RESEARCH ARTICLE

NeurEx: digitalized neurological examination offers a novel high-resolution disability scale

Peter Kosa[1], Christopher Barbour[1,2], Alison Wichman[1], Mary Sandford[1], Mark Greenwood[2] & Bibiana Bielekova[1,]

[1]Neuroimmunological Diseases Section (NDS), National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, Bethesda, Maryland
[2]Department of Mathematical Sciences, Montana State University, Bozeman, Montana

Correspondence

Bibiana Bielekova, Neuroimmunological Diseases Section (NDS), National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Building 10, Room 5N248D, Bethesda, Maryland 20892. Tel: (240) 669-2724; Fax: (301) 496-5608; E-mail: bibi.bielekova@nih.gov

Abstract

Objective: To develop a sensitive neurological disability scale for broad utilization in clinical practice. **Methods**: We employed advances of mobile computing to develop an iPad-based App for convenient documentation of the neurological examination into a secure, cloud-linked database. We included features present in four traditional neuroimmunological disability scales and codified their automatic computation. By combining spatial distribution of the neurological deficit with quantitative or semiquantitative rating of its severity we developed a new summary score (called NeurEx; ranging from 0 to 1349 with minimal measurable change of 0.25) and compared its performance with clinician- and App-computed traditional clinical scales. **Results**: In the cross-sectional comparison of 906 neurological examinations, the variance between App-computed and clinician-scored disability scales was comparable to the variance between rating of the identical neurological examination by multiple sclerosis (MS)-trained clinicians. By eliminating rating ambiguity, App-computed scales achieved greater accuracy in measuring disability progression over time ($n = 191$ patients studied over 880.6 patient-years). The NeurEx score had no apparent ceiling effect and more than 200-fold higher sensitivity for detecting a *measurable* yearly disability progression (i.e., median progression slope of 8.13 relative to minimum detectable change of 0.25) than Expanded Disability Status Scale (EDSS) with a median yearly progression slope of 0.071 that is lower than the minimal measurable change on EDSS of 0.5. **Interpretation**: NeurEx can be used as a highly sensitive outcome measure in neuroimmunology. The App can be easily modified for use in other areas of neurology and it can bridge private practice practitioners to academic centers in multicenter research studies.

Introduction

Disability scales are used in drug development and, ideally, for data-driven therapeutic decisions in clinical practice. Multiple sclerosis (MS) disability scales are either insensitive (e.g., only about 10% of MS patients exhibit yearly progression on Expanded Disability Status Scale[1] [EDSS]) or too cumbersome to use in daily practice. Therefore, a broadly applicable and sensitive disability scale is needed.

Disability scales condense selected features of neurological examination and patient's history into a single number. Conceptually, disability scale development consists of two functions - selecting features that contribute to the scale and devising an algorithm that translates selected features into a number. Customarily, these functions were performed by a single[1,2] or several[3–5] domain experts. However, it is unclear if expert-defined scales are optimal. Additionally, inter-rater differences stemming from ambiguous algorithms lower scales' utility[6].

As an alternative to expert-based clinical scales, Combinatorial Weight-Adjusted Disability Scale (CombiWISE[6]) was developed using statistical learning. Although CombiWISE outperformed 57 clinical, imaging, and electrophysiological outcomes in a power analysis, its development was restricted to existing clinical scales and their cumbersome acquisition limits CombiWISE's broad clinical use.

Thus, we present the NeurEx App, which conveniently documents neurological examination and automatically computes several traditional scales, as well as a new highly sensitive disability scale.

Materials and Methods

Cohort characteristics

906 electronically documented neurological exams representing 262 MS patients and 908.6 patient-years (average of 3.5 visits per patient) were transcribed into the NeurEx App by board-certified neurologists/nurse practitioners with MS specialization. All patients (Table S1) participated in the protocol "Comprehensive Multimodal Analysis of Neuroimmunological Diseases of the Central Nervous System" (ClinicalTrials.gov Identifier: NCT00794352) and provided written informed consent. The protocol was approved by the Combined Neuroscience Institutional Review Board of the National Institutes of Health.

Out of 906 neurological exams, 787 were graded retrospectively (i.e., from previously documented structured neuro exam notes in NIH electronic medical records) and 119 were graded at the time of the neurological exam. 191 patients with a minimum of three visits over at least 1 year represent the longitudinal sub-cohort.

NeurEx development

The NeurEx App was developed on the FileMaker Pro platform. Although NeurEx can be used as a stand-alone module, we have combined it with a research database that integrates all aspects of clinical, regulatory, and research data.

To codify calculation of traditional disability scales, we mapped relevant NeurEx features into subsystems of four disability scales commonly utilized in neuroimmunology: Expanded Disability Status Scale (EDSS)[1], Scripps Neurological Rating Scale (SNRS)[3], Ambulation Index (AI)[2], and Instituto de Pesquisa Clinica Evandro Chagas (IPEC)[4] disability scale. For example, for EDSS the NeurEx App calculates Kurtzke's Functional Systems[1] first, then assigns the EDSS score following the rules narrated for each step of the EDSS scale. Although the published rules are ambiguous[1] (i.e., identical neurological examinations may be translated to more than one score), automatic computation of clinical scales through defined code is explicit (i.e., identical neurological examinations provide only one score for each scale). Therefore, we had to select exact rules for translating NeurEx features to existing clinical scales. These rules were optimized by an iterative process described in the Supplementary Material.

