


# Addressing the targeting range of the ABILHAND-56 in relapsing–remitting multiple sclerosis: A mixed methods psychometric study

Sophie Cleanthous, Sara Strzok, Farrah Pompilus, Stefan Cano, Patrick Marquis, Stanley Cohan, Myla D Goldman, Kiren Kresa-Reahl , Jennifer Petrillo , Carmen Castrillo-Viguera, Diego Cadavid and Shih-Yin Chen

Multiple Sclerosis Journal—  
Experimental, Translational  
and Clinical

April–June 2018, 1–12

DOI: 10.1177/  
2055217318776990

© The Author(s), 2018.  
Reprints and permissions:  
[http://www.sagepub.co.uk/  
journalsPermissions.nav](http://www.sagepub.co.uk/journalsPermissions.nav)

## Abstract

**Background:** ABILHAND, a manual ability patient-reported outcome instrument originally developed for stroke patients, has been used in multiple sclerosis clinical trials; however, psychometric analyses indicated the measure's limited measurement range and precision in higher-functioning multiple sclerosis patients.

**Objective:** The purpose of this study was to identify candidate items to expand the measurement range of the ABILHAND-56, thus improving its ability to detect differences in manual ability in higher-functioning multiple sclerosis patients.

**Methods:** A step-wise mixed methods design strategy was used, comprising two waves of patient interviews, a combination of qualitative (concept elicitation and cognitive debriefing) and quantitative (Rasch measurement theory) analytic techniques, and consultation interviews with three clinical neurologists specializing in multiple sclerosis.

**Results:** Original ABILHAND was well understood in this context of use. Eighty-two new manual ability concepts were identified. Draft supplementary items were generated and refined with patient and neurologist input. Rasch measurement theory psychometric analysis indicated supplementary items improved targeting to higher-functioning multiple sclerosis patients and measurement precision. The final pool of Early Multiple Sclerosis Manual Ability items comprises 20 items.

**Conclusion:** The synthesis of qualitative and quantitative methods used in this study improves the ABILHAND content validity to more effectively identify manual ability changes in early multiple sclerosis and potentially help determine treatment effect in higher-functioning patients in clinical trials.

**Keywords:** Manual ability, multiple sclerosis, ABILHAND, patient-reported outcomes, Rasch measurement theory

Date received: 31 October 2017; Revised received 26 March 2018; accepted: 13 April 2018

## Introduction

In addition to walking disability, cognitive problems, depression, and fatigue, manual disability is a prominent problem for many people with multiple sclerosis (MS)<sup>1–4</sup> that affects the ability to perform essential activities of daily living efficiently and independently.<sup>1,5</sup> Manual disability is common<sup>2–4,6</sup> even in the early or mild stages of the disease, with

up to 60% of patients reporting symptoms in the first year post-diagnosis.<sup>2</sup> Therefore, change in manual ability is an important aspect to monitor in clinical practice for disease progression or therapeutic effect. Traditionally, in clinical trials, manual ability has been assessed using performance outcome measures, such as the Nine-Hole Peg Test (9HPT).<sup>7,8</sup> These assessments, although practical

Correspondence to:  
**Jennifer Petrillo**,  
225 Binney Street,  
Cambridge, MA 02142,  
USA.

[jennifer.petrillo@  
biogen.com](mailto:jennifer.petrillo@biogen.com)

**Sophie Cleanthous**,  
Modus Outcomes, UK

**Sara Strzok**,  
Modus Outcomes, USA



**Farah Pompilus,**  
Modus Outcomes, USA

**Stefan Cano,**  
Modus Outcomes, UK

**Patrick Marquis,**  
Modus Outcomes, USA

**Stanley Cohan,**  
Providence Multiple  
Sclerosis Center, USA

**Myla D Goldman,**  
University of Virginia, USA

**Kiren Kresa-Reahl,**  
Providence Multiple  
Sclerosis Center, USA

**Jennifer Petrillo,**  
Biogen, USA

**Carmen Castrillo-Viguera,**  
Biogen, USA

**Diego Cadavid,**  
Fulcrum Therapeutics, USA

**Shih-Yin Chen,**  
Biogen, USA

for use in clinical settings, are not by themselves informative about the daily life impact of MS (and potential treatment benefit) on patients' manual ability. Therefore, more robust patient-reported outcomes (PROs) of manual ability are needed for pivotal clinical trials and in the usual care setting to assess treatment benefit from the patients' perspective.

ABILHAND is a PRO instrument originally developed to assess manual disability in stroke<sup>9</sup> but has recently been used in clinical trials for MS.<sup>10–12</sup> It is essential to evaluate the extent to which any PRO instrument provides valid measurement, and appropriately reflects the patient experience in any new context of use.<sup>13,14</sup> This may be achieved through the discipline of psychometrics<sup>15</sup> where three paradigms exist: traditional psychometrics based on classical test theory (CTT),<sup>16</sup> and modern psychometrics including Rasch measurement theory (RMT)<sup>17,18</sup> and item response theory (IRT). A previous CTT study of ABILHAND-23 in MS suggested adequate reliability and validity.<sup>19</sup> However, subsequent RMT evaluations of ABILHAND-23<sup>19,20</sup> and ABILHAND-56<sup>20</sup> indicated limited measurement range and precision (i.e., increased error associated with measurement) in MS patients with Expanded Disability Status Scale (EDSS) levels between 0–2, which impact ABILHAND's ability to detect differences in manual ability in higher-functioning MS patients. Additional item fit analyses further suggested that there is probably more than one clinical concept related to manual ability underlying the scale; these concepts are “fine motor” (dexterity) and “power.”<sup>20</sup>

Given these limitations, the goal of the study presented here was to troubleshoot the ABILHAND-56 to increase its applicability to the broadest possible population of patients with MS. As ABILHAND-56 is used on an ongoing basis in a specific drug development program, addressing ABILHAND's measurement limitations in higher-functioning MS patients is important to improve measurement range, precision, and potential to detect treatment effect, and subsequently confirm the item clarity and relevance in MS. In this multi-phase, mixed methods study, we aimed to build on previous work by identifying additional candidate items to build on the two clinical concepts underpinning the ABILHAND-56, and thus to improve its ability to detect differences in manual ability in higher-functioning MS patients.

