include in the toolbox. Conversely, a frequently used measure such as the modified

Ashworth Scale was excluded as the relevance to arm rehabilitation following stroke was considered to be limited. However, consensus meeting participants agreed that measures of resistance to passive movement (including spasticity) may be important for trials that focus specifically on these constructs, in which case trialists should select an appropriate additional measure. Alternative tests for spasticity and related constructs were deemed to lack feasibility for most rehabilitation trials and therefore not supported by stakeholders. However, the SMART toolbox does not preclude use of any such measures, but encourages researchers to first select measures from within the toolbox before looking to supplement these with more trial-specific measures if necessary. For example, if a trial focuses on quality of movement of the upper limb, researchers should consider using performance assays or one functional task as recommended by SRRR⁸ and complement these with measures from the SMART toolbox such as patient reported and/or activity/participation/not classified within the ICF measures. Thus, it is important to match the aims of the trial with the measures used to capture primary and secondary outcomes, taking into consideration the outcomes that matter to stakeholders.

Finally, not all measures in the SMART toolbox meet all the recommendations for key psychometric properties described by Prinsen et al.¹¹ However, psychometric criteria were not the primary consideration in this study, instead this study focused on outcomes of importance to stroke survivors, carers and clinicians. Therefore, this study highlights areas where more research is needed and allows researchers to make an informed decision when selecting measures. It was out with the scope of the project to

conduct a separate systematic review of each measure's psychometric properties or to undertake a COSMIN quality assessment. However, systematic reviews for measures with unknown psychometric properties and COSMIN quality assessment of measures could be considered in future research.

Conclusion

The SMART toolbox is a collection of outcome measures for use in RCTs on post-stroke arm rehabilitation interventions. It was developed using a systematic and robust approach, with consensus reached amongst a diverse range of key international stakeholders including researchers, clinicians, stroke survivors and carers. It comprises measures that are considered important by these stakeholders, considers outcomes of importance to stroke survivors, carers and clinicians, spans the ICF domains and considers key psychometric criteria. Use of this toolbox, supported by international consensus, will enhance inter-study comparisons and pooling of data for efficacy analyses. With greater comparability, individual patient-level data could be used in datasets to conduct subgroup analyses on different populations and stroke severities. This could identify which intervention works best and for whom, which would inform guidelines and influence clinical practice. This would negate the need for lengthy additional trials in different populations, thus reducing research waste and ultimately improving clinical outcomes in arm rehabilitation after stroke.

Acknowledgements: The authors would like to thank the SMART project PhD advisory group, stroke survivor advisory group, and study participants, consensus meeting speech and language therapist and note taker.

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by a PhD Studentship from Glasgow Caledonian University and NMAHP Research Unit.

Declaration of Conflicting Interests: The Authors declare that there is no conflict of interest

All authors read and approved the final version of the manuscript.

References

- (1) Pollock A, St George B, Fenton M, Firkins L. Top ten research priorities relating to life after stroke. *Lancet Neurology* 2012;11:209-209.
- (2) Sacco RL, Sandercock P, Endres M, Feigin V, Pandian J, Shinohara Y, et al. Review and prioritization of stroke research recommendations to address the mission of the World Stroke Organization: a call to action from the WSO Research Committee. *International Journal of Stroke* 2015;10(A100):4-9.
- (3) Duncan Millar J, Van Wijck F, Pollock A, Ali M. Outcome measurement in stroke arm rehabilitation trials: do existing measures capture outcomes that are important to stroke survivors and carers, and clinicians? *Clin Rehab* 2019; 33(4): 737-749.

(4) Santisteban L, Térémetz M, Bleton J, Baron J-C, Maier M, Lindberg P.

Upper Limb Outcome Measures Used in Stroke Rehabilitation Studies: A Systematic Literature Review. *PloS one* 2016;11(5):e0154792.

(5) Ali M, English C, Bernhardt J, Sunnerhagen K, Brady M. More outcomes than trials: a call for consistent data collection across stroke rehabilitation trials. *International Journal of Stroke* 2013;8(1):18-24.

(6) Saver JL, Warach S, Janis S, Odenkirchen J, Becker K, Benavente O, et al. for the National Institute of Neurological Disorders and Stroke Common Data Element Working Group. Standardizing the Structure of Stroke Clinical and Epidemiologic Research Data: The National Institute of Neurological Disorders and Stroke (NINDS) Stroke Common Data Element (CDE) Project. *Stroke* 2012 April 01;43(4):967-973.

(7) Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *International Journal of Stroke* 2017;12(5):451-461.

(8) Kwakkel G, van Wegen E, Burridge J, Winstein C, van Dokkum L, Alt Murphy M, et al. Standardized measurement of quality of upper limb movement after stroke: Consensus-based core recommendations from the Second Stroke Recovery and Rehabilitation Roundtable. *International Journal of Stroke* 2019 33(11):951-958.

- (9) Alt Murphy M, Resteghini C, Feys P, Lamers I. An overview of systematic reviews on upper extremity outcome measures after stroke. *BMC Neurology* 2015(1):29.
- (10) Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *J Health Serv Res Policy* 2012 Jan;17(1):1-2.
- (11) Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set”—a practical guideline. *Trials* 2016;17(1):449.
- (12) Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS medicine* 2011;8(1):e1000393.
- (13) Moher D, Schulz KF, Simera, Altman D. Guidance for developers of health research reporting guidelines. *PLoS Medicine*. 2010; 7(2): e1000217.
- (14) Pollock A, Farmer SE, Brady MC, Langhorne P, Mead G, Mehrholz J, et al. Interventions for improving upper limb function after stroke. *Cochrane Database of Systematic Reviews* 2014(11).
- (15) Lemmens RJ, Timmermans AA, Janssen-Potten YJ, Smeets R, Seelen H. Valid and reliable instruments for arm-hand assessment at ICF activity level in persons with hemiplegia: a systematic review. *BMC Neurol* 2012 Apr 12;12:21-2377-12-21.

- (16) Gethin G, Killeen F, Devane D. Heterogeneity of wound outcome measures in RCTs of treatments for VLU: a systematic review. *J Wound Care* 2015; 05/02: 211-226.
- (17) Cieza A, Fayed N, Bickenbach J, Prodinger B. Refinements of the ICF Linking Rules to strengthen their potential for establishing comparability of health information. *Disabil Rehabil* 2016:1-10.
- (18) De Ruyter K. Focus versus nominal group interviews: a comparative analysis. *Marketing Intelligence & Planning* 1996;14(6):44-50.
- (19) Platz T, Pinkowski C, van Wijck F, Kim I, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. *Clin Rehabil* 2005; 19(4):404-411.
- (20) Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. *Trials* 2017;18(1):122.
- (21) Salinas J, Sprinkhuizen SM, Ackerson T, Bernhardt J, Davie C, George M, et al. An International Standard Set of Patient-Centered Outcome Measures After Stroke. *Stroke* 2016 Jan;47(1):180-186.
- (22) Stewart JC, Cramer SC. Patient-reported measures provide unique insights into motor function after stroke. *Stroke* 2013 04;44(4):1111-1116.
- (23) Lees KR, Bath PM, Schellinger PD, Kerr D, Fulton R, Hacke W, et al. Contemporary outcome measures in acute stroke research: choice of primary outcome measure. *Stroke* 2012 Apr;43(4):1163-1170.