Statistical analyses

Concordance Correlation Coefficients (CCC[7]) and Bland-Altman plots[8] were used to assess the inter-rater reliability between two clinicians and App versus clinician ratings for each scale. Cronbach's alpha coefficient[9] was used to assess the reliability of NeurEx components in measuring underlying neurological dysfunctions. Since Cronbach's alpha assumes independent observations, the last time point for each patient was selected for the calculations. The CCC and Bland-Altman plots were generated in the open-source statistical software R[10], and Cronbach's alpha were generated using the *psych* R Package[11].

The correlation between NeurEx and EDSS or Combi-WISE was evaluated using linear regression models (GraphPad Prism 7), reporting a coefficient of determination (R^2; the proportion of variance explained by the linear model, where higher number signifies closer fit of experimental data points to the linear regression line of the model), slope, and it's 95% confidence interval (CI).

The test statistics for disability slopes (ratios of slopes to standard errors) were calculated on the slope coefficients from the simple linear regression models in the longitudinal cohort. The statistical significance of group differences in test statistics were evaluated using the paired-rank Friedman test with Dunn's multiple comparison test (GraphPad Prism 7).

Results

The NeurEx app characteristics

The NeurEx App divides the neurological examination into 17 panels (each panel displayed on one iPad page) addressing: 1. Cognitive functions, 2. Eyes, 3. Eye movements, 4. Visual fields, 5. Brainstem and upper cranial nerves, 6. Brainstem and lower cranial nerves, 7. Pyramidal signs and motor fatigue, 8. Upper extremity strength, 9. Lower extremity strength, 10. Reflexes, 11. Muscle atrophy, 12. Cerebellar functions, 13. Sensory exam (dermatome/stocking-glove distribution), 14. Sensory exam (pains and paresthesias), 15. Positive phenomena, 16. Stance and gait, and 17. Bowel, bladder, sexual functions (Fig. 1, Fig. S1, and Video S1).

The guiding principle for NeurEx development was to maximize the speed and convenience of the bedside documentation of a neurological examination, while assuring that the App contains information necessary for unequivocal computing of the four neuroimmunological disability scales: EDSS, SNRS, AI, and IPEC. To expand the dynamic range of NeurEx and avoid the ceiling effect that limits the use of current disability scales in patients with high disability, the NeurEx App captures broad types of
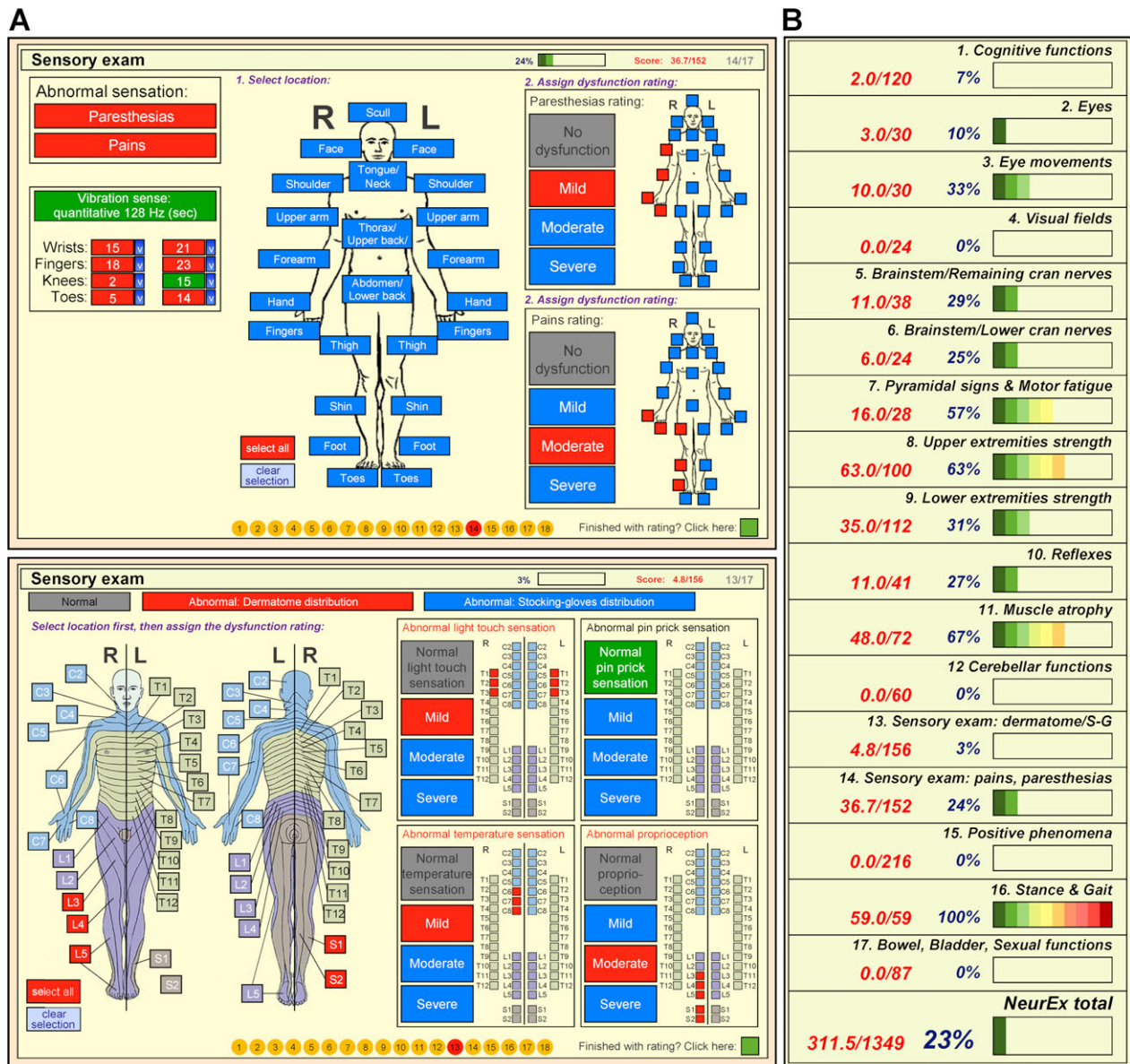
Figure 1. The NeurEx App. (A) A snapshot of the sensory exam part of the NeurEx App showing a human body diagram that allows recording of spatial information of deficiencies, as well as semiquantitative (mild, moderate, severe) information about recorded disability. (B) A snapshot of the summary of the NeurEx that depicts individual components of the neurological examination with graphical representation showing the amount of accumulated disability.