## Materials and Methods

### Study Design Overview

We used a step-wise mixed methods design strategy comprising two waves of patient interviews, a combination of qualitative and quantitative analytic techniques, and consultation interviews with three clinical neurologists specializing in MS (see Figure 1).

Mixed methods design is broadly defined as the combination and comparison of multiple data sources, data collection, analytical procedures, or research methods.<sup>21</sup> In psychometric research, mixed methods specifically refers to the synthesis of qualitative and quantitative methods to identify, define and operationalize PRO instruments as measures of a given concept of interest in a specific context of use.<sup>14</sup>

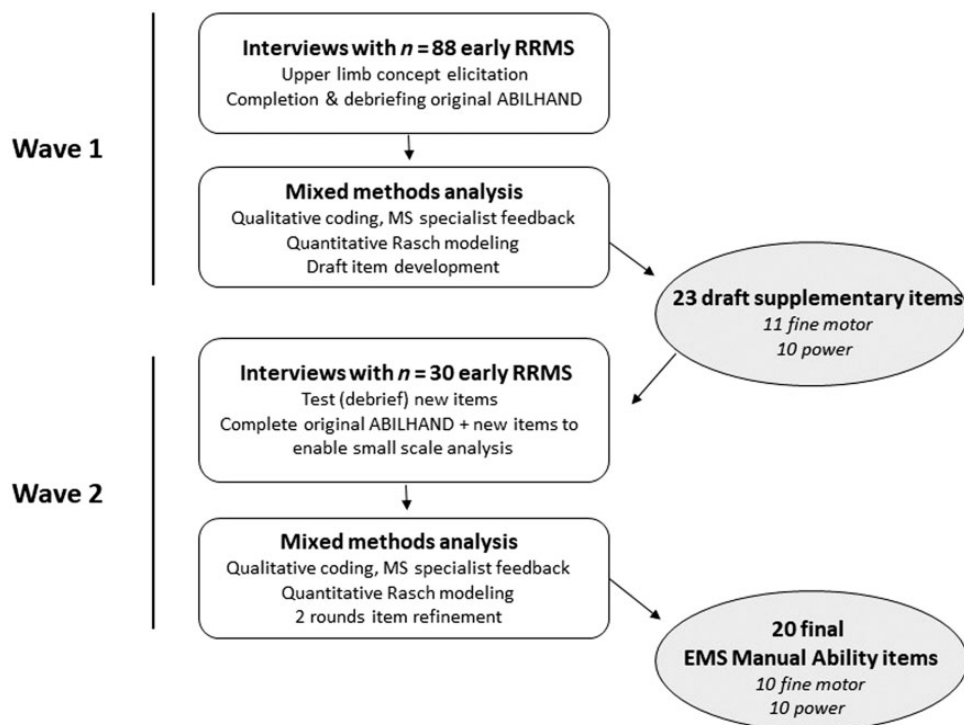
### Study Population and Recruitment Process

Institutional review board approval was obtained, and written informed consent was provided by all study participants. Early relapsing–remitting MS (RRMS) patients were recruited through the study sponsor's patient services department and through a social media site for MS patients. Patients were eligible to participate if they were diagnosed with RRMS within the last two years and had a Patient Determined Disease Steps (PDDS)<sup>22</sup> score of 0–1 (no to mild disability). The PDDS range was selected to coincide with the EDSS 0–2 levels where previous research indicated limitations in the ABILHAND's measurement range and precision.

### Patient Interviews

In Wave 1, concept elicitation interviews were used to identify aspects of manual ability relevant to this patient sample. This was to guide identification of new items that could be used to supplement the ABILHAND. We then asked patients to complete the ABILHAND-56 to further assess its relevance in early RRMS.

In Wave 2, we conducted cognitive debriefing interviews to establish relevance, clarity, and ease of completion of the draft supplementary items that were generated in Wave 1. A “think aloud” process was followed where patients were asked to complete the items while thinking aloud and specifically noting any queries, problems, or ambiguities of the questionnaire.<sup>23</sup> All interviews were conducted over the telephone; the ABILHAND-56 and supplementary items were displayed on patients' computer screens and item responses captured via an online platform. Interviews were audio-taped and transcribed verbatim. In



**Figure 1.** Study overview. EMS: Early Multiple Sclerosis; MS: multiple sclerosis; RRMS: relapsing–remitting multiple sclerosis.

addition, consultation interviews with three neurologists specializing in MS (SCohan, MDG, KKR) were conducted at each of the two waves.

### Materials

Based on the findings of our previous psychometric analysis,<sup>20</sup> an expanded four-level response scale, *very easy*, *easy*, *difficult*, and *impossible*, was used to improve the ABILHAND-56's potential to capture manual disability in this early RRMS sample.<sup>9</sup>

### Data Analysis

#### *Qualitative analysis – concept elicitation.*

Transcripts were analyzed thematically<sup>24</sup> using detailed line-by-line coding<sup>25</sup> to examine, compare, and develop treatment benefit conceptual domains using ATLAS.ti software.<sup>26</sup> Coding was targeted to manual ability. Codes and quotations were inductively categorized into overarching domains that reflected their conceptual underpinning. Each code was compared with the rest of the data to create analytical domains and sub-domains. Saturation was assessed by ordering interviews chronologically, then grouping these into quantiles and comparing concepts emerging by each sequential quantile to assess whether saturation had been reached (i.e., no new concepts emerged).

