

UCSF

UC San Francisco Previously Published Works

Title

An electronic, unsupervised patient-reported Expanded Disability Status Scale for multiple sclerosis.

Permalink

<https://escholarship.org/uc/item/6d45b620>

Journal

Multiple sclerosis (Houndmills, Basingstoke, England), 27(9)

ISSN

1352-4585

Authors

Romeo, Andrew R
Rowles, William M
Schleimer, Erica S
[et al.](#)

Publication Date

2021-08-01

DOI

10.1177/1352458520968814

Peer reviewed



Published in final edited form as:

Mult Scler. 2021 August ; 27(9): 1432–1441. doi:10.1177/1352458520968814.

An electronic, unsupervised Patient Reported Expanded Disability Status Scale for Multiple Sclerosis

Andrew R Romeo^{*1}, William M Rowles^{*1}, Erica S Schleimer¹, Patrick Barba¹, Wan-Yu Hsu¹, Refujia Gomez¹, Adam Santaniello¹, Chao Zhao¹, Jennifer R Pearce², J.B. Jones³, Bruce C Cree¹, Stephen L Hauser¹, Jeffrey M Gelfand¹, UCSF MS Epic Study Group¹, Walter F. Stewart⁴, Douglas S Goodin¹, Riley M Bove¹

¹UCSF MS and Neuroinflammation Center, Weill Institute for Neurosciences, Department of Neurology, Division of Neuroinflammation and Glial Biology, University of California San Francisco, San Francisco, CA

²Plain Language Health, San Francisco, CA

³Sutter Health, Palo Alto Medical Foundation Research Institute, Walnut Creek, CA

⁴Medcurio Inc., Oakland, CA

Abstract

Background.—In persons with multiple sclerosis (MS), the Expanded Disability Status Scale (EDSS) is the criterion standard for assessing disability, but its in-person nature constrains patient participation in research and clinical assessments.

Objective.—To develop and validate a scalable, electronic, unsupervised patient-reported EDSS (ePR-EDSS) that would capture MS-related disability across the spectrum of severity.

Methods.—We enrolled 136 adult MS patients, split into a preliminary testing Cohort 1 (n=50), and a validation Cohort 2 (n=86), which was evenly distributed across EDSS groups. Each patient completed an ePR-EDSS either immediately before or after a MS clinician's Neurostatus EDSS evaluation.

Corresponding Author Riley Bove, Riley.Bove@ucsf.edu, UCSF MS and Neuroinflammation Center, Weill Institute for Neurosciences, Department of Neurology, Division of Neuroinflammation and Glial Biology, University of California San Francisco, 675 Nelson Rising Lane, San Francisco, CA 94158, USA, Telephone: 415 595 2795, Fax: N/A.

*These two authors contributed equally

Disclosures

A. Romeo, W. Rowles, E. Schleimer, P. Barba, R. Gomez, A. Santaniello, J. Pearce, WY Hsu, C. Zhao: none

J. Jones: Dr. Jones has received research support from the California Initiative to Advance Precision Medicine, Roche Genentech, AstraZeneca, Boehringer Ingelheim, and the Hilton Foundation.

B. Cree: Dr. Cree has received personal compensation for consulting from Akili, Alexion, Atara, Biogen, EMD Serono, Novartis, Sanofi and TG Therapeutics.

S. Hauser: Dr. Hauser serves on the Board of Directors for Neuron; Scientific Advisory Board for Alector, Annexon, Bionure, Molecular Stethoscope, and Symbiotix; he has also received nonfinancial support from F. Hoffmann-La Roche Ltd and Novartis AG.

J. Gelfand: Dr. Gelfand reports research support to UCSF from Genentech for a clinical trial and personal fees for consulting from Biogen and Alexion and personal fees for medical legal consulting.

W.F. Stewart: Dr. Stewart serves as consultant to Amgen, Dr. Ready, Allergan and Grifol.

D. S. Goodin: Dr. Goodin reports consultancy, Novartis; speaker honoraria, EMD, Serono, Novartis, Sanofi Genzyme.

R. Bove: Dr. Bove has received research support from the National Multiple Sclerosis Society, the Hilton Foundation, the California Initiative to Advance Precision Medicine, the Sherak Foundation and Akili Interactive. RB has also received personal compensation for consulting from Alexion, Biogen, EMD Serono, Novartis, Sanofi Genzyme, Roche Genentech and Pear Therapeutics.

Results.—In Cohort 2, mean age was 50.6 years (range 26–80) and median EDSS was 3.5 (IQR 1.5, 5.5). The ePR-EDSS and EDSS agreed within 1-point for 86% of examinations; kappa for agreement within 1-point was 0.85 ($p < 0.001$). The correlation coefficient between the two measures was 0.91 (< 0.001).

Discussion.—The ePR-EDSS was highly correlated with EDSS, with good agreement even at lower EDSS levels. For clinical care, the ePR-EDSS could enable the longitudinal monitoring of a patient’s disability. For research, it provides a valid and rapid measure across the entire spectrum of disability and permits broader participation with fewer in-person assessments.