deficits, including positive phenomena (i.e., pains, paresthesias, extensor/flexor spasms, seizures), muscle atrophy, or autonomic dysfunctions. Consequently, the NeurEx contains 670 features with a theoretical maximum score of 1349.

All features are set as "normal" (highlighted in green color) by default, resulting in a NeurEx score of 0. By selecting features/grading other than "normal," abnormalities are highlighted in red and NeurEx points are

accumulated. The maximum NeurEx points for each feature were selected proportional to the number of features that constitute specific systems to assure relative equanimity between the systems (i.e., to avoid a possibility of a single system dominating the NeurEx score, in contrast to how walking ability dominates the EDSS). Consequently, the smallest NeurEx point is 0.25 (e.g., mild paresthesias in one body location) and the largest NeurEx point per single feature is 20 (e.g., coma, Fig. S1).

The combination of finger gestures (touch, swipe) and human body diagrams that display features in real spatial distribution provides fast (less than 10 minutes) and intuitive recording of a patient's examination (Video S1). NeurEx combines spatial information with a quantitative (e.g., vibration with 128 Hz tuning fork in specific appendages), or semiquantitative (e.g., mild, moderate, and severe) grading of selected deficits to derive final scores of each subsystem. To avoid ambiguity, NeurEx provides definitions of the scoring systems utilized for muscle strength (Fig. S1; panels 8 and 9) and reflexes (Fig. S1; panel 10).

The last page of the NeurEx App (Fig 1B and Fig. S1; panel 18) provides automatically real-time computed EDSS, SNRS, AI, and IPEC, including their subscores. This last page also contains scores for all 17 sections of the neurological examination, including the total NeurEx score. The NeurEx subsystem scores include heatmap-coded progress bars for a visual representation of the distribution of neurological deficits.

The NeurEx App-based and clinician-assigned disability scores are comparable

To evaluate the variance in clinical scales caused by rating ambiguity, MS-trained clinicians (i.e., Clinician #2) retrospectively assigned four selected disability scales to 787 neurological examinations previously documented by their colleagues (i.e., Clinician #1) in the structured neuroimmunology examination module of the NIH electronic medical records (Fig. 2A). Afterwards, Clinician #2 retrospectively transcribed the original neurological exam into the NeurEx App, which computed App-derived disability scores. Concordance correlation coefficients (CCC; representing the strength of the 1-1 relationship between two pairs of observations with values closer to one indicating stronger concordance) between two clinicians grading identical neurological examinations ranged from 0.943 for SNRS to 0.968 for AI. Thus, a small percent of variance is attributable to the ambiguity in rating procedure.

Next, we explored the CCCs between the NeurEx- and clinician-based disability ratings in a prospective manner, using 119 clinical visits. The CCCs were even higher compared to the retrospective cohort (i.e., ranging from 0.968 for SNRS and EDSS to 0.987 for AI; Fig. 2B).

Bland-Altman plots, which display the differences between measurements versus the average of measurements to examine if the magnitude of rater differences changes based on the magnitude of the measurements, also showed excellent concordance between clinician and App-based ratings (Fig. S2).

