medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Analytical validation of a multi-protein, serum-based assay for disease activity assessments in multiple sclerosis

Ferhan Qureshi,[1]* Wayne Hu,[1]* Louisa Loh,[1] Hemali Patel,[1] Maria DeGuzman,[1] Michael Becich,[1†] Fatima Rubio da Costa,[1†] Victor Gehman,[1†] Fujun Zhang,[1†] John Foley,[2] Tanuja Chitnis[3]

[1]Octave Bioscience, Inc., Menlo Park, CA, USA

[2]Rocky Mountain Multiple Sclerosis Clinic, Salt Lake City, UT, USA

[3]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

*These authors contributed equally to this work

[†]Employee of Octave Bioscience, Inc., at the time the study was completed

Correspondence

Ferhan Qureshi

1440 Obrien Drive, Suite B

Menlo Park, CA 94025, USA

Email: fqureshi@octavebio.com

Phone: 650-224-3072

Fax: 650-479-8876

Abbreviations:

%CV, percent coefficient of variation; APLP1, amyloid beta precursor-like protein 1; CCL20, C-C motif chemokine ligand 20; CD6, cluster of differentiation 6; CDCP1, CUB domain-containing protein 1; C_{max}, maximum concentration; CNS, central nervous system, CNTN2, contactin 2; COL4A1, collagen type IV alpha-1; conc, concentration; CSF, cerebrospinal fluid; CXCL9, chemokine (C-X-C motif) ligand 9 (MIG); CXCL13, chemokine (C-X-C motif) ligand 13; DMT, disease-modifying therapy; DNA, deoxyribonucleic acid; FLRT2, fibronectin leucine-rich repeat transmembrane protein; Gd+, gadolinium positive; GFAP, glial fibrillary acidic protein; GH, growth hormone; HAMA, human anti-mouse antibodies; HCl, hydrochloride; IL-12β, interleukin-12 subunit beta; LLOQ, lower limit of quantitation; LOQ, limit of quantitation; mAb, monoclonal antibody; Max, maximum; Min, minimum; MOG,

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

myelin oligodendrocyte glycoprotein; MRI, magnetic resonance imaging; MS, multiple sclerosis; MSDA, Multiple Sclerosis Disease Activity; Na, sodium; NfL, neurofilament light chain; OPG, osteoprotegerin; OPN, osteopontin; PCR, polymerase chain reaction; PEA, Proximity Extension Assay; PRTG, protogenin; qPCR, quantitative polymerase chain reaction; R^2, coefficient of determination; RA, rheumatoid arthritis; RF, rheumatoid factor; SD, standard deviation; SERPINA9, serpin family A member 9; TNFRSF10A, tumor necrosis factor receptor superfamily member 10A (TRAIL-R1); TNFSF13B, tumor necrosis factor superfamily member 13B (BAFF); ULOQ, upper limit of quantitation; VCAN, versican core protein

Key words:

Analytical characterization, analytical validation, biomarker, multiple sclerosis, proximity extension assay

Total word count: 3964 (5000 maximum)

Running head: Analytical Validation of a Serum-Based Assay for Disease Activity Assessments in Multiple Sclerosis

Previous presentation: Part of this work, namely analytical validation of the individual biomarkers in the MSDA Test, was previously presented at the ACTRIMS 2021 Forum, Virtual (February 25□27, 2021) and analytical validation of the Disease Activity Score and 4 Disease Pathway Scores, was previously presented at the ACTRIMS 2022 Forum, West Palm Beach, Florida (February 24□26, 2022)

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Clinical relevance

Multiple sclerosis (MS) is a chronic, neurodegenerative, immune-mediated disease of the CNS. MS has a complex disease course with variable clinical outcomes. Although many treatments are effective in early stages of the relapsing/remitting form of the disease, early diagnosis and treatment are critical to managing disease activity and slowing disease progression. One of the major areas of focus in MS research is the identification of biomarkers in biological fluids, such as cerebrospinal fluid or blood, to track pathogenesis, disease activity, and disease progression, which can lead to individualized disease management and improved quality of care. Currently, there are no validated clinical tests that leverage multiple blood biomarkers to track disease activity or progression in patients with MS. Herein, we describe the analytical characterization and validation of a multi-protein, serum-based assay panel developed using Olink® PEA methodology. We demonstrate the extensive characterization of this multi-protein, serum-based assay and establish its accuracy, precision, sensitivity, and robustness. This report will be followed by a complementary clinical validation study investigating the correlation between the proteomic assay results and relevant clinical and radiographic endpoints for patients with MS.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Purpose: To characterize and analytically validate the MSDA Test, a multi-protein, serum-based biomarker assay developed using Olink® PEA methodology.

Experimental design: Two lots of the MSDA Test panel were manufactured and subjected to a comprehensive analytical characterization and validation protocol to detect biomarkers present in the serum of patients with MS. Biomarker concentrations were incorporated into a final algorithm used for calculating four Disease Pathway scores (Immunomodulation, Neuroinflammation, Myelin Biology, and Neuroaxonal Integrity) and an overall Disease Activity score.

Results: Analytical characterization demonstrated that the multi-protein panel satisfied the criteria necessary for a fit-for-purpose validation considering the assay's intended clinical use. This panel met acceptability criteria for 18 biomarkers included in the final algorithm out of 21 biomarkers evaluated. VCAN was omitted based on factors outside of analytical validation; COL4A1 and GH were excluded based on imprecision and diurnal variability, respectively. Performance of the four Disease Pathway and overall Disease Activity scores met the established acceptability criteria.

Conclusions and clinical relevance: Analytical validation of this multi-protein, serum-based assay is the first step in establishing its potential utility as a quantitative, minimally invasive, and scalable biomarker panel to enhance the standard of care for patients with MS.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

What is known and what is new in your work?