Keywords

Multiple sclerosis; patient reported outcome measures; eHealth; disability

INTRODUCTION

The Expanded Disability Status Scale (EDSS), originally developed by Kurtzke¹, has long been an accepted standard for summarizing and quantifying neurologic disability due to multiple sclerosis (MS). Despite this acceptance, however, this scale has undergone many modifications since its introduction as the DSS in the 1950s.² In fact, over time, investigators have sought to modify even its “expanded” version¹ in order to make the scale more “objective”. The culmination of this process (so far) is the Neurostatus EDSS (NS-EDSS), which is now widely utilized to assess disability in MS clinical trials.² For example, the original (1982) description of EDSS=4.5 states that “the patient must be able to walk without aid or rest for some 300 meters and to work a full day in a position of average difficulty. The patient is up and about most of the day, but some limitation of full activity separates this from step 4.0”. In the NS-EDSS, such descriptions have been changed to exclude the functional requirements for activities of daily life and to make the ambulation distance requirements more explicit. In addition, both scores require that these abilities be observed. There is no doubt that changes such as these have made the EDSS a more objective measure. Whether these modifications have improved the scale, however, is less clear. For example, characterizing someone as “fully ambulatory” because they can walk unaided for 500 meters down a hallway, which has handrails on both sides for safety, misrepresents the disability of a patient who, because of instability, requires a cane whenever they are outside the home and relies on furniture when inside. Moreover, the original EDSS estimates (albeit subjectively) a patient’s integrated function over time. By contrast, the NS-EDSS represents the measured function of a patient by a particular investigator, on a particular day, at a particular time and does this for a disease that is characterized by wide fluctuations in its clinical manifestations depending upon factors such as the specific day, the time of day, the ambient temperature, and the presence or absence of intercurrent illness. Regardless of any such concerns over what is being measured by each of these modified versions of the EDSS scale, however, all of these instruments require an in-person assessment and all suffer from having high inter- and intra-rater variability, especially at the lower disability levels.³ In addition, in large clinical trials, it is usually not feasible to perform in-person evaluations more frequently than once every three months. This limitation would be even greater for any population-based survey of MS in the community.

As a result, there has been considerable interest in developing valid instruments to measure patient-reported functioning and disability.^{4–8} Thus, an electronically-administered patient-reported EDSS (ePR-EDSS) that is valid across a wide EDSS range could, potentially, decrease utilization of resources and personnel time, permit the more frequent assessment of disability status during clinical trials, increase patient access to MS clinical research and, perhaps, also provide a more reliable and stable measure of function over time. We began with the first-reported PR-EDSS (Goodin, 1998⁵), which was validated against the Kurtzke EDSS; because the NS-EDSS is now widely utilized and experience with this scale is so well established, any PR-EDSS for now needs to be validated by using this measure as the nominal “gold standard”. We applied principles of human-centered design^{9, 10} to iteratively refine the original Goodin questionnaire into an online freely accessible ePR-EDSS, that could be performed in an unsupervised fashion. We specifically sought to improve performance in individual cerebral and brainstem functional system scores (FS), and then to validate the ePR-EDSS using the measured NS-EDSS across a wide disability range.

METHODS

Participants

This study included 136 adults (ages 18–80 years) with a diagnosis of MS (by 2017 International Panel criteria)¹¹ or clinically isolated syndrome. Each participant was recruited either from the UCSF longitudinal Expression Proteomics Imaging Clinical (EPIC) study or from the UCSF Multiple Sclerosis and Neuroinflammation Center (February 2018 – January 2020).¹² Patients presenting for an annual research evaluation or a routine clinical follow up (which included an NS-EDSS examination by an MS neurologist previously trained and certified in NS-EDSS) were invited to participate. At the time of the visit, each participant completed the ePR-EDSS on a tablet. Of 138 individuals invited, 136 agreed to participate and two declined because of time constraints. To avoid potential recall or anchoring biases from undergoing a recent in-person neurological examination, participants randomly completed the ePR-EDSS questionnaire either before or after the examination (84 before and 52 after). The delay between the two evaluations was less than 4 hours in 98.5% of cases, and within 24 hours for 2 cases (1.5%).

ePR-EDSS

Development: 3 phases.—An instrument initially developed by Goodin⁵ was iteratively revised using principles of human-centered design. During the first phase (December 2015–June 2016), because of limitations in the previous assessment of “cerebral” and “brainstem” functional system, questions relating to these functional systems were added to the written questionnaire, and this new version was pilot-tested in 20 participants. Examples of changes include adding questions about hearing and swallowing to the questionnaire, which contributed to brainstem FS scoring; and making response options to questions about bladder/bowel and cerebral functions more detailed (vs. leaving the patient to determine whether symptoms were “mild” or “moderate”). During the second phase (September 2016–June 2017), stakeholder input was obtained through a series of interviews (18 individuals with MS, 4 MS clinicians) and focus groups. Focus groups (3 sessions, each with 3–

8 participants with MS, including members of support groups, both for MS generally and also, more specifically, for African-Americans and Hispanic-Americans with MS) were conducted between January and June 2017. A health literacy expert (JP) reviewed language and user experience. Important suggestions from stakeholders included developing a screening page to avoid providing detailed responses for a functional system in which there was no impairment. The instrument was revised based on stakeholder input and the questionnaire was then converted to an openly-available MS BioScreen computerized platform (a Ruby on Rails application; openmsbioscreen.ucsf.edu), which permits MS patients to track their care and disease status.¹³ The original scoring algorithm developed by Goodin⁵ was also iteratively revised to match the changes in the questions. The final algorithm was automated so that patients' responses to questions automatically generate FS and global ePR-EDSS scores online, requiring no human intervention. All code of the customized-program was based on the specifications and logic developed by the authors (DSG, RB, and WFS). The third phase consisted of iterative testing with two cohorts.

Initially, (Cohort 1, n=50, June 2017 to January 2018), 50 adults with MS completed the ePR-EDSS questionnaire that began with screening questions addressing each functional system. Affirmative answers to screening questions for a particular system prompted additional detailed questions, whereas negative answers allowed the participant to "screen out" for that system. Based on participant feedback and statistical analyses, both the logic of the program and wording of the questions were further adjusted. Implementation of screening questions had led to a systematic bias in calculating lower ePR-EDSS scores relative to the NS-EDSS in Cohort 1. Additional modifications included reducing the influence of self-reported hearing impairment on brainstem scores; using a "must use pads" rubric to capture more accurately bowel/bladder dysfunction; and a re-wording of the motor fatigability rubric to exclude generalized fatigue.