Cronbach's alpha was used to assess reliability of individual NeurEx components (e.g., subscores for panels 1–17) in measuring underlying neurological disability using independent observations of last visits for all 262 patients. The estimated Cronbach's alpha of 0.876 indicates a high internal consistency in NeurEx components.

NeurEx correlates highly with EDSS and CombiWISE in cross-sectional cohort

To assess clinical utility of NeurEx, we explored the relationship between NeurEx and EDSS (most utilized scale) and CombiWISE (most sensitive scale) in a cross-sectional cohort of MS patients (Fig. 2C). A linear regression model for EDSS with NeurEx shows R^2 of 0.682 with a slope of 0.023 (95% CI: 0.022–0.024, $P < 0.0001$). The same analysis performed on CombiWISE and NeurEx revealed R^2 of 0.813 with a slope of 0.214 (95% CI: 0.207–0.221; $P < 0.0001$).

Analysis of histograms for EDSS, CombiWISE, and NeurEx values showed more evenly distributed values for NeurEx, compared to the bimodally distributed[12] EDSS and, less prominently, CombiWISE scores (to which EDSS contributes approximately 30%). Visual examination of linear regression slopes also demonstrated a ceiling effect for EDSS and CombiWISE, as most disabled patients approached the theoretical maximums of these scales. In contrast, the maximum achieved NeurEx score in our cross-sectional cohort was 374.9, which represents only 27.8% of its theoretical maximum of 1349. This suggested that even for the most disabled patients, NeurEx has the potential to measure (linear) progression of disability. This was tested in the longitudinal subcohort of MS patients.

NeurEx outperforms EDSS and CombiWISE in measuring disease progression

One of the validated advantages of CombiWISE over EDSS is its increased sensitivity in detecting

Figure 2. The NeurEx App performs comparable to clinicians. Concordance correlation coefficients (CCC) calculated to assess the inter-rater reliability between ratings performed by two clinicians in a retrospective cohort (A) and App versus Clinician in the prospective cohort (B). The red lines represent the perfect 1-1 fit between two observations (CCC = 1). EDSS and AI datapoints were slightly jittered to avoid overplotting. (C) Linear regression models of NeurEx and EDSS (left) and NeurEx and CombiWISE (right) with coefficients of determination (R^2) displayed above the best fit line (in red). The histograms on the y-axes for EDSS (orange) and CombiWISE (purple) show a typical bimodal distribution of scores in the cohort. This bimodal distribution is noticeably reduced in the NeurEx scale (blue histograms) on the x-axes.
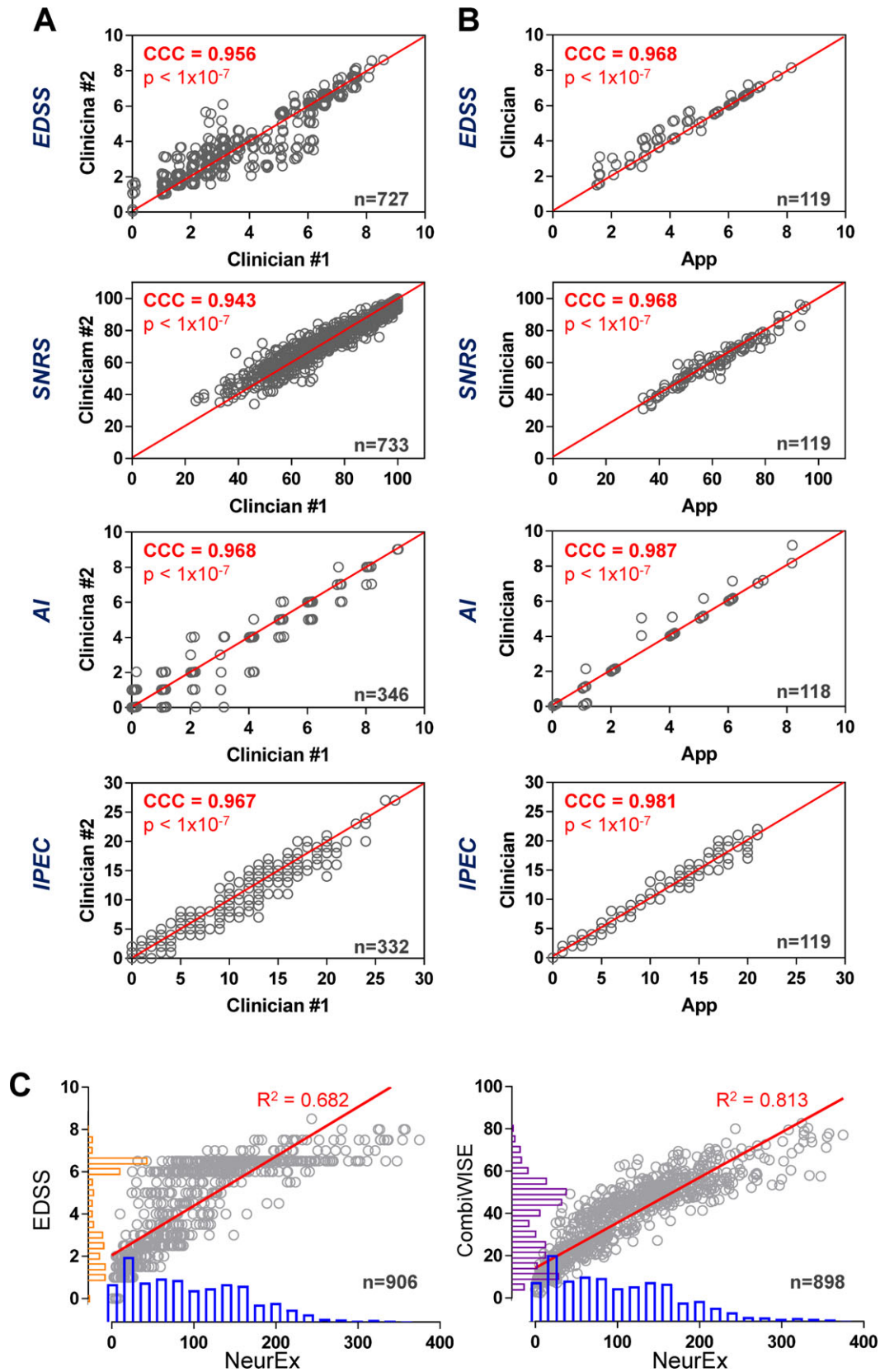
individualized disability progression[6] in 1–2 years. While linear regression slopes derived from cross-sectional data demonstrated the higher dynamic range of NeurEx in comparison to both EDSS and CombiWISE, only a longitudinal cohort can correctly estimate accuracy of individualized progression measurements.