*Qualitative analysis – cognitive debriefing.* This analysis aimed to identify any potential wording ambiguities and assess relevance and acceptability in relation to each question item, response scale and set of instructions as well as identify additional items that could expand the measurement of manual disability in early RRMS.<sup>23</sup>

*Item generation.* Item generation followed item construction principles,<sup>13,27–29</sup> aiming to have an adequate range of items to cover the conceptual breadth within each of the upper limb mobility sub-domains. Concepts chosen for item development were activities that were applicable to the broadest range of people with MS. Lay language was used in item constructions, using as many of the patients' own words as possible while aiming for brevity and minimal semantic overlap.

*Quantitative data analysis.* A small-scale RMT analysis was performed on data available for the ABILHAND-56 at Wave 1 and ABILHAND-56 as well as supplementary items at Wave 2 using RUMM2030 analytical software.<sup>30</sup> RMT analysis compares observed data against the stringent criteria of the Rasch model, broadly aiming to assess the sample-to-scale targeting, the measurement continuum, and sample measurement.<sup>31,32</sup> Considering the

small sample size, which would not permit any confirmatory conclusions to be made about the items' measurement properties, the focus of this quantitative analysis was to improve to scale targeting. Targeting refers to the match between the distribution of a construct (e.g., manual disability) in the sample and the range of the construct measured by a PRO instrument.<sup>33,34</sup> The better this match is, the greater the potential for accurate evaluation of a PRO instrument and accurate person measurement. Results were interpreted with reference to published criteria wherever possible.<sup>32</sup>

## Results

### Study Sample

RRMS patients ( $n=88$ ), with an RRMS diagnosis <27 months, participated in Wave 1 interviews,

69.3% ( $n=61$ ) of whom reported difficulties with manual ability at screening (Table 1).

### Wave 1 Qualitative Results

**Concept elicitation.** Eighty-two unique codes related to manual disability were identified. Seventy-five of these emerged as “upper limb” concepts in initial coding; seven additional upper limb concepts were identified in retrospective review of activity limitation concepts. Inductive categorisation of these concepts into higher order sub-domains and domains replicated the two-level manual disability conceptual structure suggested in earlier work.<sup>20</sup> Early RRMS patients indicated issues with upper limb mobility related to dexterity that were categorised under the “fine motor” sub-domain as well as issues related to strength categorized under the “power” sub-domain (Table 2). Consultation with the three neurologists

**Table 1.** Sample characteristics.

Patient demographic and clinical characteristics	Wave 1 concept elicitation sample ( $n = 88$ )	Wave 1 RMT analysis sample ( $n = 29$ )	Wave 2 debriefing and RMT sample ( $n = 30$ )
PDSS score ( $n, \%$ )			
0 – normal	44 (50%)	18 (62.1%)	13 (43.3%)
1 – mild disability	44 (50%)	11 (37.9%)	17 (56.7%)
Age in years (mean $\pm$ SD)	40.0 ( $\pm$ 8.72)	38.51 ( $\pm$ 7.66)	35.07 ( $\pm$ 8.11)
Gender ( $n, \%$ )			
Male	23 (26.1%)	7 (24.1%)	7 (23.3%)
Female	65 (73.9%)	22 (75.9%)	23 (76.7%)
Race/ethnicity ( $n, \%$ )			
White	76 (86.4%)	26 (89.7%)	19 (63.3%)
Asian	1 (1.1%)	0 (0%)	1 (3.3%)
Black/African-American	5 (5.7%)	1 (3.4%)	5 (16.7%)
Hispanic/Latino	5 (5.7%)	1 (3.4%)	2 (6.7%)
Mixed race or “other”	1 (1.1%)	1 (3.4%)	3 (10.0%)
Education ( $n, \%$ )			
High school	11 (12.5%)	4 (13.8%)	2 (6.7%)
Some college/associate degree/trade certification	28 (31.8%)	7 (24.1%)	12 (40%)
Bachelor's degree	32 (36.4%)	11 (37.9%)	7 (23.3%)
Post-graduate degree	17 (19.3%)	7 (24.1%)	9 (30.0%)
Employment status ( $n, \%$ )			
Full time	57 (64.8%)	20 (68.9%)	22 (73.3%)
Part time	14 (15.9%)	1 (3.4%)	5 (16.7%)
Not employed	10 (11.4%)	5 (17.2%)	2 (6.7%)
Student	2 (2.2%)	1 (3.4%)	1 (3.3%)
Homemaker	5 (5.7%)	2 (6.9%)	0 (0.0%)

PDSS: Patient Determined Disease Steps; RMT: Rasch measurement theory; SD: standard deviation.

**Table 2.** Examples of patient descriptions under fine motor and power sub-domains.

Upper limb mobility sub-domain	Concept inductive code	Example quote
Fine motor	Brushing teeth	<i>Brushing one's teeth – I would say very easy right now, but when my hands are really numb, it's difficult. – BI-H-88</i>
Fine motor	Computer: mouse use	<i>But when I work with the computer, I can't use the mouse with my right hand. My wrist just gets an attitude, and it just goes wherever it wants. So, I have to use my left hand. – BI-H-55</i>
Fine motor	Using keys	<i>There's things like holding a key to put into a keyhole can be challenging or even making sure I have a good grip on my keys, so I don't drop them. – BI-W-28</i>
Power	Holding telephone	<i>Honestly, when I'm on my cell phone – you know how you just lay on the couch with your phone? I can't (laughter) hold it up with my left arm. I have to prop my arm up and look at my phone. – BI-H-66</i>
Power	Lifting things	<i>I wasn't able to lift the boxes down, put them back up. BI-H-02</i>

specializing in MS was supportive of the two-domain structure.

Saturation analysis indicated that the 88 interviews produced a comprehensive set of concepts with relation to manual disability in higher-functioning people with RRMS; 66 of 75 of the initially identified upper limb mobility concepts arose within the first 30 interviews and the remaining nine concepts either echoed concepts derived from earlier interviews, were not generalizable to the entire MS population, or already existed in the ABILHAND-56.