What's known

- Multiple sclerosis (MS) has a complex disease course with variable clinical outcomes; early diagnosis and treatment are critical to management of MS.
- One key focus in MS research is the identification of biomarkers in biological fluids, such as cerebrospinal fluid or blood, to track pathogenesis, disease activity, and disease progression, which may lead to individualized disease management and improved quality of care.
- There currently are no validated clinical tests that leverage multiple blood biomarkers to track disease activity or progression in patients with MS.

What's new

- The MS Disease Activity (MSDA) Test is a multi-protein, serum-based biomarker assay designed to quantitatively measure disease activity using the protein levels of biomarkers present in the serum of patients with MS.
- In this study, we evaluated 21 biomarkers, 18 of which were selected for inclusion in the MSDA Test, and extensively characterized the MSDA Test (individual biomarkers and algorithmic scores) by establishing the accuracy, precision, sensitivity, and robustness of the assay.
- This study serves as a critical first step in the validation of this multi-protein, serum-based assay, which will be a quantitative, minimally invasive, and scalable tool to improve MS disease management.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

1 INTRODUCTION

Multiple sclerosis (MS) is a chronic, neurodegenerative, immune-mediated disease of the CNS, characterized by inflammatory demyelination and neuronal damage.[1,2] MS has a complex disease course with variable symptoms or manifestations that can range from mild and self-limiting to severe.[1] The clinical course, after the first clinical manifestation of the disease, or clinically isolated syndrome, can vary.[3] The damage caused by MS typically leads to relapses, or acute attack of symptoms, followed by progressive disease.[4] Most treatments are effective in early stages of the relapsing/remitting form of the disease;[4,5] however, a delay in treatment can lead to irreversible damage.[6] Studies show that the extent of remyelination in early MS is greater than in chronic MS.[7] Clinical studies are underway to explore treatments targeting remyelination, which may slow or offset disease progression.[8]

The McDonald Criteria, designed to improve the accuracy of MS diagnosis, established the use of MRI to show the accrual of lesions over time and space.[9] The revised McDonald Criteria substituted CSF oligoclonal immunoglobulin G bands for the second clinical/MRI finding.[10] Nonetheless, use of any of these assessments do not always accurately predict disease activity, course, progression, recurrence, or response to treatment.[11-13] As such, there is an unmet clinical need for objective and quantitative measures that can accurately diagnose MS, monitor disease activity, and promote individualized disease management.[13,14]

One major area of focus in MS is the identification of biomarkers in biological fluids, such as CSF or blood, to track pathogenesis, disease activity, and progression.[14,15] One of the key therapeutic strategies in MS is to reduce relapse, lesions, and brain atrophy at all disease stages.[4] As a result, new biomarkers for early MS diagnosis and disease activity monitoring can lead to prevention of disease

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

progression, potentially reducing the patient's level of disease worsening.[14] The dynamic range of proteins in CSF presents challenges when differentiating small disease-specific changes from inherent inter-individual differences, especially as it relates to methodological variations.[16,17] CSF collection also requires invasive procedures, such as lumbar puncture. On the other hand, blood-based collection of biomarkers allows for safe, quick, and easy collection.[14] With these considerations, detection of biomarkers in blood is a viable and attractive option for the accurate diagnosis and assessment of disease activity and progression in MS. However, there currently are no validated clinical tests that leverage multiple blood biomarkers to track disease activity or progression in patients with MS.[18]

Development of multi-protein assays can be challenging. Each protein biomarker requires specific conditions and methodologies for optimal quantification. The optimal multi-protein assay should be designed so that stability and integrity of all biomarker proteins are maintained and optimized to eliminate cross-reactivity.[19] Larger scale, proteomic techniques allow higher throughput of samples and more timely readout. However, maintaining robustness, repeatability, and sensitivity is challenging, yet critical, to the validation of a multi-protein biomarker panel.[20]

Analysis of multiple proteins may more accurately represent the various pathways, processes, and cell types involved in complex disease states and has the potential to deliver more personalized medicine for MS.[20-23] Single proteins may not perform well alone as diagnostic or prognostic markers. However, as part of a multi-protein assay, they may contribute to a clinically useful model when combined with other proteins and biomarkers.[21] Therefore, multi-protein assay platforms have been characterized and validated for complex disease states.[19,21,22,24]

The MSDA Test is a multi-protein, serum-based biomarker assay designed to quantitatively measure disease activity using the protein levels of biomarkers present

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

in the serum of patients with MS. Our custom multi-protein assay panel was developed using the Olink® PEA (Olink Proteomics, Uppsala, Sweden) methodology described previously (**Figure S1**).[19] Herein, we describe the comprehensive analytical characterization and validation of the MSDA Test to satisfy the criteria necessary for a fit-for-purpose validation considering the assay's intended clinical use.

2 EXPERIMENTAL SECTION

2.1 Assay development

Twenty-one biomarkers were selected for inclusion in the custom assay panel based on statistical associations with clinical and radiographic endpoints as demonstrated in feasibility studies for which >1400 proteins were screened using 2 immunoassay platforms **(Table S1)**. These feasibility studies investigated biomarker associations (single-protein and multi-protein) in both cross-sectional and longitudinal samples relative to several radiographic and clinical MS endpoints, including clinically defined relapse versus remission (exacerbation versus quiescence), the presence and count of gadolinium-enhanced lesions on a matched MRI, annualized relapse rate, and Expanded Disability Status Scale. From these studies, the custom panel of 21 proteins was selected with a primary focus on the detection and prediction of disease activity status. The 21 proteins were chosen based on their statistical significance relative to the aforementioned endpoints and with the intent to comprehensively survey the biological pathways, mechanisms, and cell types associated with MS pathophysiology as determined via literature review, protein-protein interaction modeling, gene set enrichment, and spatial expression profiling.[25] Dynamic range of the individual protein assays was considered, as well as the intent to develop a

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

single multi-protein immunoassay panel for which each protein could be measured in an undiluted serum sample. The MSDA Test algorithm consisting of 18 biomarkers included in the panel was finalized in a subsequent clinical validation study for which independent sample sets were analyzed. The final model was trained and validated relative to the presence and count of gadolinium-enhanced lesions.