Therefore, in a longitudinal cohort of 191 MS patients with at least three clinic visits (average 4.3, SD 1.4) spanning at least 1 year (average 4.6, SD 2.3) we evaluated the ability of EDSS, CombiWISE, and NeurEx to detect disability progression using linear regression models.

The progression slope of NeurEx relative to its smallest detectable change was more than 200-fold higher than the progression slope measured by EDSS relative to its smallest detectable change (i.e., median disability accumulation of 8.13/0.25 NeurEx points/year versus 0.07/0.5 EDSS points/year) (Fig. 3A). NeurEx also achieved the highest coefficient of determination (R^2), explaining 80% of variance in individualized linear regression models (median $R^2 = 0.799$), followed by CombiWISE (median $R^2 = 0.797$) and EDSS (median $R^2 = 0.450$) (Fig. 3B).

The longitudinal cohort also allowed comparison of NeurEx-derived, explicitly coded disability scales to more ambiguous ratings by different clinicians. For both EDSS and CombiWISE, we observed higher progression slopes (Fig. 3C) and higher R^2 (Fig. 3D) for NeurEx App-generated data. Specifically, EDSS slope increased from a median of 0.07 for clinician-based to 0.09 for App-based, while CombiWISE slopes increased from a median of 1.40 in clinician-based to 1.65 in App-based. The improvement in slopes was accompanied by increased variance explained in both EDSS (median R^2 for clinician-based 0.445 and App-based 0.513) and CombiWISE (median R^2 for clinician-based 0.797 and App-based 0.840).

Representative examples of two patients followed for over 8 years visually depict the advantages of the NeurEx-based codification of disability scales (Fig. 3E). While patient #1 doesn't show disability progression on EDSS (slope of 0), both CombiWISE and NeurEx detect accumulation of disability (slope of 0.76 and 10, respectively). Patient #2 illustrates the increase of the coefficient of determination of the linear regression if the disability is measured by EDSS ($R^2 = 0.64$), CombiWISE ($R^2 = 0.86$), and NeurEx ($R^2 = 0.93$).

To compare the power of the scales to detect changes in disease progression within individual patients, *t*-statistics on the slope coefficient from a simple linear regression were quantified for each scale. Statistical analysis showed a significant increase in the sensitivity of NeurEx compared to EDSS ($P < 0.0001$, Fig. 3F) with the median *t*-statistic 1.8 units larger for NeurEx than EDSS.

NeurEx App also allows detailed analysis of the progression slopes between neurological subsystems (Fig. 3G). Subsystems that dominate traditional MS disability scales such as strength, stance, and gait exhibited high progression rates. However, progression in subsystems that do not contribute to EDSS, SNRS, or AI but are part of IPEC (such as pains, paresthesias, and quantitative vibration sense) achieved comparatively high progression. Specifically, pains and paresthesias showed the highest progression of all components (median slope of 1.70 points per year), followed by stance and gait (1.06 points per year), lower extremity strength (0.93 points per year), and cerebellar functions (0.59 points per year).

Discussion

EDSS[1], developed in 1983 by Dr. John Kurtzke, remains the most utilized scale in MS field. Its drawback is the discrete nature with only 19 possible "levels" of progression, leading to a median annualized EDSS change measured in large natural history cohorts of 0.1.[13] This rate of progression is *unmeasurable* by EDSS in intervals shorter than 5-10 years, as the minimal measurable EDSS step is 0.5. In clinical trials, only approximately 10% of MS patients will have *measurable* yearly EDSS progression, making EDSS insensitive for most reasonably sized research applications.