For the final validation (Cohort 2), 86 adults with MS completed the final version of the ePR-EDSS that removed the functional system screening page tested in Cohort 1. These subjects were block-enrolled to ensure that there were approximately 20 participants in each NS-EDSS category of: [0–1.5], [2–3.5], [4–5.5] and [6.0+]. Because some participants' NS-EDSS had changed between their prior clinical visit and the validation visit, enrollment was increased to 86 so that each of the 4 NS-EDSS categories had at least 20 subjects at the time of the validation visit.

The final ePR-EDSS version includes 23 questions, takes between 7–12 minutes to complete (based on time measured for Cohort 2 participants), and can be accessed at <https://openmsbioscreen.ucsf.edu/predss/about>. Participants were provided no instructions other than instructions included in the actual instrument; a study coordinator remained in the room solely to address technical issues (e.g. with the online form or tablet) should they arise.

Statistical Analysis

Analyses were completed to describe demographic and disease characteristics of Cohorts 1 and 2, to determine agreement between ePR-EDSS and NS-EDSS scores, and to assess sources of variability in the score differences. To quantitatively compare these two measures,

we used the kappa statistic and the Spearman's rank correlation coefficient.³ As sensitivity analyses, we adjusted for age, sex and education. Data were analyzed using R 3.6.0.

IRB Approval

This study was approved by the UCSF Institutional Review Board (IRB# 15–18362).

RESULTS

Participants

A total of 136 participants participated in this study (Testing Cohorts 1 and 2). Their mean age was 50.4 years (range 26–80), 69.9% of participants were women, and 91.2% were Caucasian-American. Mean disease duration was 15.9 years and median NS-EDSS was 2.8 (range 0–8). At the time of assessment, the majority of participants (66.2%) had a relapsing course; 28.7% were progressive and 5.1% were classified as having clinically isolated syndrome (n=2) or a yet- unclear disease course (n=5). In addition, 73.5% were on a disease modifying therapy for MS (12.5% self-injectable, 26.5% oral, and 34.6% by infusion), as shown in Table 1.

Agreement

We first sought to determine whether agreement between NS-EDSS and ePR-EDSS was sufficient to enable interchangeability during data collection and/or comparison between datasets, as previously articulated.³ Because of our development strategy, as expected, Cohort 2 outperformed Cohort 1 in terms of agreement between ePR-EDSS and NS-EDSS (Table 2). Therefore, as planned, Cohort 2 was used for comparisons between ePR-EDSS and NS-EDSS. The mean absolute difference between ePR-EDSS and NS-EDSS was 0.67, and there was an agreement within 1 point for 86.0% of all examinations. When we restricted the analysis to the participants with NS-EDSS below 6.0 (n=66), a sample more representative of MS trial participants, 83.3% of examinations were within 1 point of agreement. A Bland-Altman plot (Figure 1), which presents the differences between ePR-EDSS and NS-EDSS, revealed a mean difference of 0.12, and showed increasing agreement with increasing EDSS. The difference between the 95% limits of agreement was 3.63, which is greater than the cutoff for clinically significant change in NS-EDSS, as defined for most current randomized clinical trials of MS therapies³ (i.e., a 1-point change for an EDSS of 5.0 or less and a 0.5-point change otherwise).

The kappa statistic was 0.24 for complete agreement between the two scales ($p < 0.001$) and 0.85 for agreement within 1 point ($p < 0.001$). The intra-class correlation coefficient for the overall scores in Cohort 2 was 0.90 ($p < 0.001$). For individual functional systems, complete agreement was highest for the brainstem score (55.8%) and lowest for the sensory score (31.4%). In sensitivity analyses adjusted for NS-EDSS, the absolute difference between ePR-EDSS and NS-EDSS was not significantly related to age, sex, disease duration, years of education, or the timepoint at which the ePR-EDSS tool was assessed (before/after neurological exam). MS phenotype, however, did show a significant relationship with the magnitude of absolute disagreement between ePR-EDSS and NS-EDSS; a progressive

(vs. relapsing) MS phenotype was associated with greater absolute disagreement, in either direction ($\beta = 0.48$, $p = 0.012$) (Supplementary Figure 1).

Correlation

Finally, we determined whether the correlation between ePR-EDSS and EDSS was high enough to enable substitution of an ePR-EDSS for Neurostatus EDSS throughout data collection for a given study. The Spearman's rank correlation coefficient for the overall scores was 0.91 ($p < 0.001$, Figure 2), and for individual functional systems was highest for pyramidal, bowel and bladder, and cerebral scores (Table 3).

DISCUSSION

In this study, we have developed and validated an online, openly-available ePR-EDSS. This instrument iteratively incorporated participant perspectives to shorten the tool and to avoid suggesting to patients with early MS and low disability the possibility of certain unfavorable outcomes. These changes were then tested and the final ePR-EDSS was improved by the iterative amendments, with 86% of assessments having agreement between ePR-EDSS and NS-EDSS that was within 1-point.