Serum pools (*n*=4) were included on all runs during assay discovery and development. They were procured in large volumes, aliquoted, stored at −65°C, and run in triplicate. Serum pools were used solely to assess the analytical performance of the assays and served as process controls to determine acceptability of future analytical runs. The SD of repeated measurements was applied to the expected concentrations. Two assay kit lots of the panel were manufactured for which critical reagents were varied to the extent possible.

2.2 Description of the two-layer stacked classifier algorithm for determination of the overall Disease Activity score

A two-layer, L2-penalized logistic regression stacked classifier model was developed and clinically validated in a separate study that optimized the model's performance to classify serum samples based on the presence of gadolinium-enhancing lesions (0 lesions or ≥1 lesions) on an MRI administered within 60 days of blood draw.[26] In the first layer of the model, individual protein concentrations in $\log_{10}$ which were demographically corrected for age and sex and LOQ-imputed (referred to as adjusted concentrations) were used as inputs into the four Disease Pathway models (Immunomodulation, Neuroinflammation, Myelin Biology, and Neuroaxonal Integrity). The second layer of the model used the adjusted protein concentrations and the

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

output (eg, the probability) of the Disease Pathway models as meta features to calculate an overall Disease Activity score (**File S1, Supporting Information**). Thresholds were established, which corresponded to low (1.0□4.0), moderate (4.5□7.0), and high (7.5□10.0) Disease Activity scores. Analytical characterization and validation of the individual biomarkers were factors used to determine inclusion of those biomarkers in the algorithm.

2.3 Incurred sample reanalysis

Incurred sample reanalysis was performed to characterize precision and robustness for the individual biomarkers and the Disease Activity and Disease Pathway scores. Forty-eight individual samples from patients with MS were repeatedly analyzed across 10 plates over ≥5 days with varied equipment, reagents, location, and personnel. Acceptability criteria for individual biomarkers was an average %CV ≤20%, and average SD at all established Disease Activity score levels of ≤1.0 units. The 48 samples broadly represented the expected range of biomarker values and Disease Activity scores in the real-world MS population.

2.4 Assay accuracy, precision, and sensitivity

Accuracy for each analyte was determined by mixing serum samples at different ratios and evaluating the percent recovery of the observed concentration relative to the expected concentration. Sample mixing enabled the accuracy assessment to be performed using endogenous protein versus a recombinant protein source. Expected concentrations were calculated by applying the targeted ratios of unmixed samples.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The ratios of sample mixtures with two samples were 25%:75%, 50%:50%, and 75%:25%. The ratios of sample blends for mixtures with four samples were 25%:25%:25%:25% and 40%:10%:40%:10%. Additionally, accuracy was also evaluated for the Disease Pathway and Disease Activity algorithms by correlating observed scores with expected scores using the same sample mixtures created for the individual analyte assessments.

Intra- and inter-assay precision was measured for each analyte. The %CV was determined using serum pools enabling the assessment to be performed using endogenous protein. Serum pools were manufactured to represent patients with shorter and longer MS disease duration, those with inflammatory disease (rheumatoid arthritis), and one healthy control. Acceptability criteria for intra- and inter-assay precision was established as %CV ≤15% and ≤20%, respectively.

Sensitivity was defined as the assay's ability to accurately and precisely detect low concentrations of a given substance in biological specimens. To establish the ULOQ and LLOQ, a LOQ panel was manufactured during assay development. For each analyte, four levels were targeted near the anticipated upper limit (ULOQ 1–4) and four levels were targeted near the anticipated lower limit (LLOQ 5–8). The targeted concentrations were based on expected real-world MS patient sample distributions, the shape of the standard curve, and location of asymptotes. The LOQ

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

panel was run in triplicate over two lots (≥5 runs per lot) and fit to the standard curve. Accuracy, defined as 80–120% recovery relative to the expected concentration and precision (inter-assay %CV ≤20%), were used to establish the acceptability criteria and determine the LLOQ and ULOQ of each analyte. Additionally, individual LOQs were assessed and established separately for each kit lot. The most conservative LOQ levels with acceptable accuracy and precision parameters for both lots were used to establish the final LLOQ and ULOQ.

Undiluted serum samples were run in the MSDA Test and as a result, no dilution factor was accounted for in the sensitivity analysis. Therefore, the LLOQ and ULOQ define both the analytical measurement range and the reportable range of the assay. Serum samples that recovered either above the ULOQ or below the LLOQ were reported at the established LOQ concentration (referred to as LOQ imputation). MS serum samples were used to establish MS reference ranges for each biomarker. A diverse set of patient samples were used throughout the assay development process and for the analytical validation studies. A total of 1645 samples from nine deeply phenotyped cohorts were analyzed primarily for evaluating associations of biomarkers with MS disease activity and disease progression endpoints. Additional samples from both patients with MS and other disease states were procured for specific analytical characterization experiments. The 1645 samples that were analyzed for the associations of biomarkers with MS endpoints were combined in the subsequent analysis to establish MS reference ranges. These samples were collected both retrospectively and prospectively from nine US and international sites

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and broadly represent the real-world MS population. The mean ± SD age of these patients at the time of the blood draw was 40.85 ± 11.0 years, with a mean ± SD disease duration of 8.39 ± 8.0 years; 72.8% of the patients were female. For race, the top 3 categories were White (81.4%), unknown/not reported (13.5%), and Black/African American (2.7%). The primary endpoint used to train the finalized MSDA Test algorithm was the presence and count of gadolinium-enhancing lesions on an MRI administered within close proximity to the blood draw. For the 1645 patient samples, 1326 had available gadolinium-positive (Gd+) lesion counts and 53.0% of the patient samples had ≥1 Gd+ lesion. The linear interpolation method was used to establish the 95% interval (2.5th and 97.5th percentiles).[27] The percentile relative to these reference ranges are presented with their protein concentrations.