SNRS[3], developed concurrently with EDSS by a group of domain experts from the Scripps Institute has five times higher dynamic range with 100 possible levels of disability. Nevertheless, SNRS never achieved broad use, likely due to the cumbersome nature of its computation.

Charged with developing a sensitive, objective scale, an expert panel assembled Multiple Sclerosis Functional Composite (MSFC)[5] in 1999. MSFC unites three functional tests (timed 25-foot walk [25FW], nine-hole peg test [9HPT], and Paced Auditory Serial Addition Test [PASAT]) to a number that relates patient's performance to a reference population. MSFC is a continuous scale that takes considerable investment of resources to perform; therefore, it is rarely administered outside of clinical trials. Most importantly, MSFC sensitivity to measure yearly disability progression in MS cohorts does not meaningfully outperform EDSS[6,14].

Consequently, most recent MS trials adopted composite outcomes, consisting of sustained progression of disability on one of several scales, such as EDSS, 25FT, and 9HPT.[15,16] While such outcomes enhance sensitivity of group comparisons, they do not measure *severity* of MS progression on a patient level.

CombiWISE[6], a continuous, statistical learning-optimized scale that combines SNRS, EDSS, 25FW, and non-dominant hand 9HPT, has excellent sensitivity;
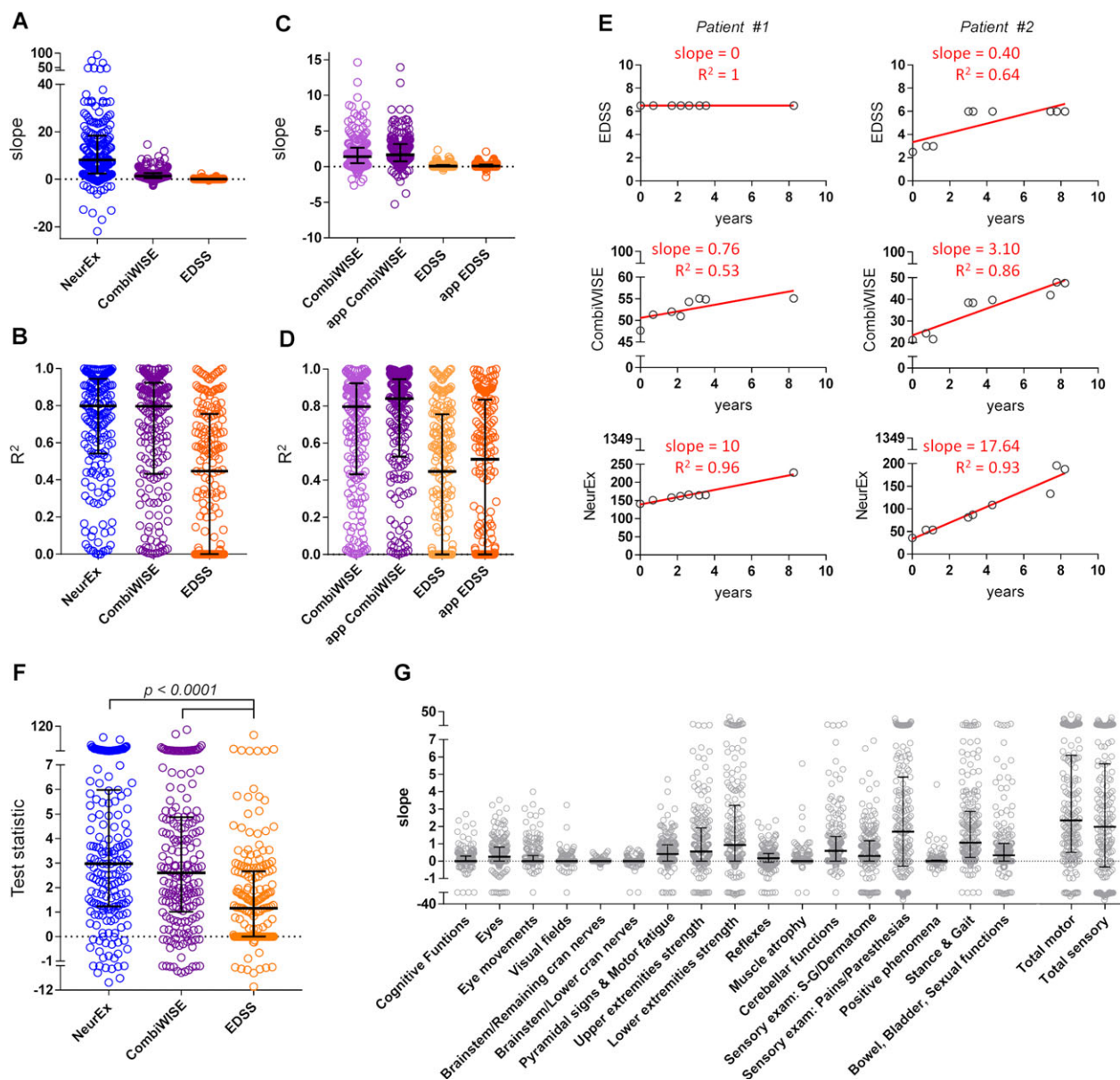
Figure 3. NeuroEx outperforms CombiWISE and EDSS in detection of disability progression. (A) Slopes of diseases progression and (B) coefficients of determination (R^2) measured by linear regression of values for NeurEx (blue), CombiWISE (purple), and EDSS (orange) in a cohort of 185 MS patients with at least three visits over at least 1 year show the highest median for NeurEx, followed by CombiWISE and EDSS. (C,D) Comparison of progression slopes and R^2 values for linear regression models on clinician-based scores (CombiWISE – light purple, EDSS – light orange) and app-based scores (app CombiWISE – purple, app EDSS – orange) showed improvement in both the slopes (C) and the R^2 values (D) in App-based scores compared to clinician-based scores. (E) Examples of two MS patients followed over 8 years and their disability progression measured by EDSS, CombiWISE, NeurEx shows increased sensitivity (slope) and specificity (R^2) of NeurEx compared to both EDSS and CombiWISE. The red lines represent linear regression model. (F) *t*-statistics on the slope coefficient from a simple linear regression (SLR) for NeurEx (blue) CombiWISE (purple), and EDSS (orange) to assess the sensitivity to detect changes in disease progression within patients. Friedman test with Dunn's adjustment for multiple comparisons revealed statistically significant difference between NeurEx and EDSS scale and between CombiWISE and EDSS scale. (G) Slopes of disability progression in the 17 components of the NeurEx exam plus combined total score for motor and sensory dysfunction show different levels of disability progression in different parts of the neurological exam. Thick black line – median, black whiskers – 1st and 3rd quartiles.