The ePR-EDSS is unique compared to other published tools - it can be accessed and performed by the patient without any supervision, is freely and openly available, has built-in logic to calculate functional system and total scores, and is validated over a wide NS-EDSS range. Previously, both patient-reported and examiner-administered instruments have also demonstrated a similarly greater variability at lower levels of disability (i.e. $EDSS \leq 5.5$)^{3,14}. Although the test-retest reliability of the NS-EDSS is unknown for this low-disability sub-group, a recent study comparing different formats of examiner-administered NS-EDSS reported high inter-rater inconsistencies on many of the functional systems scores, which ultimately determine the NS-EDSS score.¹⁴ Clearly, also, the high degree of correlation between ePR-EDSS and NS-EDSS (0.91) in our cohort was driven, to a large extent, by data at the extremes of disability. Consequently, further validation of the ePR-EDSS for individuals within the disability range of 1.0–5.5, including both a consideration of the sensitivity to change over time and a determination of the comparative validity of the ePR-EDSS and the NS-EDSS, would be very useful. Moreover, the greater discrepancy noted in progressive patients (after adjusting for EDSS itself) suggests that physicians and patients differ in how they assess the “invisible” symptoms, which many MS patients experience. For example, a recently published “self-reported disability status scale”, based entirely on mobility, patients often over-estimated their disability compared to physician-assessment.¹⁵

Discrepancies such as these between any patient-reported scale and the NS-EDSS (including discrepancies in the estimated FS scores) are only to be expected. Indeed, all patient-reported outcomes, including the ePR-EDSS, are intended to (and likely do) capture overall function around the time that the scale is administered. Thus, for example, when queried, patients typically report how they have been functioning in general, and not how they are functioning at any particular moment in time. In this sense, the ePR-EDSS is more like the original EDSS (1982) in that its measures will integrate a patient's abilities over some period of time – even though both of these scales capture the patient's, rather than the

clinician's, evaluation of these abilities. Presumably, both the ePR-EDSS and the NS-EDSS would be able to capture changes over time although, which is the more reliable, remains to be determined. Also, patients will almost certainly report events differently than they are measured by the NS-EDSS. One important source of disagreement, of course, is the fact that clinicians may be under-recognizing and therefore under-scoring the severity of symptoms that they cannot observe directly and for which they may not use consistent, granular, codified questions to elicit response. This could explain discrepancies between the patients and clinicians, for example, in the bowel/bladder FS, or the cerebral FS when it includes fatigue. In fact, it is common that patients' reported scores are higher than clinicians'.¹⁴ There are other sources of heterogeneity related to the scales themselves. As an example, deficits in proprioceptive functions (e.g., joint position, joint movement, and vibratory senses) will, typically, not be perceived by patients as sensory loss. Rather, if these deficits are perceived at all, they will likely be reported as problems with balance, coordination, and/or ambulation. Therefore, any PR-EDSS intended to be validated against the NS-EDSS will need to incorporate algorithms to convert, logically, a patient's answers to specific questions into predicted functional systems and overall scores to be compared with the measured NS-EDSS. For this reason, both the initial PR-EDSS⁵ and the current ePR-EDSS included such scoring. To address the rare and possibly inevitable cases where there will be a significant discrepancy between the two scores (exemplified by one outlier in our study with a measured NS-EDSS of 4.5 but an ePR-EDSS score of 0.0), incorporating additional patient-generated data, such as step count measured using a commercial accelerometer,¹⁶ could provide additional objective information about the patient's true ambulatory activity. Consequently, it seems likely that no PR-EDSS instrument can be used interchangeably with the NS-EDSS.³ Moreover, which of these instruments provides the best (i.e., most valid) measure of "true" disability cannot be decided without an actual "gold standard". Moreover, in MS clinical trials, because predictive validity is more important than concurrent validity, to actually compare these two measures, as discussed above, will require longitudinal assessment.¹⁷⁻¹⁹ Further, its performance at the time of clinical change, such as a relapse, should be evaluated.

Limitations in the development and testing of ePR-EDSS include selection of a convenience cohort rather than a population-based cohort. Thus, although our study population included patients with all disability levels and, despite the inclusion of more diverse sets of patients in the development of the ePR-EDSS, the validation cohort we actually used was quite homogeneous with respect to its racial and ethnic composition. In addition, the predictive validity of the both ePR-EDSS and NS-EDSS will need to be established, longitudinally, by their correlation with long-term function using, as gold-standard, "hard" disability outcomes such as unremitting EDSS>6.0, unremitting need for a wheel chair, or death due MS-related morbidity.¹⁷⁻¹⁹

Nevertheless, there are several clear advantages to the ePR-EDSS. First, it is highly correlated with the NS-EDSS, which, currently, is the standard assessment tool. Therefore, although it is not interchangeable with the NS-EDSS, it could be used as a screening tool for clinical trial recruitment potentially helping to identify prospective participants who might qualify for any study, which is using the NS-EDSS as an entry criterion. Second, it could also be used as a convenient way to monitor a patient's function longitudinally

(perhaps monthly or every few months) and, thereby, to identify any changes that might occur. Third, administration of the ePR-EDSS does not require input from, or participation by, medically-trained personnel and is therefore less costly. Fourth, disability status can be assessed much more frequently during the course of any clinical trial and, perhaps, this will provide a more reliable measure of function over time. And, finally, using the ePR-EDSS may allow more MS patients, over a broader geographical range and having greater ethnic diversity, to participate in MS clinical research (including longitudinal studies) by reducing the onerous time-commitments that are currently required by these studies. There is a promising and expanding plethora of digital tools being deployed to evaluate and monitor patient function over time, offering a more naturalistic and continuous picture of patient function and activity.²⁰ However, until one agreed-upon composite snapshot emerges of overall patient functional status, the research community remains tethered to the EDSS and its associated measures (PR-EDSS³, tele-EDSS²¹).