*Item generation.* Of the identified concepts, 40 of 82 were not covered by existing ABILHAND items; of these, neurologist feedback suggested that 20 of these 40 were more clinically relevant to MS patients with less severe manual disability. This led to the drafting of 23 items: 11 “fine motor” and 12 “power” items. We identified these item sets as Early Multiple Sclerosis Manual Ability – Fine Motor and Early Multiple Sclerosis Manual Ability – Power.

*Cognitive debriefing, item reduction and refinement.* Findings from Wave 2 interviews suggested that 20 of the 23 supplemental items were well-understood and acceptable to patients. However, three items appeared to overlap in

sub-domains. Patients interpreted “washing hair in the shower” and “holding a full bag of groceries” as relating to both lower limb and manual ability. “Holding the steering wheel while driving for a long time” was deemed unclear as patients associated this item with multiple actions (including turning the wheel and shifting gears). Subsequent consultation with neurologists led to removal of the three items not focused on manual ability and to wording revisions of the remaining supplementary items. For example, “inserting a cable into a USB port” was changed to the more widely-applicable task of “inserting a cell phone charging cable into a cell phone.”

*Final supplementary items for ABILHAND in early MS.* Findings from Wave 2 supported a final item pool comprising 10 “fine motor” and 10 “power” Early Multiple Sclerosis Manual Ability items (Table 3).

*Quantitative Results: RMT Psychometric Analysis*  
In line with previous findings,<sup>20</sup> endorsement frequencies indicated that none of the patients endorsed the “impossible” response option for 49 of the 56 ABILHAND items in Wave 1 and 69 of the 79 ABILHAND plus supplemental items in Wave 2. As this lack of endorsement of one of the four categories could artificially inflate the extent of

**Table 3.** ABILHAND plus Early Multiple Sclerosis (MS) Manual Ability items, by theorized sub-scale.

ABILHAND 56-items			
ABILHAND Fine Motor		ABILHAND Power	
AB1	Turning over the pages of a book	AB8	Taking the metallic cap off a bottle
AB2	Pulling up the zipper of trousers	AB11	Closing a door
AB3	Peeling onions	AB12	Washing one's face
AB4	Sharpening a pencil manually	AB17	Opening a screw-topped jar
AB5	Using a spoon	AB20	Tearing open a bag of chips
AB6	Using a screwdriver	AB22	Combing one's hair
AB7	Picking-up a can	AB24	Hammering a nail
AB9	Filing one's nails	AB27	Making pancake batter
AB10	Grasping a coin on a table	AB30	Washing one's hands
AB13	Peeling potatoes with a knife	AB31	Handling a stapler
AB14	Turning off a faucet	AB32	Winding up a wrist watch
AB15	Buttoning up trousers	AB35	Brushing one's hair
AB16	Dialing on a keypad phone	AB42	Cutting meat
AB18	Cutting one's nails	AB43	Eating a sandwich
AB19	Turning on a radio	AB50	Shelling hazel nuts
AB21	Turning on the switch of a lamp	AB51	Screwing a nut on
AB23	Unwrapping a chocolate bar	AB54	Squeezing toothpaste on a toothbrush
AB25	Replacing a light bulb		
AB26	Inserting a diskette into a drive		
AB28	Spreading butter on bread		
AB29	Counting paper money		
AB33	Turning a key in a keyhole		
AB34	Turning on a television set		
AB36	Drawing		
AB37	Ringling a door bell		
AB38	Placing a glass on a table		
AB39	Drinking a glass of water		
AB40	Buttoning up a shirt		
AB41	Threading a needle		
AB44	Handling 4-color ballpoint pen		
AB45	Blowing one's nose		
AB46	Wrapping up gifts		
AB47	Fastening the zipper of a jacket		
AB48	Fastening a snap		
AB49	Writing a sentence		
AB52	Opening mail		
AB53	Typing		
AB55	Taking a coin out of the pocket		
AB56	Brushing one's teeth		
Early MS Manual Ability items			
Fine Motor		Power	
FM01	Using a standard computer mouse	P01	Holding up a book or tablet while reading
FM02	Removing a credit card from slots/ pockets in a wallet	P02	Holding a phone up to one's ear for a long time

(continued)

**Table 3.** Continued

Early MS Manual Ability items			
Fine Motor		Power	
FM03	Removing a single piece of paper from a file folder	P03	Putting heavy items on a shelf above head
FM04	Pushing buttons on a TV remote control or similar device	P04	Taking a heavy item down from a shelf above head
FM05	Texting on a cell/mobile phone	P05	Pulling the cap off a pen
FM06	Opening the metallic tab of a soda can	P06	Opening a safety cap on a medicine bottle
FM07	Plugging an electrical plug into a wall outlet that is easy to reach	P07	Lifting a full pot of water with one handle off stove
FM08	Attaching a cell phone to a charging cable	P08	Filling a kettle with water
FM09	Inserting a key into a keyhole	P09	Lifting a 20-lb weight one time
FM10	Accurately pouring liquids into a measuring cup	P10	Blow drying one's hair