2.5 Assay interference

Assay interference was defined as the effect of a substance present in the sample altering the correct value of the result or the recovery of samples in the assay. Since patients with MS may be treated with a variety of drugs, potential interference of drugs was tested to determine if their presence would affect measurement of the individual protein biomarkers. Concentrations of common prescriptions, over-the-counter drugs, common MS drugs, and DMTs were spiked into serum samples (**Table S2**). Concentrations of common prescription and over-the-counter drugs were determined by Sun Diagnostics (New Gloucester, ME, USA) using a commercially available test kit. DMTs were targeted at two times C_{max} from pharmacokinetic studies, or the highest possible concentration allowable for spiking with the procured interferent stock. Finally, a universal mAb standard was tested at two concentrations (424 and 7.93 µg/mL) to cover the two times C_{max} of several mAb DMTs. Endogenous substances (hemoglobin, bilirubin, and lipids) and heterophilic antibodies (RF and HAMA) were also measured. For most interferent substances, the

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

acceptability threshold, or median recovery, for the interference assessment was established as 80□120% relative to a corresponding spike control, except for HAMA for which percent recovery of sample mixtures was evaluated (**File S2, Supporting Information**).

2.6 Diurnal variability

Patient serum samples were collected at days 1□5 and day 12 to characterize biomarker level fluctuations. For each of the six time points per patient, the %CV and the percentage difference of the observed protein concentration relative to the average concentration at all time points were calculated.

2.7 Sample stability

In an initial experiment, stability studies for four serum samples were performed to determine the effect that storage and processing conditions can have in a clinical setting. Stability was assessed at the following four temperatures: −65°C or below (−80°C), −10°C or below (−20°C), 2–8°C (4°C), and room temperature (18–25°C) at the following time points: 4 hours (for 4°C and room temperature) and days 1, 3, 7, 14, and 28 (for −20°C, 4°C, and room temperature). The results from −20°C, 4°C, and room temperature were compared with the control storage condition (−80°C). In a follow-up study, the stability of storage of 14 samples was evaluated at 4°C at days 1–3 and 7 compared with a control storage condition (−80°C) to establish the duration of time samples that can be stored at 4°C. Five freeze-thaw cycles,

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

performed at −65°C or below, were also evaluated using four MS serum samples compared with fresh samples.

3 RESULTS AND DISCUSSION

3.1 Analytical characterization and validation

Experiments were performed between July 2020 and July 2021. Fifty-one plates were run (40 and 11 plates using the first and second lots of manufactured kits, respectively).

Based on the analytical validation and characterization of individual biomarkers described below, the 18 out of 21 biomarkers that were included in the algorithm were determined to have acceptable analytical performance. GH and COL4A1 were excluded from the algorithm based on the analytical characterization studies described below. VCAN was not incorporated into the final algorithm due to biostatistical factors unrelated to analytical performance.

3.2 Incurred sample reanalysis

All individual biomarkers were determined to have a mean %CV <20% and met established acceptability criteria (**Figure 1A**). The Disease Activity score and the four Disease Pathway scores demonstrated reproducible results throughout the range of scores (**Figures 1B–1F)**. For the Disease Activity score, the average SD across 48 samples was observed to be 0.3 score units, which is less than one interval (0.5) on the reportable scale, and as a result, met acceptability criteria. Additionally, incurred sample reanalysis showed robustness and equivalency of the assay between lots and laboratories, with the exception of COL4A1 (**Table S3**).

3.3 Assay accuracy, precision, and sensitivity

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Samples for the accuracy assessment were selected from an internal MS cohort (*n*=64) to target both high and low concentrations for the individual biomarkers relative to the MS population. Twenty mixed samples from four selected samples were analyzed for each biomarker. Minimum percent recovery for each biomarker ranged from 78% to 89%; the maximum percent recovery for each biomarker ranged from 99% to 124%. The median percent recovery ranged from 91% to 100% (**Figure 2A**). Additionally, the Disease Pathway and overall Disease Activity scores were calculated for both observed and expected concentrations of the various sample mixtures. The observed calculated scores correlated with the expected scores; R^2 ≥0.85 was established as the acceptability criteria (**Figures 2B□2F**).

Twelve replicates per serum pool were analyzed on a single plate for the intra-assay precision assessment; ≤51 values per serum pool were analyzed across 51 plates spanning 2 lots of reagent kits. The intra- and inter-assay precision satisfied the criteria for meeting the precision parameter with most analytes passing the established criteria. Of note, COL4A1 was found to have inferior inter- and intra-assay precision that ranged from 7% to 47% and 15% to 59%, respectively. Based on these findings, COL4A1 was removed from consideration for inclusion in the algorithm. MS serum samples (*N*=1645) were analyzed during the assay development and validation process and used to establish the MS reference ranges for each analyte. Sensitivity analysis demonstrated that the LLOQ and ULOQ of each analyte met the sensitivity requirements established for the assay. The maximum

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

percentage of samples requiring imputation at any LOQ was 1.8% (for NfL at LLOQ) (**Table 1**).

3.3 Assay interference

Most biomarker interactions with interferent combinations, such as common MS drugs, DMTs, and mAbs produced a median recovery that ranged from 80% to 120% (**Figure 3**). A lower percentage recovery was observed for two biomarkers, COL4A1 and CCL20, demonstrating a potential alteration in the presence of the sample for individual drugs. COL4A1 produced a low percent recovery for several drugs that ranged from 71% to 79%, which was likely an artifact of established assay imprecision (**Figure 3** and **Figure S2**). For CCL20, cefoxitin spiked at 660 mg/dL resulted in a median percent recovery of 77% (**Figure S2**). Additional assay interferents are shown in **Figures S2** (common drugs) and **S3** (routine endogenous interferents and heterophilic antibodies).