Consequently, valid PR-EDSS measures offer considerable promise as a means to enhance clinical research. In addition, our findings have practical implications for the clinical setting, in which time-pressures often preclude a full NS-EDSS assessment. This would not only save time but also allow in-person encounters to focus on counseling and shared decision-making and improve, thereby, the quality of clinical care that our MS patients receive.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

REFERENCES

1. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*1983; 33: 1444–1452. [PubMed: 6685237]
2. Kurtzke JF. A new scale for evaluating disability in multiple sclerosis. *Neurology*1955; 5: 580–583. 1955/08/01. DOI: 10.1212/wnl.5.8.580. [PubMed: 13244774]
3. Collins CD, Ivry B, Bowen JD, et al. A comparative analysis of Patient-Reported Expanded Disability Status Scale tools. *Mult Scler*2016; 22: 1349–1358. 2015/11/14. DOI: 10.1177/1352458515616205. [PubMed: 26564998]
4. Bowen J, Gibbons L, Gianas A, et al. Self-administered Expanded Disability Status Scale with functional system scores correlates well with a physician-administered test. *Mult Scler*2001; 7: 201–206. 2001/07/28. DOI: 10.1177/135245850100700311. [PubMed: 11475445]
5. Goodin DS. A questionnaire to assess neurological impairment in multiple sclerosis. *Multiple sclerosis*1998; 4: 444–451. 1998/12/05. [PubMed: 9839306]
6. Cheng EM, Hays RD, Myers LW, et al. Factors related to agreement between self-reported and conventional Expanded Disability Status Scale (EDSS) scores. *Mult Scler*2001; 7: 405–410. 2002/01/25. DOI: 10.1177/135245850100700610. [PubMed: 11795463]
7. Lechner-Scott J, Kappos L, Hofman M, et al. Can the Expanded Disability Status Scale be assessed by telephone? *Mult Scler*2003; 9: 154–159. 2003/04/24. DOI: 10.1191/1352458503ms884oa. [PubMed: 12708811]
8. Leddy S, Hadavi S, McCarren A, et al. Validating a novel web-based method to capture disease progression outcomes in multiple sclerosis. *J Neurol*2013; 260: 2505–2510. 2013/06/29. DOI: 10.1007/s00415-013-7004-1. [PubMed: 23807152]
9. IDEO. *The Field Guide to Human-Centered Design: Design Kit*. IDEO (Firm), 2015, p.189.

10. Matheson GO, Pacione C, Shultz RK, et al. Leveraging human-centered design in chronic disease prevention. *American journal of preventive medicine* 2015; 48: 472–479. 2015/02/24. DOI: 10.1016/j.amepre.2014.10.014. [PubMed: 25700655]
11. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol* 2018; 17: 162–173. 2017/12/26. DOI: 10.1016/s1474-4422(17)30470-2. [PubMed: 29275977]
12. University of California, San Francisco MS-EPIC Team, Cree BA, Gourraud PA, et al. Long-term evolution of multiple sclerosis disability in the treatment era. *Ann Neurol* 2016; 80: 499–510. 2016/07/28. DOI: 10.1002/ana.24747. [PubMed: 27464262]
13. Schleimer E, Pearce J, Barneut A, et al. Applying human-centered design to develop a precision medicine tool for patients with multiple sclerosis: the Open MS BioScreen. *JMIR* 2020 4th, 2020. DOI: 10.2196/preprints.15605.
14. Goodkin DE, Cookfair D, Wende K, et al. Inter- and intrarater scoring agreement using grades 1.0 to 3.5 of the Kurtzke Expanded Disability Status Scale (EDSS). *Multiple Sclerosis Collaborative Research Group. Neurology* 1992; 42: 859–863. 1992/04/01. DOI: 10.1212/wnl.42.4.859.
15. Kaufmann M, Salmen A, Barin L, et al. Development and validation of the self-reported disability status scale (SRDSS) to estimate EDSS-categories. *Mult Scler Relat Disord* 2020; 42: 102148. DOI: 10.1016/j.msard.2020.102148. [PubMed: 32371376]
16. Block VJ, Bove R, Zhao C, et al. Association of Continuous Assessment of Step Count by Remote Monitoring With Disability Progression Among Adults With Multiple Sclerosis. *JAMA network open* 2019; 2: e190570. 2019/03/16. DOI: 10.1001/jamanetworkopen.2019.0570. [PubMed: 30874777]
17. Groves RM. *Survey errors and survey costs*. New York: Wiley, 1989, p.xxi, 590 p.
18. Goodin DS, Reder AT, Ebers GC, et al. Survival in MS: a randomized cohort study 21 years after the start of the pivotal IFNbeta-1b trial. *Neurology* 2012; 78: 1315–1322. 2012/04/13. DOI: 10.1212/WNL.0b013e3182535cf6. [PubMed: 22496198]
19. Goodin DS, Traboulsee A, Knappertz V, et al. Relationship between early clinical characteristics and long term disability outcomes: 16 year cohort study (follow-up) of the pivotal interferon beta-1b trial in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2012; 83: 282–287. 2011/12/24. DOI: 10.1136/jnnp-2011-301178. [PubMed: 22193561]
20. Block VJ and Bove R. We should monitor our patients with wearable technology instead of neurological examination – Yes. *Multiple Sclerosis Journal*; 0: 1352458520922762. DOI: 10.1177/1352458520922762.
21. Bove R BC, Crabtree E, Zhao C, Gomez R, Garcha P, Morrissey J, Dierkhising J, Green AJ, Hauser SL, Cree BAC, UCSF MS-EPIC Study, Wallin MT, Gelfand JM. Towards a low-cost, in-home, telemedicine-enabled assessment of disability in multiple sclerosis. *Multiple Sclerosis Journal* 2018.