**Table 4.** Overview of Rasch measurement theory (RMT) sample-to-scale targeting results.

ABILHAND scale version	Sample measurement range <sup>a</sup>	Sample measurement mean (SD) <sup>a</sup>	Standard error range	Sample measurements % beyond the scale ceiling <sup>b</sup>
Wave 1 ABILHAND-56	1.35–48.74	40.59 (8.98)	1.68–9.67	3.70% ( <i>n</i> =1)
Wave 1 Fine Motor-39	3.09–50.49	40.39 (9.34)	2.12–9.94	7.49% ( <i>n</i> =2)
Wave 1 Power-17	10.05–49.90	40.50 (7.77)	1.98–10.11	7.49% ( <i>n</i> =2)
Wave 2 ABILHAND-56	9.66–47.45	36.65 (10.95)	1.80–5.46	20.00% ( <i>n</i> =6)
Wave 2 ABILHAND-56 + draft items	21.37–51.41	39.01 (7.75)	1.78–2.83	13.33% ( <i>n</i> =4)
Wave 2 ABILHAND-56 + final items <sup>c</sup>	21.10–51.84	38.87 (7.92)	1.79–2.89	13.33% ( <i>n</i> =4)
Wave 2 Fine motor-39	12.84–46.24	36.88 (10.64)	2.18–5.61	16.67% ( <i>n</i> =5)
Wave 2 Fine motor-39 + draft-items	17.92–47.93	37.86 (9.18)	2.24–3.95	13.33% ( <i>n</i> =4)
Wave 2 Fine motor-39 + final items <sup>c</sup>	16.47–48.42	37.51 (9.76)	2.25–4.18	16.67% ( <i>n</i> =5)
Wave 2 Power-17	2.68–51.27	35.39 (12.72)	3.70–11.61	23.33% ( <i>n</i> =7)
Wave 2 Power-17 + draft items	26.64–58.34	40.82 (6.85)	2.90–3.97	3.33% ( <i>n</i> =1)
Wave 2 Power-17 + final items <sup>c</sup>	27.78–59.07	41.01 (6.64)	2.96–3.94	0.00% ( <i>n</i> =0)

SD: standard deviation.  
<sup>a</sup>Where the scale item range is set to range from 0–100 and item mean always at 50; <sup>b</sup>patients for whom the scale items are too easy; <sup>c</sup>final items as available at Wave 2.

sub-optimal targeting for these analyses, the four-level response scale was rescored into three levels, merging the two higher categories (“very easy” – “easy” – “difficult/impossible”) for this analysis.

Table 4 details the sample-to-scale targeting for the different scale versions at Wave 1 and Wave 2. Findings are presented in an interval 0–100 transformed score, based on the interval logit metric

produced by RMT analysis. In alignment with the sample's PDSS scores, the sample mean was consistently below the scale mean (<50), indicating that these patients lie on the lower end of the manual disability continuum. The supplementary items both in their draft and final form shift the sample measurement means closer to the scale mean for all three different versions of the scale (36.65 to 38.87, 36.88 to 37.51 and 35.39 to 41.01 for the

ABILHAND, fine motor, and power scales respectively).

The range of standard error (SE) associated with measurement is also reduced by the added supplementary items, indicating precision associated with measurement is increased. The highest SE associated with measurement is reduced from 5.46 to 2.89, 5.61 to 4.18 and 11.61 to 3.64 for the three respective scales (Table 4). Finally, the percentage of people at the ceiling (people for whom the scale items are too easy) is reduced by the supplementary items for the ABILHAND-56 and the Early MS Manual Ability sub-scales. Figures 2–4 display the relative improvements to sample-to-scale targeting graphically.

### Discussion

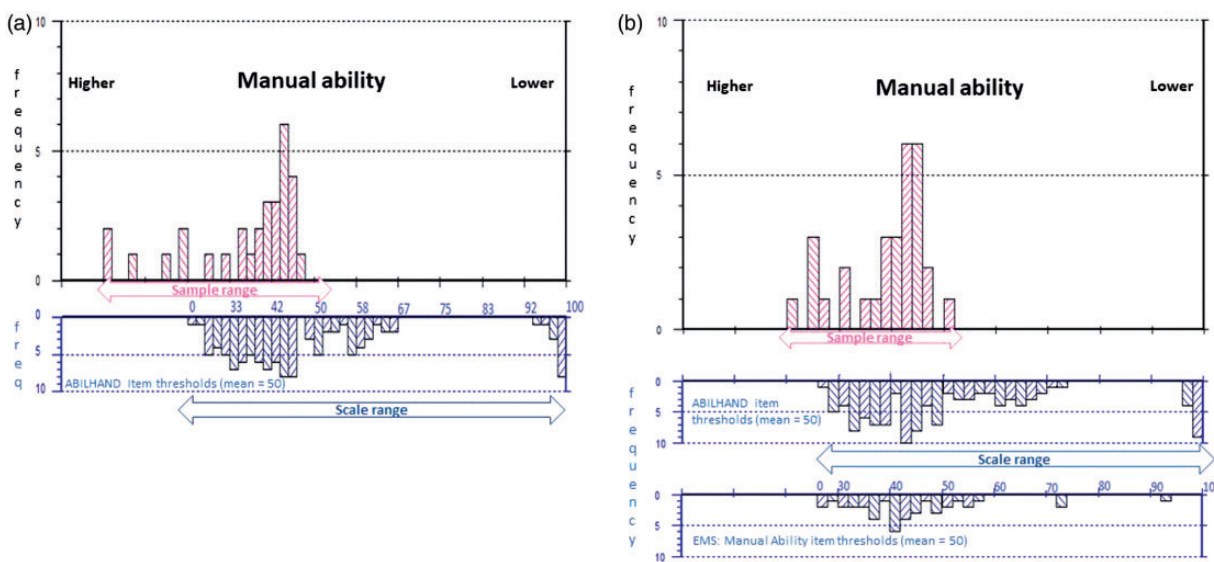
In this multi-phase, mixed-methods psychometric study, we identified 20 additional candidate items to help improve the ABILHAND-56’s ability to detect differences in manual ability in higher-functioning early RRMS patients. The robust development process included patient and clinician feedback as well as modern psychometric analysis.

Wave 1 in-depth qualitative research findings indicated that the majority of existing ABILHAND-56 items were well-understood and appropriate to this

MS sample, confirming the ABILHANDs relevance in this clinical population. In addition, we identified a rich pool of relevant manual ability concepts aligning with the previously-identified two-level fine motor and power manual ability conceptual framework.<sup>20</sup> Clinical neurologists helped ensure that item development focused on the most clinically relevant additional supplementary items to expand the ABILHAND’s measurement range. Wave 2 patient interviews ensured relevance, understanding, and acceptability of the supplementary items, in addition to providing evidence for revision and refinement.