3.4 Diurnal variability

Diurnal variation was evaluated in eight patients over six time points (**Figure S4**). Mean and median percent differences for each biomarker and patient were observed to be within ±20%; mean and median %CV was found to be <30% for 19 of the 21 biomarkers (**Table 2**). Of note, there were some individual samples that were outside of the acceptable range (±30%; data not shown). In addition, mean and median diurnal variability ≥30% was observed for COL4A1, which may have been due to the imprecision of the assay to detect this biomarker. GH was also found to be more variable compared with the other biomarkers, which is not surprising, as GH has

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

been previously reported to have a high degree of ultradian and diurnal variability.[28] For this reason, GH was removed from consideration for inclusion in the algorithm.

3.5 Sample stability

In the initial stability study, all biomarkers were stable for up to 1 day at room temperature and at 4°C, and for 28 days at −20°C. For those samples stored at room temperature, CXCL13, IL-12β, and TNFSF13B decreased beyond −20% at 3 days. During a follow-up study, all biomarkers were found to meet acceptability criteria when stored at 4°C, and consistent with the initial study as well as the control condition (−80°C) at follow-up (**Table S4**). In a study to examine the stability of samples after freeze-thaw, most biomarkers met acceptability criteria when compared with fresh sample. Of note, GFAP concentrations decreased beyond −20% for freeze-thaw cycles 4 and 5 (**Table S5**). Finally, score level analysis showed that test conditions were within 3 SDs (±1.5 score difference) from the control conditions during the initial study (**Table S6**) and at follow-up (**Table S7**). From these findings, we showed that biomarker levels were found to be most affected above certain thresholds (room temperature for 24 hours, 4°C for 7 days, −20°C for 28 days, and three freeze-thaws). These data can be used to establish allowable sample handling and storage conditions. Beyond these empirical estimations of protein stability, it is

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

also important to note that statistically meaningful associations of biomarkers with multiple MS endpoints were observed using samples that had been stored at −80°C for extended periods of time. This suggests that the target epitopes for the proteins that were selected for inclusion in the custom assay panel and the final algorithm were sufficiently stable to derive clinically meaningful insights.

4 CONCLUSION

The accuracy, precision, and reproducibility of a biomarker assay are critical to its utility as a diagnostic and prognostic tool in the management of complex neurodegenerative disorders such as MS. Additionally, such an assay should be insensitive to external factors such as assay interferents and sample collection, processing, and storage. The Clinical and Laboratory Standards Institute and the United States Food and Drug Administration issued guidance on the development and validation of assays for the detection of serum-based biomarkers.[27,29,30] Parameters such as accuracy, precision, recovery, sensitivity and specificity, quality control, and sample stability need to be optimized for the assay to be properly validated.[27,29,30] Results from our analytical validation experiments to characterize the MSDA Test support that the assay is accurate, precise, sensitive, specific, and robust at determining individual biomarker levels and algorithmic scores, regardless of assay interferents, and validated in terms of sample stability. Our findings of high accuracy and precision for the MSDA Test assay align with those of other validation studies of multi-protein assays utilizing the same,[31,32] as well as alternative[21,22] platforms.

PEA demonstrated high sensitivity, specificity, reproducibility, and repeatability with low intra- and inter-assay variability, which has allowed for large-scale, high

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

throughput screening of up to 92 proteins in 96 samples simultaneously, with low sample consumption and cost.[19] This platform detected novel protein biomarkers and biomarker combinations for many complex disease states, such as cardiovascular disease,[33-37] cancer,[32,38-40] Alzheimer's disease,[41] and inflammatory diseases such as atopic dermatitis and lupus[42,43]; the platform has also proven useful in aging research.[44] For the MSDA Test, we demonstrated that a focused panel of MS biomarkers can be developed and optimized on the PEA platform with absolute quantitation of the proteins to support a fit-for-purpose analytical validation, thereby enabling clinical use of the assay.

Thus far, there are no validated clinical tests that leverage multiple blood biomarkers to track disease activity or progression in patients with MS. This is critical for a disease such as MS, which has a complicated clinical course varying from mild, self-limiting to severe.[1] Although MS disease prognosis is primarily based on clinical evidence, such as relapse rate and disability progression, and diagnostic tests (eg, brain MRI or the presence of oligoclonal immunoglobulin G bands in the CSF),[14] neither can consistently and accurately predict disease course, activity, or prognosis.[13] Given the emphasis on early diagnosis and the efficacy of therapies to treat early stages of the relapsing/remitting form of the disease,[4,5] validation of a biomarker panel remains an unmet need in clinical practice, and use of this biomarker tool should provide diagnostic and prognostic value for the treatment of MS. This study demonstrated identification of biomarkers for this complex disease using the PEA platform. With further clinical validation, this assay can potentially be used to track disease activity and progression of MS, allowing a more personalized approach to MS treatment.

A limitation of using a multi-protein assay is that the conditions established for one biomarker are not always uniform across the full panel of biomarkers. Our

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

findings show that the MSDA Test was optimized for assessment of 18 out of the 21 included biomarkers and the analytical validation paradigm that we described demonstrates a high level of accuracy, sensitivity, and precision with minimal cross-reactivity and interference by substances commonly seen in patients with MS.