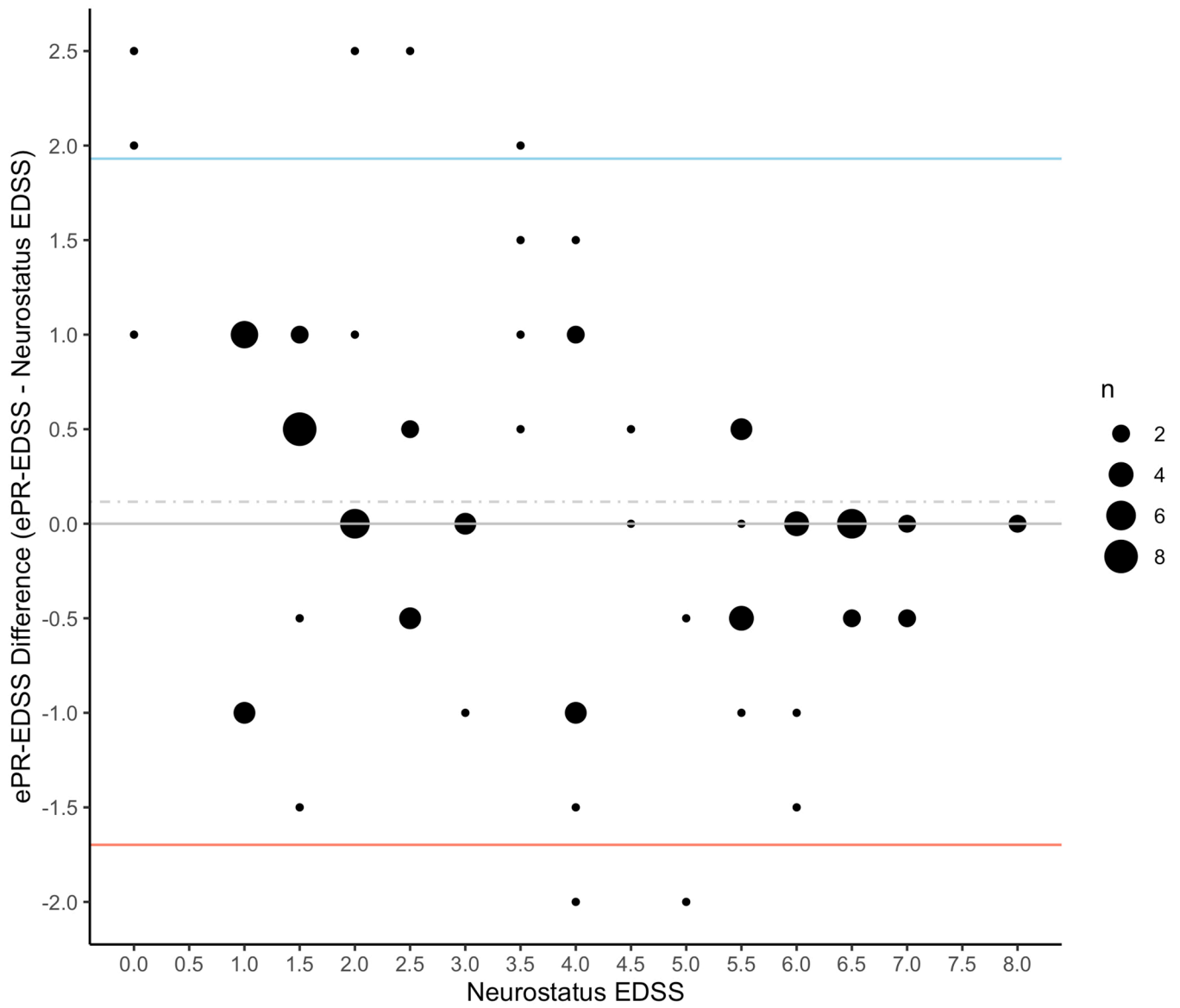


FIGURE 1. Bland-Altman Plot for the differences found between the ePR-EDSS and NS-EDSS measures.

The blue and red lines respectively indicate the upper- and lower-limits of agreement (-1.7, 1.9).

The Dot-dash is drawn at the mean of the difference between the ePR-EDSS and NS-EDSS measures (+0.12).