CombiWISE outperformed 57 clinical, electrophysiological and imaging outcomes[6] in detecting yearly MS disability progression. Nevertheless, CombiWISE does not eliminate the problem of cumbersome acquisition, even though a public website allows its automatic computation: https://bielekovalab.shinyapps.io/msdss/. This

impediment limits broad adaptation of CombiWISE and prompted us to develop the NeurEx App.

An ideal scale combines high accuracy (power) with practical utility. To achieve high accuracy, the scale must have high dynamic range (i.e., high sensitivity) and low error rate (i.e., high specificity). To attain a broad utilization, scale should be obtained with a reasonable investment of time or resources. Rather than a priori selecting specific features of the neurological examination, we opted for highly convenient documentation of the neurological examination in its entirety. To optimize specificity, NeurEx uses explicitly defined muscle strength and reflex ratings and exact time(s) the patient feels vibrations of defined frequency in prespecified appendages. In the development and validation of the Combinatorial MRI scale of CNS tissue destruction (i.e., COMRIS-CTD[17]) we have confirmed the hypothesis that noise stemming from ambiguity of semiquantitative ratings (such as "mild, moderate, and severe") may be successfully abated by combining semiquantitative features with spatial information in the scale assembly. We used the same principle in the NeurEx App, where semiquantitative ratings of sensory modalities combine with spatial information to enhance accuracy. Using touch screen and human body diagrams on portable platforms such as iPads, allows clinician to document even the most complex neurological examination in <10 min at bedside. NeurEx App also contains embedded quality assurance skills, such as automatic prompts for the clinician to complete severity of grading if she/he has selected an abnormality in a specific subsystem.

Comparing clinician-derived scores of four traditional scales with the NeurEx App-derived scores demonstrated strong, highly significant concordance, and comparable variances between App-based and two different clinicians-based scores.

Please, note that the measured variance between two clinicians represents only variance in *translating* documented neurological examination to a number, since Clinician #2 scored structured neurological examinations documented by the Clinician #1 in the electronic medical records, rather than reexamining the patient. This variance is effectively eliminated by the NeurEx App, which always assigns identical number to an identical examination. This represents important advantage in circumstances where several raters may examine a patient, e.g., within a clinical trial or shared private practice. The structured flow of the App as well as legends defining rules for grading different features (e.g., grading of deficits in muscle strength and reflexes) help to document neuroexam correctly. However, the NeurEx App is just a tool dependent on the input of data from the clinician; it cannot compensate for errors performed during examination or for the examination that is incomplete. It is possible to compute disability scores for less detailed examinations and these scores will be consistent for individual rater(s) if she/he omits identical parts of neurological examination each time. However, such omission of examination features will result in inter-rater variability if the second rater performed full examination and identified previously unmeasured disability.