This study serves as a critical first step in the validation of a multi-protein, serum-based assay. The next step in the validation of the MSDA Test is clinical validation, which will support and confirm the association between the serum-based MSDA Test and clinical and radiographic MS endpoints. Upon completion of clinical validation of the assay, the final Disease Activity and Disease Pathway algorithms will use the ensemble of validated proteins to expand the use of the assay by evaluating biomarker correlations with endpoints associated with additional MS disease assessments, selection of therapy, and differential diagnosis of patients with MS. Upon successful clinical validation, this MSDA Test will be a quantitative, minimally invasive, and scalable tool to improve disease management for patients with MS and their physicians.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ACKNOWLEDGMENTS

The authors wish to thank the following team members from Olink Proteomics (Uppsala, Sweden) who were involved in the development of the Multiple Sclerosis Disease Activity Test assay: Erika Assarsson, Sandra Ohlsson, Martin Lundberg, Jessica Bergman, and Niklas Nordberg. All authors contributed to and approved the manuscript for submission. Writing and editorial assistance were provided by Jennifer L. Venzie, PhD, and Bu Reinen, PhD, CMPP, of The Lockwood Group (Stamford, CT, USA), and were funded by Octave Bioscience, Inc.

CONFLICT OF INTEREST STATEMENT

F. Qureshi, W. Hu, L. Loh, H. Patel and M. DeGuzman are employees of Octave Bioscience, Inc. M. Becich, F. Rubio da Costa, V. Gehman, and F. Zhang were employees of Octave Bioscience, Inc. at the time the study was completed. J. Foley has received research support from Biogen, Novartis, Adamas, Octave Bioscience, Inc., Genentech, and Mallinckrodt, received speakers' honoraria and acted as a consultant for EMD Serono, Genzyme, Novartis, Biogen, and Genentech, and has equity interest in Octave Bioscience Inc., and is the founder of InterPro Biosciences. T. Chitnis has received compensation for consulting from Biogen, Novartis Pharmaceuticals, Roche Genentech, and Sanofi Genzyme, and received research support from the National Institutes of Health, National MS Society, US Department of Defense, EMD Serono, I-Mab Biopharma, Mallinckrodt ARD, Novartis Pharmaceuticals, Octave Bioscience, Inc., Roche Genentech, and Tiziana Life Sciences.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FUNDING

The study was funded by Octave Bioscience, Inc. and in part by the U.S. Department of Defense (W81XWH2110633 to T Chitnis).

DATA AVAILABILITY STATEMENT

Access to data can be provided after a research proposal is submitted to the corresponding author and a data sharing agreement is in place.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

REFERENCES

[1] Compston, A., Coles, A. (2008). Multiple sclerosis. *Lancet*, *372*, 1502-1517.

[2] Weiner, H. L. (2004). Multiple sclerosis is an inflammatory T-cell–mediated autoimmune disease. *Arch Neurol*, *61*, 1613-1615.

[3] Miller, D. H., Chard, D. T., Ciccarelli, O. (2012). Clinically isolated syndromes. *Lancet Neurol*, *11*, 157-169.

[4] Giovannoni, G., Butzkueven, H., Dhib-jalbut, S., Hobart, J., Kobelt, G., Pepper, G., . . . Vollmer, T. (2016). Brain health: time matters in multiple sclerosis. *Mult Scler Relat Disord*, *9*, S5-S48.

[5] Macaron, G., Ontaneda, D. (2019). Diagnosis and management of progressive multiple sclerosis. *Biomedicines*, *7*, 56.

[6] Miller, J. R. (2004). The importance of early diagnosis of multiple sclerosis. *J Manag Care Pharm*, *10*, S4-S11.

[7] Goldschmidt, T., Antel, J., König, F. B., Brück, W., Kuhlmann, T. (2009). Remyelination capacity of the MS brain decreases with disease chronicity. *Neurology*, *72*, 1914-1921.

[8] Cunniffe, N., Coles, A. (2021). Promoting remyelination in multiple sclerosis. *J Neurol*, *268*, 30-44.

[9] McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F. D., . . . Wolinsky, J. S. (2001). Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann Neurol*, *50*, 121-127.

[10] Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., . . . Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*, *17*, 162-173.

[11] Tintoré, M., Rovira, A., Río, J., Nos, C., Grivé, E., Téllez, N., . . . Montalban, X. (2006). Baseline MRI predicts future attacks and disability in clinically isolated syndromes. *Neurology*, *67*, 968-972.

[12] Teixeira, M., Seabra, M., Carvalho, L., Sequeira, L., Abreu, P., Mendonça, T., . . . Guimarães, J. (2020). Clinically isolated syndrome, oligoclonal bands and multiple sclerosis. *Clin Exp Neuroimmunol*, *11*, 33-39.

[13] Jafari, A., Babajani, A., Rezaei-Tavirani, M. (2021). Multiple sclerosis biomarker discoveries by proteomics and metabolomics approaches. *Biomark Insights*, *16*, 11772719211013352.

[14] Ziemssen, T., Akgün, K., Brück, W. (2019). Molecular biomarkers in multiple sclerosis. *J Neuroinflammation*, *16*, 272.

[15] Gul, M., Jafari, A. A., Shah, M., Mirmoeeni, S., Haider, S. U., Moinuddin, S., Chaudhry, A. (2020). Molecular biomarkers in multiple sclerosis and its related disorders: a critical review. *Int J Mol Sci*, *21*, 6020.

[16] Hühmer, A. F., Biringer, R. G., Amato, H., Fonteh, A. N., Harrington, M. G. (2006). Protein analysis in human cerebrospinal fluid: physiological aspects, current progress and future challenges. *Dis Markers*, *22*, 3-26.

[17] Zhang, J. (2007). Proteomics of human cerebrospinal fluid – the good, the bad, and the ugly. *Proteomics Clin Appl*, *1*, 805-819.

[18] Lublin, F. D., Reingold, S. C., Cohen, J. A., Cutter, G. R., Sørensen, P. S., Thompson, A. J., . . . Polman, C. H. (2014). Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*, *83*, 278-286.

[19] Assarsson, E., Lundberg, M., Holmquist, G., Björkesten, J., Bucht Thorsen, S., Ekman, D., . . . Fredriksson, S. (2014). Homogenous 96-plex PEA

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*, *9*, e95192.