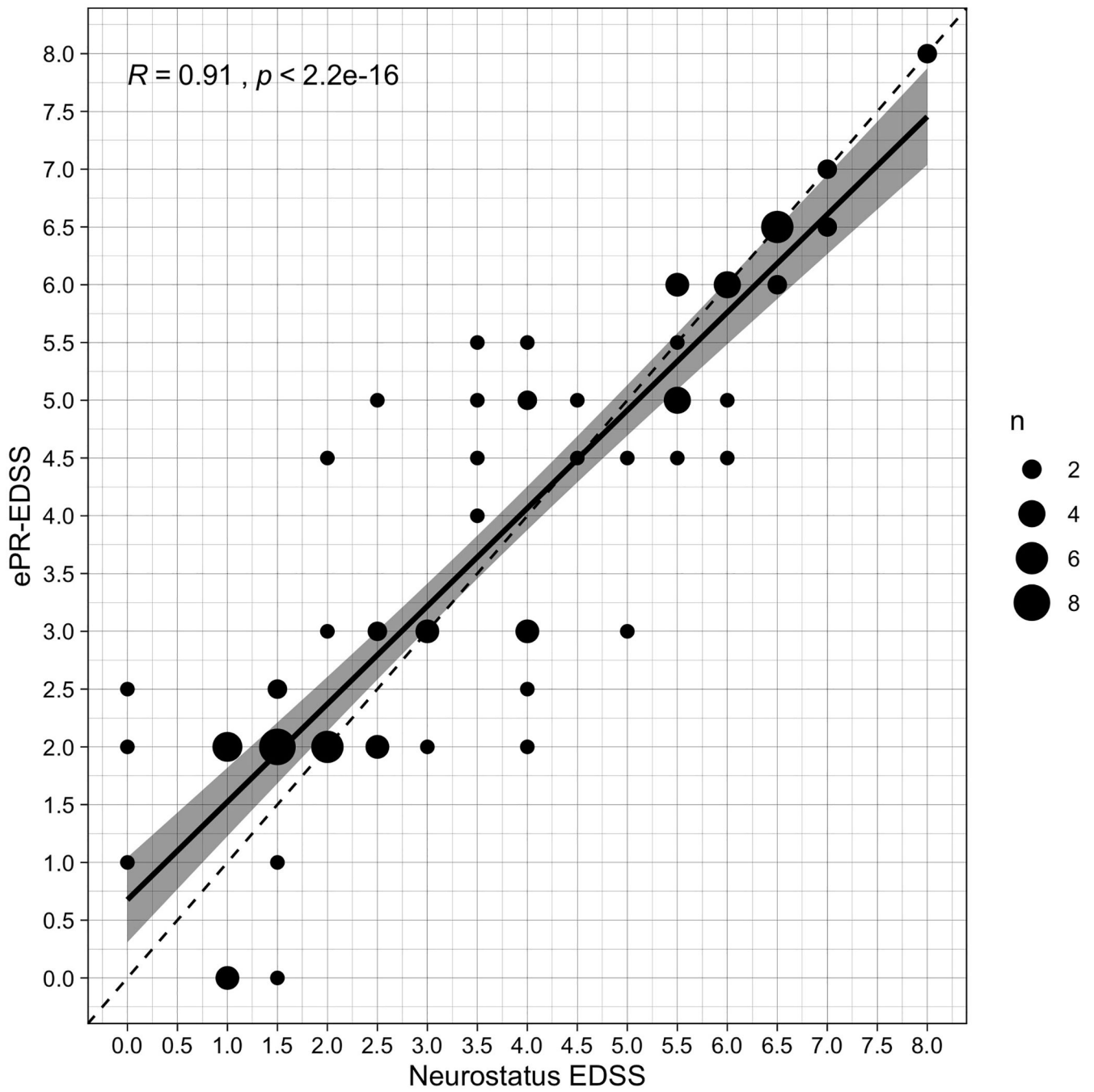


FIGURE 2. Correlation between ePR-EDSS and NS-EDSS in 86 adults with MS.
 The point size indicates the number of observations at each coordinate; the dashed line is drawn at perfect ePR-EDSS—NS-EDSS agreement.

TABLE 1.

Clinical and demographic characteristics of included study participants (n = 136)

	Cohort 1 (n = 50)	Cohort 2 (n = 86)	Combined (n = 136)	p-value
Age				0.76
Mean (SD)	49.9 (13.3)	50.6 (12.3)	50.4 (12.6)	
Median [IQR]	49.5 [40.0, 59.0]	51.2 [41.8, 58.1]	50.7 [41.0, 58.6]	
Sex				0.58
Female	33 (66.0%)	62 (72.1%)	95 (69.9%)	
Male	17 (34.0%)	24 (27.9%)	41 (30.1%)	
Race				0.82
Asian	1 (2.0%)	2 (2.3%)	3 (2.2%)	
Black	1 (2.0%)	3 (3.5%)	4 (2.9%)	
Other	1 (2.0%)	4 (4.7%)	5 (3.7%)	
White	47 (94.0%)	77 (89.5%)	124 (91.2%)	
Ethnicity				1.0
Hispanic or Latino	4 (8.0%)	6 (7.0%)	10 (7.4%)	
Not Hispanic or Latino	46 (92.0%)	80 (93.0%)	126 (92.6%)	
MS Type				0.12
CIS	2 (4.0%)	0 (0.0%)	2 (1.5%)	
PP	3 (6.0%)	16 (18.6%)	19 (14.0%)	
RR	35 (74.0%)	55 (64.0%)	90 (66.2%)	
SP	8 (16.0%)	12 (14.0%)	20 (14.7%)	
UNC	2 (4.0%)	3 (3.5%)	5 (3.7%)	
Disease Duration				0.97
Mean (SD)	15.9 (11.9)	15.9 (12.3)	15.9 (12.1%)	
Median [IQR]	15.0 [5.2, 22.2]	14.4 [5.2, 22.7]	15.0 [5.0, 22.9]	
DMT				0.42
Infused	13 (26.0%)	34 (39.5%)	47 (34.6%)	
Self-Injectable	7 (14.0%)	10 (11.6%)	17 (12.5%)	
None	14 (28.0%)	22 (25.6%)	36 (26.5%)	
Oral	16 (32.0%)	20 (23.3%)	36 (26.5%)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Cohort 1 (n = 50)	Cohort 2 (n = 86)	Combined (n = 136)	p-value
NS-EDSS				0.018
Mean (SD)	2.8 (1.7)	3.6 (2.2)	3.3 (2.0)	
Median [IQR]	2.5 [1.5, 3.5]	3.5 [1.5, 5.5]	2.8 [1.5, 5.5]	

TABLE 2.

Measures of ePR-EDSS — NS-EDSS agreement in Cohorts 1 and 2 (n = 50 and n = 86, respectively)