The macro-level psychometric analysis of the addition of the new items, based on RMT, suggests improved targeting in this higher-functioning RRMS sample, with lower ceiling effects and greater precision (the ability to discriminate different levels of manual ability). The analysis also provided evidence that an altered response scale to further improve targeting for higher-functioning patients is needed; this adaptation should therefore be considered for future MS studies using this scale.

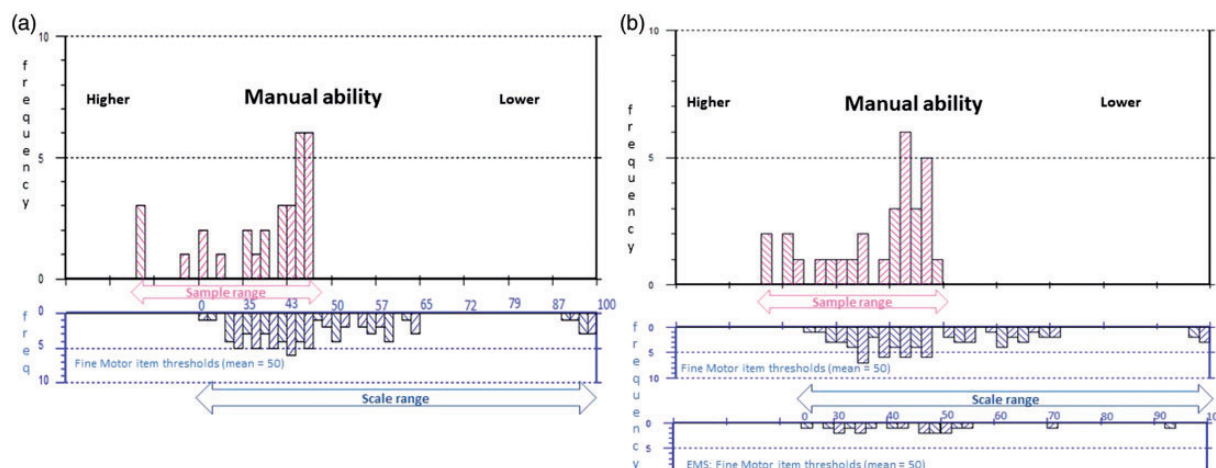
A mixed method psychometric approach advances our understanding of content validity and helps ensure that a PRO instrument adequately reflects the patient experience in a given context.<sup>13,14</sup> This process is vital to maximize clinical interpretability, particularly when scores derived from PROs are



**Figure 2.** ABILHAND-56 sample to scale targeting.

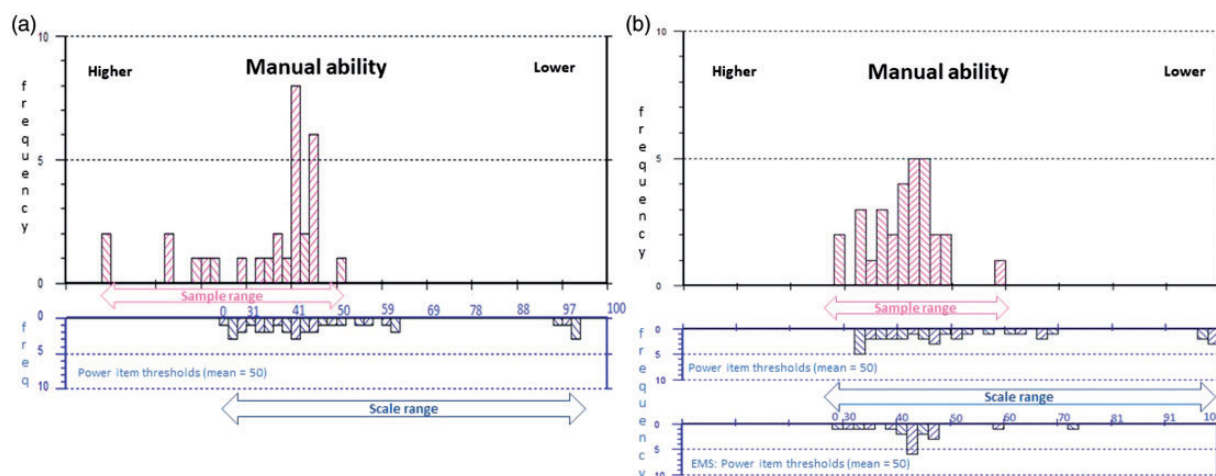
The upper histograms represent the sample distribution, and the lower histograms the scale item threshold distribution plotted on the same interval metric continuum of manual ability. (a) Displays sample-to-scale targeting of the original ABILHAND-56 items and (b) the improvements to the match between sample and scale targeting introduced by merging the original with the Early Multiple Sclerosis (EMS): Manual Ability items. Sample measurements falling off the 0–100 range of the scale indicate patients for whom the scale remains too easy.





**Figure 3.** Fine motor sample to scale targeting.

The upper histograms represent the sample distribution, and the lower histograms the scale item threshold distribution plotted on the same interval metric continuum of manual ability. (a) Displays sample-to-scale targeting of the original Fine Motor 39 items and (b) the improvements to the match between sample and scale targeting introduced by merging the original with the Early Multiple Sclerosis (EMS): Fine Motor items. Sample measurements falling off the 0–100 range of the scale indicate patients for whom the scale remains too easy.