[20] Schubert, O. T., Röst, H. L., Collins, B. C., Rosenberger, G., Aebersold, R. (2017). Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc*, *12*, 1289-1294.

[21] Dillon, R., Croner, L. J., Bucci, J., Kairs, S. N., You, J., Beasley, S., . . . Wilcox, B. (2018). Analytical validation of a novel multiplex test for detection of advanced adenoma and colorectal cancer in symptomatic patients. *J Pharm Biomed Anal*, *154*, 85-94.

[22] Eastman, P. S., Manning, W. C., Qureshi, F., Haney, D., Cavet, G., Alexander, C., Hesterberg, L. K. (2012). Characterization of a multiplex, 12-biomarker test for rheumatoid arthritis. *J Pharm Biomed Anal*, *70*, 415-424.

[23] Chitnis, T., Prat, A. (2020). A roadmap to precision medicine for multiple sclerosis. *Multiple Sclerosis Journal*, *26*, 522-532.

[24] Bhawal, R., Oberg, A. L., Zhang, S., Kohli, M. (2020). Challenges and opportunities in clinical applications of blood-based proteomics in cancer. *Cancers (Basel)*, *12*, 2428.

[25] Chitnis, T., Becich, M., Bove, R., Cree, B. A. C., Gehman, V., Gomez, R., . . . Baranzini, S. E. Development of a custom multivariate proteomic serum based assay for association with radiographic and clinical endpoints in MS. Presented at: Americas Committee for Treatment and Research in Multiple Sclerosis and European Committee for Treatment and Research in Multiple Sclerosis; September 11-13, 2020; Virtual.

[26] Chitnis, T., Foley, J., Ionete, C., El Ayoubi, N., Saxena, S., Gaitan-Walsh, P., . . . Khoury, S. J. Clinical validation study results of a multivariate proteomic serum based assay for disease activity assessments in multiple sclerosis. Presented at: 37th Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS); October 13-15, 2021; Virtual.

[27] Clinical and Laboratory Standards Institute. (2018). EP283CE. Defining, establishing, and verifying reference intervals in the clinical laboratory. 3rd edition.

[28] Hancox, T. P. M., Skene, D. J., Dallmann, R., Dunn, W. B. (2021). Tick-tock consider the clock: the influence of circadian and external cycles on time of day variation in the human metabolome—a review. *Metabolites*, *11*, 328.

[29] United States Food and Drug Administration. Bioanalytical method validation: guidance for industry. Accessed September 13, 2021. https://www.fda.gov/media/70858/download

[30] Clinical and Laboratory Standards Institute. (2016). EP07. Interference testing in clinical chemistry. 3rd edition.

[31] Xu, Y. W., Peng, Y. H., Chen, B., Wu, Z. Y., Wu, J. Y., Shen, J. H., . . . Xu, L. Y. (2014). Autoantibodies as potential biomarkers for the early detection of esophageal squamous cell carcinoma. *Am J Gastroenterol*, *109*, 36-45.

[32] Yang, X., Suo, C., Zhang, T., Yin, X., Man, J., Yuan, Z., . . . Ye, W. (2021). Targeted proteomics-derived biomarker profile develops a multi-protein classifier in liquid biopsies for early detection of esophageal squamous cell carcinoma from a population-based case-control study. *Biomark Res*, *9*, 12.

[33] Hijazi, Z., Wallentin, L., Lindbäck, J., Alexander, J. H., Connolly, S. J., Eikelboom, J. W., . . . Siegbahn, A. (2020). Screening of multiple biomarkers associated with ischemic stroke in atrial fibrillation. *J Am Heart Assoc*, *9*, e018984.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[34] Sanders-van Wijk, S., Tromp, J., Beussink-Nelson, L., Hage, C., Svedlund, S., Saraste, A., . . . Shah, S. J. (2020). Proteomic evaluation of the comorbidity-inflammation paradigm in heart failure with preserved ejection fraction: results from the PROMIS-HFpEF study. *Circulation*, *142*, 2029-2044.

[35] Pol, T., Hijazi, Z., Lindbäck, J., Oldgren, J., Alexander, J. H., Connolly, S. J., . . . Wallentin, L. (2021). Using multimarker screening to identify biomarkers associated with cardiovascular death in patients with atrial fibrillation. *Cardiovasc Res*, cvab262.

[36] Feldreich, T., Nowak, C., Fall, T., Carlsson, A. C., Carrero, J. J., Ripsweden, J., . . . Ärnlöv, J. (2019). Circulating proteins as predictors of cardiovascular mortality in end-stage renal disease. *J Nephrol*, *32*, 111-119.

[37] Nowak, C., Carlsson, A. C., Östgren, C. J., Nyström, F. H., Alam, M., Feldreich, T., . . . Ärnlöv, J. (2018). Multiplex proteomics for prediction of major cardiovascular events in type 2 diabetes. *Diabetologia*, *61*, 1748-1757.

[38] Leandersson, P., Åkesson, A., Hedenfalk, I., Malander, S., Borgfeldt, C. (2020). A multiplex biomarker assay improves the diagnostic performance of HE4 and CA125 in ovarian tumor patients. *PLoS One*, *15*, e0240418.

[39] Enroth, S., Berggrund, M., Lycke, M., Broberg, J., Lundberg, M., Assarsson, E., . . . Gyllensten, U. (2019). High throughput proteomics identifies a high-accuracy 11 plasma protein biomarker signature for ovarian cancer. *Commun Biol*, *2*, 221.

[40] Berggrund, M., Enroth, S., Lundberg, M., Assarsson, E., Stålberg, K., Lindquist, D., . . . Gyllensten, U. (2019). Identification of candidate plasma protein biomarkers for cervical cancer using the multiplex proximity extension assay. *Mol Cell Proteomics*, *18*, 735-743.