	Cohort 1					Cohort 2				
	All (n=50)	EDSS 0 – 1.5 (n=15)	EDSS 2 – 3.5 (n=24)	EDSS 4 – 5.5 (n=7)	EDSS 6.0+ (n=4)	All (n=86)	EDSS 0 – 1.5 (n=23)	EDSS 2 – 3.5 (n=22)	EDSS 4 – 5.5 (n=21)	EDSS 6.0+ (n=20)
<i>Bland Altman</i>										
Mean difference	-0.20	-0.067	-0.063	-0.93	-0.25	0.12	0.50	0.43	-0.31	-0.23
Paired t-test (p)	0.27	0.84	0.81	0.12	0.70	0.25	0.020	0.046	0.16	0.025
Upper 95% limit	2.3	2.5	2.4	1.7	2.1	1.9	2.4	2.3	1.6	0.58
Lower 95% limit	-2.7	-2.6	-2.5	-3.5	-2.6	-1.7	-1.4	-1.4	-2.2	-1.0
%N within 95% limits of agreement	96.0	93.3	100	100	100	91.9	91.3	90.9	100	95.0
<i>Mean Absolute Difference</i>	1.0	1.0	1.1	1.1	0.75	0.67	0.93	0.66	0.83	0.23
<i>Percentage agreement</i>										
Complete	14.0	13.3	8.3	28.6	25.0	29.1	0	40.9	9.5	70.0
Within 0.5	44.0	46.7	33.3	57.1	75.0	61.6	39.1	68.2	52.4	90.0
Within 1	64.0	60.0	62.5	71.4	75.0	86.0	87.0	81.8	81.0	95.0
Within 1.5	84.0	93.3	83.3	71.4	75.0	91.9	91.3	86.4	90.5	100
<i>Kappa for agreement</i>										
Complete agreement (p)	0.060 (0.15)	-0.083 (0.44)	-0.041 (0.52)	0.17 (0.20)	0.077 (0.51)	0.24 (<0.001)	-0.056 (0.18)	0.25 (0.005)	0.0099 (0.86)	0.58 (<0.001)
0.5 agreement (p)	0.38 (<0.001)	0.30 (0.014)	0.21 (0.0062)	0.46 (0.0056)	0.64 (0.015)	0.59 (<0.001)	0.20 (0.024)	0.59 (<0.001)	0.40 (<0.001)	0.86 (<0.001)
1.0 agreement (p)	0.60 (<0.001)	0.44 (0.0018)	0.52 (<0.001)	0.62 (<0.001)	0.64 (0.015)	0.85 (<0.001)	0.79 (<0.001)	0.77 (<0.001)	0.73 (<0.001)	0.93 (<0.001)
<i>Spearman's rank correlation coefficient</i> (p)	0.69 (<0.001)	-0.38 (0.16)	0.36 (0.080)	0.68 (0.093)	-0.82 (0.18)	0.91 (<0.001)	0.17 (0.45)	0.60 (0.0034)	0.59 (0.0052)	0.89 (<0.001)
<i>Intra-class coefficient</i> [conf. int.]	0.75 [0.59, 0.85] (<0.001)	-0.36 [-0.80, 0.213] (0.89)	0.32 [-0.098, 0.64] (0.065)	0.31 [-0.28, 0.81] (0.17)	-1.5 [-1.7, 0.074] (0.97)	0.90 [0.86, 0.94] (<0.001)	0.047 [-0.27, 0.40] (0.39)	0.44 [0.060, 0.72] (0.012)	0.51 [0.13, 0.77] (0.005)	0.80 [0.51, 0.92] (<0.001)

	Cohort 1					Cohort 2				
	All (n=50)	EDSS 0 – 1.5 (n=15)	EDSS 2 – 3.5 (n=24)	EDSS 4 – 5.5 (n=7)	EDSS 6.0 + (n=4)	All (n=86)	EDSS 0 – 1.5 (n=23)	EDSS 2 – 3.5 (n=22)	EDSS 4 – 5.5 (n=21)	EDSS 6.0 + (n=20)
(p)										

* In the Table EDSS refers to the NS-EDSS

TABLE 3.

Spearman’s rank correlation between ePR-EDSS — NS-EDSS functional system scores in Cohorts 1 and 2 (n = 50 and n = 86, respectively)

Functional System	Cohort 1					Cohort 2				
	All (n=50)	EDSS 0 – 1.5 (n=15)	EDSS 2 – 3.5 (n=24)	EDSS 4 – 5.5 (n=7)	EDSS 6.0 + (n=4)	All (n=86)	EDSS 0 – 1.5 (n=23)	EDSS 2 – 3.5 (n=22)	EDSS 4 – 5.5 (n=21)	EDSS 6.0 + (n=20)
Visual	0.17 (0.23)	0.38 (0.16)	0.10 (0.63)	0.11 (0.82)	NA	0.38 (<0.001)	-0.25 (0.24)	0.33 (0.15)	0.35 (0.12)	0.58 (0.0079)
Brainstem	0.46 (<0.001)	0.42 (0.12)	0.40 (0.050)	-0.35 (0.44)	1 (<0.001)	0.37 (<0.001)	0.40 (0.056)	0.14 (0.53)	0.31 (0.17)	0.52 (0.018)
Pyramidal	0.36(0.0093)	-0.47 (0.079)	0.26 (0.22)	0.23 (0.61)	0.82 (0.18)	0.67 (<0.001)	0.066 (0.77)	0.36 (0.099)	0.65 (0.0015)	0.21 (0.39)
Cerebellar	0.54 (<0.001)	-0.13 (0.63)	0.34 (0.10)	0.083 (0.86)	-0.27 (0.73)	0.53 (<0.001)	-0.083 (0.71)	-0.13 (0.55)	0.27 (0.24)	0.069 (0.79)
Sensory	0.28 (0.049)	-0.20 (0.47)	0.16 (0.45)	0.0095 (0.98)	NA	0.45 (<0.001)	-0.17 (0.44)	0.59 (0.0042)	0.011 (0.96)	0.29 (0.21)
Bowel & Bladder	0.79 (<0.001)	0.33 (0.24)	0.75 (<0.001)	0.81 (0.026)	1 (<0.001)	0.71 (<0.001)	0.39 (0.066)	0.78 (<0.001)	0.24 (0.29)	0.44 (0.052)
Cerebral	0.37 (0.0079)	-0.35 (0.21)	0.52 (0.0088)	0.16 (0.74)	0.27 (0.73)	0.55 (<0.001)	0.11 (0.60)	0.39 (0.069)	0.54 (0.012)	0.63 (0.0030)

* In the Table EDSS refers to the NS-EDSS