**Figure 4.** Power sample to scale targeting.

The upper histograms represent the sample distribution, and the lower histograms the scale item threshold distribution plotted on the same interval metric continuum of manual ability. (a) Displays sample-to-scale targeting of the original Power 17 items and (b) the improvements to the match between sample and scale targeting introduced by merging the original with the Early Multiple Sclerosis (EMS): Power items. Sample measurements falling off the 0–100 range of the scale indicate patients for whom the scale remains too easy.

used to make decisions about the state of disease and treatment.<sup>35</sup> Our study used a novel mixed methods approach that demonstrates how we can efficiently conduct psychometric research to empirically troubleshoot legacy PRO instruments to ensure they appropriately capture the targeted concept of interest in a specific context of use.<sup>14</sup>

Traditionally, PRO instruments are developed via a three-step approach moving through qualitative concept elicitation and cognitive debriefing, to

quantitative field testing.<sup>36,37</sup> However, we suggest this standard linear methodology limits our ability to efficiently construct items, elaborate upon response options, identify anomalies, and troubleshoot overall instrument design. Therefore, we advocate an integrated, iterative process, prior to PRO instrument field testing. Using this approach, we generated optimal supplementary items for the ABILHAND in MS, which could help improve the match between manual ability in this population and subsequently improve manual ability measurement and

interpretation in MS studies. It is important that the supplemental items only be used in conjunction with the ABILHAND items, as they do not measure the full spectrum of MS manual ability on their own.

The outcome of this study has been the development of a potential new tool, which could be used in clinical practice and clinical trials to measure changes in manual ability in MS from the patients' perspective. Attention to manual ability should be a central focus in clinical management and development of new therapeutic/clinical interventions, including emerging candidate reparative therapies.<sup>38</sup> In the current MS research and treatment landscape, it is increasingly clear that measures need to be targeted to include the highly-functioning population, and need to be sensitive to changes relevant to their functional status, particularly in studies focusing on preserving physical ability of newly diagnosed MS patients or reversing the damage caused by the disease before irreversible axonal loss takes place.<sup>19,20</sup> Findings from this multi-phase mixed methods study indicate that the Early MS Manual Ability items expand manual ability measurement to issues relevant to higher-functioning patients and therefore have the potential to increase sensitivity to detect subtle clinical change in higher-functioning MS patients. The recent treatment effects observed with natalizumab on the 9HPT components of the primary endpoint in patients with advanced non-relapsing secondary progressive multiple sclerosis (SPMS) in the ASCEND natalizumab trial highlight the importance of having robust clinical outcome assessments, including PROs, to measure treatment effects on upper extremity function.<sup>39</sup>

While our findings with Early MS Manual Ability are encouraging, they should be interpreted with consideration of the study's limitations. The structure of the ABILHAND and Early MS Manual Ability item stem ("How difficult are the following activities") is simple and function descriptions are brief; patients reported they were able to complete the items quickly, with few problems. However, given that the enhanced conceptual coverage in higher-functioning people with MS is achieved by adding 20 items to the existing ABILHAND-56, it will be worthwhile to explore the burden presented by adding additional items in future studies. Given that inclusion criteria were based on self-report information and because of the small sample size of the RMT analysis, additional analysis in a larger clinically defined sample would help confirm the validity and generalizability of these findings. The scoring structure of the ABILHAND-56 and Early

MS Manual Ability items is empirically supported by a psychometric analysis in one context and strictly requires further psychometric testing. Finally, the revised scoring structure improves but does not resolve all the measurement issues related to the original ABILHAND-56.

Through mixed methods psychometric research, we generated 20 supplementary items to improve the targeting on ABILHAND-56 in higher-functioning MS patients. The qualitative and quantitative findings support its use in measuring manual ability in MS from the patients' perspective. Further data from a larger clinically defined sample is needed to confirm the new items' measurement properties.

### **Acknowledgments**

The authors wish to acknowledge and thank the 118 patients who shared their MS stories with the research team, as well as JoAnne Liebler, Catherine Podeszwa, and Sasha Spite, project interviewers.

### **Conflict of Interest**

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Shih-Yin Chen, Jennifer Petrillo, and Carmen Castrillo-Viguera are employees of and stockholders in Biogen. Diego Cadavid was an employee of Biogen when the study was conducted. This work is not related to his current employment with Fulcrum Therapeutics. Farrah Pompilus, Sophie Cleanthous, Sara Strzok, Stefan Cano, and Patrick Marquis are employees of Modus Outcomes, which received payment from Biogen Pharmaceuticals to conduct this research. Stanley Cohan receives research support from Biogen, Novartis, Mallinckrodt, Sanofi-Genzyme, Genentech and Opexa, is a paid consultant and/or serves on advisory boards for Biogen, Sanofi-Genzyme, Novartis and has received speaking honoraria, travel expenses and meals/lodging from Biogen, Novartis, Sanofi-Genzyme, Acorda, and Genentech. Myla Goldman has received personal consultancy funds from EMD Serono, Genzyme, and Novartis, and institutional consultancy and/or research funds from Acorda, Biogen Idec, and Novartis Pharmaceuticals, and is grant supported by the National Multiple Sclerosis Society and National Institutes of Health (K23NS062898). Kiren Kresa-Reahl speaker's bureau honoraria: Biogen, Novartis, TEVA, EMDSerono, Mallinckrodt, Genzyme. Consultant services: Biogen, Genentech, EMDSerono. Research Support: Biogen, Novartis, Mallinckrodt, Genzyme, Genentech.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was funded by Biogen.