By eliminating noise between clinicians' ratings, NeurEx-based scores provided more reliable disability progression slopes in longitudinal data, as evidenced by test statistics and by NeurEx explaining the highest proportion of variance in individual longitudinal exams (i.e., 80%) among all rating scales. However, the most notable feature of NeurEx scale is its dynamic range with the maximum of 1349 points and the smallest measurable change of 0.25 points (0.02% of the maximum) providing a superior resolution to the 10-point EDSS scale with the smallest measurable change of 0.5 points (5% of the maximum). The longitudinal sub-cohort showed over a 100-fold higher median of disease progression slope measured by NeurEx compared to EDSS and confirmed the inability of EDSS to detect yearly disease progression in individual patients (i.e., 61 out of 191 patients [32%] followed for a median of 3.1 years [range 1–8.3 years] patients showed no disease progression on EDSS). The test statistics fully support this superiority of NeurEx to traditional scales. Finally, the NeurEx scale has no apparent ceiling effect, with most disabled patient in our cohort reaching only 27.8% from its theoretical maximum. This allows measuring disability progression even in the most disabled MS subjects, facilitating their participation in research studies. Finally, for research applications, NeurEx allows more meaningful correlations in neuroimaging or functional studies (e.g., correlating correct side – right or left limb with left or right side of the brain).

In conclusion, technological advances in mobile computing are providing an unprecedented opportunity for development of new types of outcomes. Their sensitivity and specificity can revolutionize drug testing, facilitate translational research, and empower data-driven approach to patient care. Modifying and/or expanding the NeurEx App to serve other neurology fields, such as neurodegenerative and movement diseases or stroke, consists of the straightforward task of assuring that the NeurEx App captures all features necessary for computation of traditional disability scales and codifying the algorithm(s) for their numerical translation. Partnership between medical experts, patients, and the commercial sector is crucial for further technical development, support, and broad distribution of these healthcare-empowering tools.

Acknowledgments

Author Contribution

Peter Kosa: Acquisition, analysis and interpretation of data, drafting of the manuscript. Christopher Barbour: analysis and interpretation of data, drafting of manuscript. Alison Wichman: acquisition of data, critical review of the manuscript. Mary Sandford: acquisition of data, critical review of the manuscript. Mark Greenwood: analysis and interpretation of data, drafting of manuscript. Bibiana Bielekova: Study design and concept, acquisition, analysis and interpretation of data, drafting of the manuscript.

Conflict of Interest

P. Kosa, C. Barbour, A. Wichman, M. Sandford, M. Greenwood and B. Bielekova have no disclosures relevant to the manuscript.

References

1. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983;33:1444–1452.
2. Hauser SL, Dawson DM, Lehrich JR, et al. Intensive immunosuppression in progressive multiple sclerosis. A randomized, three-arm study of high-dose intravenous cyclophosphamide, plasma exchange, and ACTH. N Engl J Med 1983;308:173–180.
3. Sipe JC, Knobler RL, Braheny SL, et al. A neurologic rating scale (NRS) for use in multiple sclerosis. Neurology 1984;34:1368–1372.
4. Schmidt F, Oliveira AL, Araujo A. Development and validation of a neurological disability scale for patients with HTLV-1 associated myelopathy/tropical spastic paraparesis (HAM/TSP): the IPEC-1 scale. Neurology 2012;78:P03–P258.
5. Cutter GR, Baier ML, Rudick RA, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. Brain 1999;122(Pt 5):871–882.
6. Kosa P, Ghazali D, Tanigawa M, et al. Development of a sensitive outcome for economical drug screening for progressive multiple sclerosis treatment. Front Neurol 2016;7:131.
7. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255–268.
8. Bland JM, Altman DG. Comparing two methods of clinical measurement: a personal history. Int J Epidemiol 1995;24 (Suppl 1):S7–S14.
9. Cronbach LJ, Warrington WG. Time-limit tests: estimating their reliability and degree of speeding. Psychometrika 1951;16:167–188.
10. R: A language and environment for statistical computing. R Foundation for Statistical Computing [computer program] 2014.
11. psych: Procedures for Psychological, Psychometric, and Personality Research [computer program] 2017.
12. Hohol MJ, Orav EJ, Weiner HL. Disease steps in multiple sclerosis: a simple approach to evaluate disease progression. Neurology 1995;45:251–255.
13. Lorscheider J, Buzzard K, Jokubaitis V, et al. Defining secondary progressive multiple sclerosis. Brain 2016;139:2395–2405.
14. Kragt JJ, Thompson AJ, Montalban X, et al. Responsiveness and predictive value of EDSS and MSFC in primary progressive MS. Neurology 2008;70:1084–1091.
15. Montalban X, Hauser SL, Kappos L, et al. Ocrelizumab versus placebo in primary progressive multiple sclerosis. N Engl J Med 2016;376:209–220.
16. Hauser SL, Bar-Or A, Comi G, et al. Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. N Engl J Med 2016;376:221–234.
17. Kosa P, Komori M, Waters R, et al. Novel composite MRI scale correlates highly with disability in multiple sclerosis patients. Multiple Sclerosis Relat Disord 2015;4:526–535.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Fine-tuning of the rules for translating NeurEx into the four clinical scales.
Figure S1. 18 panels of the NeurEx App specifying the number of features, maximum score and % of total NeurEx score for each panel.
Figure S2. Bland-Altman plots.
Table S1. Demographic data.
Vidoe S1. An example of NeurEx App grading of a hypothetical-MS patient.