## ORCID iD

Kiren Kresa-Reahl  <http://orcid.org/0000-0001-7095-1520>

Jennifer Petrillo  <http://orcid.org/0000-0001-8441-2157>

## References

1. Yozbatiran N, Baskurt F, Baskurt Z, et al. Motor assessment of upper extremity function and its relation with fatigue, cognitive function and quality of life in multiple sclerosis patients. *J Neurol Sci* 2006; 246: 117–122.
2. Kister I, Bacon TE, Chamot E, et al. Natural history of multiple sclerosis symptoms. *Int J MS Care* 2013; 15: 146–158.
3. Holper L, Coenen M, Weise A, et al. Characterization of functioning in multiple sclerosis using the ICF. *J Neurol* 2010; 257: 103–113.
4. Johansson S, Ytterberg C, Claesson IM, et al. High concurrent presence of disability in multiple sclerosis: Associations with perceived health. *J Neurol* 2007; 254: 767–773.
5. Bertoni R, Lamers I, Chen CC, et al. Unilateral and bilateral upper limb dysfunction at body functions, activity and participation levels in people with multiple sclerosis. *Mult Scler* 2015; 21: 1566–1574.
6. Krishnan V and Jaric S. Hand function in multiple sclerosis: Force coordination in manipulation tasks. *Clin Neurophysiol* 2008; 119: 2274–2281.
7. Mathiowetz V, Weber K, Kashman N, et al. Adult norms for the nine-hole peg test of finger dexterity. *OTJR (Thorofare N J)* 1985; 5: 24–37.
8. Lamers I, Cattaneo D, Chen CC, et al. Associations of upper limb disability measures on different levels of the International Classification of Functioning, Disability and Health in people with multiple sclerosis. *Phys Ther* 2015; 95: 65–75.
9. Penta M, Tesio L, Arnould C, et al. The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment. *Stroke* 2001; 32: 1627–1634.
10. Cano S, Cleanthous S, Marquis P, et al. Measuring upper limb function in multiple sclerosis: Enhancing the ABILHAND's performance. *Value Health* 2015; 18: A24.
11. Mikol D, Freedman MS, Goldman MD, et al. ASCEND study of natalizumab efficacy on reducing disability in patients with secondary progressive multiple sclerosis: Baseline demographics and disease characteristics. In: *29th Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS)*, Copenhagen, Denmark, 2–5 October 2013, Poster P 1087.
12. Mikol D, Goldman M, Hartung H, et al. The 9-Hole Peg Test (9-HPT) has a stronger correlation than the Expanded Disability Status Scale (EDSS) with patient-reported upper extremity impairment as assessed using ABILHAND in patients with secondary progressive multiple sclerosis (SPMS): Analysis of baseline data from the ASCEND Study. *Neurology* 2015; 84: 7.222. [http://n.neurology.org/content/84/14\\_Supplement/P7.222.short](http://n.neurology.org/content/84/14_Supplement/P7.222.short)
13. Food and Drug Administration. Patient reported outcome measures: Use in medical product development to support labelling claims, <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf> (accessed 28 February 2017).
14. US Food and Drug Administration. Qualification of Clinical Outcome Assessments (COAs). <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm> (accessed 28 February 2017).
15. McDowell I and Newell C. *Measuring health: A guide to rating scales and questionnaires*. 2nd ed. Oxford: Oxford University Press, 1996.
16. Novick MR. The axioms and principal results of classical test theory. *J Math Psychol* 1966; 3: 1–18.
17. Andrich D. *Rasch models for measurement*. Beverley Hills, CA: Sage Publications, 1988.
18. Rasch G. *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Education Research, 1960.
19. Barrett LE, Cano SJ, Zajicek JP, et al. Can the ABILHAND handle manual ability in MS? *Mult Scler* 2013; 19: 806–815.
20. Barrett L, Cano S, Zajicek J, et al. Lending a hand: Can DASH items help ABILHAND improve manual ability measurement in multiple sclerosis? *Mult Scler* 2015; 21: 612–621.
21. Morse J. Evolving trends in qualitative research: Advances in mixed-methods design. *Qual Health Res* 2005; 15: 583–585.
22. Hohol MJ, Orav EJ and Weiner HL. Disease steps in multiple sclerosis: A longitudinal study comparing disease steps and EDSS to evaluate disease progression. *Mult Scler* 1999; 5: 349–354.
23. Blair J and Presser S. Survey procedures for conducting cognitive interview to pretest questionnaires: A review of theory and practice. In: *Proceedings of the Section on Survey Research Methods, Annual Meetings of the American Statistical Association*, 1993; 370: 75. Alexandria, VA: American Statistical Association.
24. Braun V and Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006; 3: 77–90.
25. Bryman A and Burgess B. *Analyzing qualitative data*. New York: Routledge, 2002.
26. Friese S. *ATLAS.ti 7 User Guide and Reference*. Berlin: ATLAS ti Scientific Software Development GmbH.
27. Kline P. *Handbook of test construction: Introduction to psychometric design*. New York: Methuen, 1986.
28. Feinstein A. Clinical biostatistics XLI. Hard science, soft data, and the challenge of choosing clinical variables in research. *Clin Pharmacol Ther* 1977; 22: 485–498.

29. Fowler F. *Improving survey questions: Design and evaluation*. Thousand Oaks: Sage, 1995.
30. Andrich D and Sheridan B. *RUMM 2030*. Perth, WA: RUMM Laboratory Pty Ltd, 1997–2018.
31. Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res* 2011; 11: 571–585.
32. Hobart J and Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods. *Health Technol Assess (Rockv)* 2009; 13: 1–214.
33. Hobart JC, Riazi A, Thompson AJ, et al. Getting the measure of spasticity in multiple sclerosis: The Multiple Sclerosis Spasticity Scale (MSSS-88). *Brain* 2006; 129: 224–234.
34. Wright B and Masters G. *Rating scale analysis*. Chicago: MESA Press, 1982.
35. Hobart J and Cano S. Rating scales for clinical studies in neurology: Challenges and opportunities. *US Neurol* 2008; 4: 12–18.
36. Ware JE, Snow KK, Kosinski M, et al. *SF-36 Health Survey manual and interpretation guide*. Boston, MA: Nimrod Press, 1993.
37. Cano S and Hobart J. The problem with health measurement. *Patient Prefer Adherence* 2011; 5: 279–290.
38. Cadavid D, Balcer L, Galetta S, et al. Safety and efficacy of opicinumab in acute optic neuritis (RENEW): A randomised, placebo-controlled, phase 2 trial. *Lancet Neurol* 2017; 16: 189–199.
39. Steiner D, Arnold D, Freedman M, et al. Natalizumab versus placebo in patients with secondary progressive multiple sclerosis (SPMS): Results from ASCEND, a multicenter, double-blind, placebo-controlled, randomized phase 3 clinical trial. *Neurology* 2016; 87: E22.