[41] Whelan, C. D., Mattsson, N., Nagle, M. W., Vijayaraghavan, S., Hyde, C., Janelidze, S., . . . Hansson, O. (2019). Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol Commun*, *7*, 169.

[42] Petrackova, A., Smrzova, A., Gajdos, P., Schubertova, M., Schneiderova, P., Kromer, P., . . . Kriegova, E. (2017). Serum protein pattern associated with organ damage and lupus nephritis in systemic lupus erythematosus revealed by PEA immunoassay. *Clin Proteomics*, *14*, 32.

[43] Brunner, P. M., He, H., Pavel, A. B., Czarnowicki, T., Lefferdink, R., Erickson, T., . . . Paller, A. S. (2019). The blood proteomic signature of early-onset pediatric atopic dermatitis shows systemic inflammation and is distinct from adult long-standing disease. *J Am Acad Dermatol*, *81*, 510-519.

[44] Moaddel, R., Ubaida-Mohien, C., Tanaka, T., Lyashkov, A., Basisty, N., Schilling, B., . . . Ferrucci, L. (2021). Proteomics in aging research: a roadmap to clinical, translational research. *Aging Cell*, *20*, e13325.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIGURE LEGENDS

FIGURE 1. Incurred sample reanalysis results for (A) Individual biomarkers and (B) Overall Disease Activity score, (C) Immunomodulation, (D) Neuroinflammation, (E) Myelin Biology, and (F) Neuroaxonal Integrity pathway scores in the MSDA Test.

FIGURE 2. Accuracy of the MSDA Test to detect (A) Individual biomarkers and (B) Overall Disease Activity score, (C) Immunomodulation, (D) Neuroinflammation, (E) Myelin Biology, and (F) Neuroaxonal Integrity pathway scores.

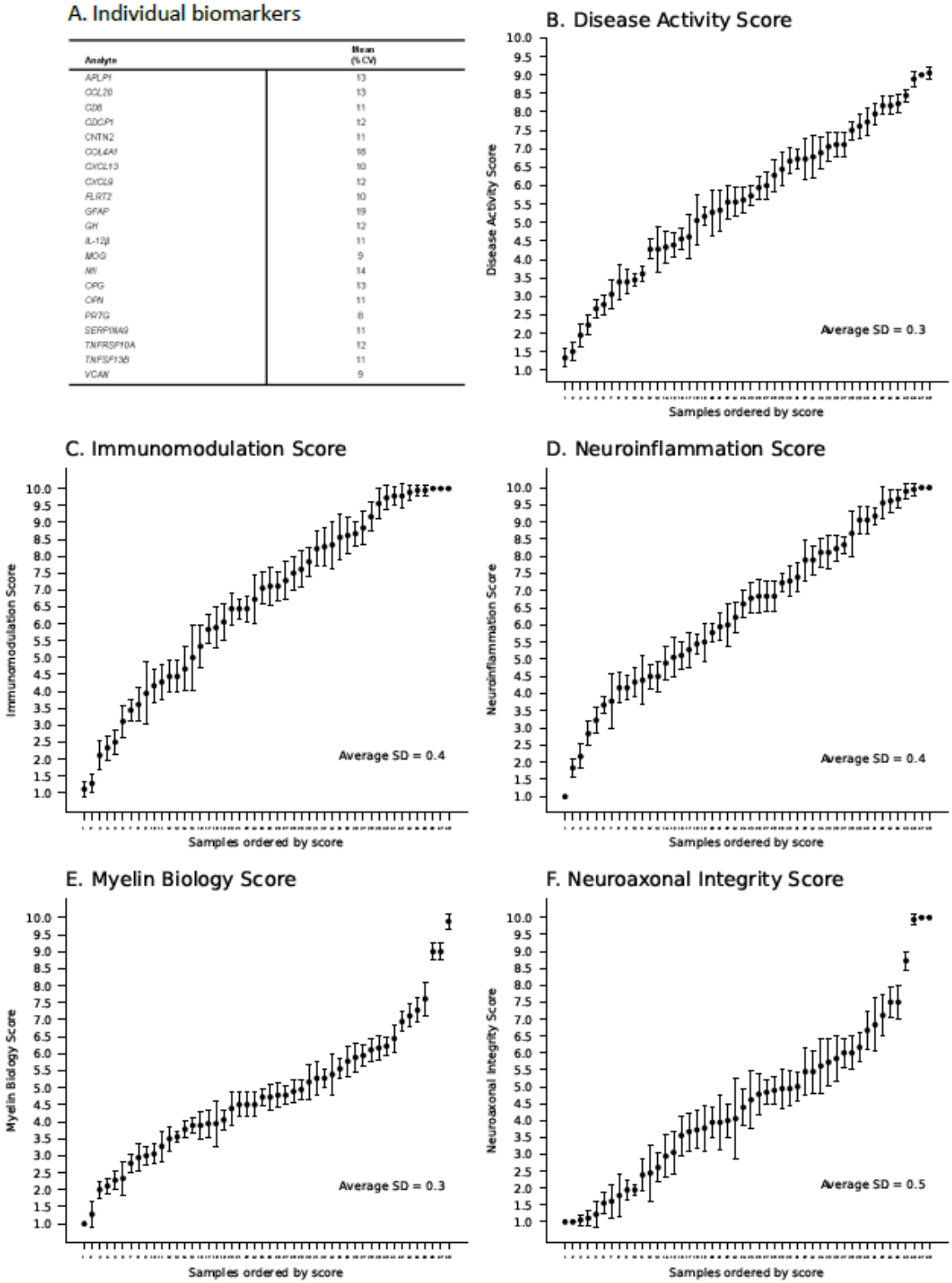
FIGURE 3. Assay interference for common MS drugs, DMTs, and the high concentration of universal mAb surrogates in the MSDA Test.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIGURE 1

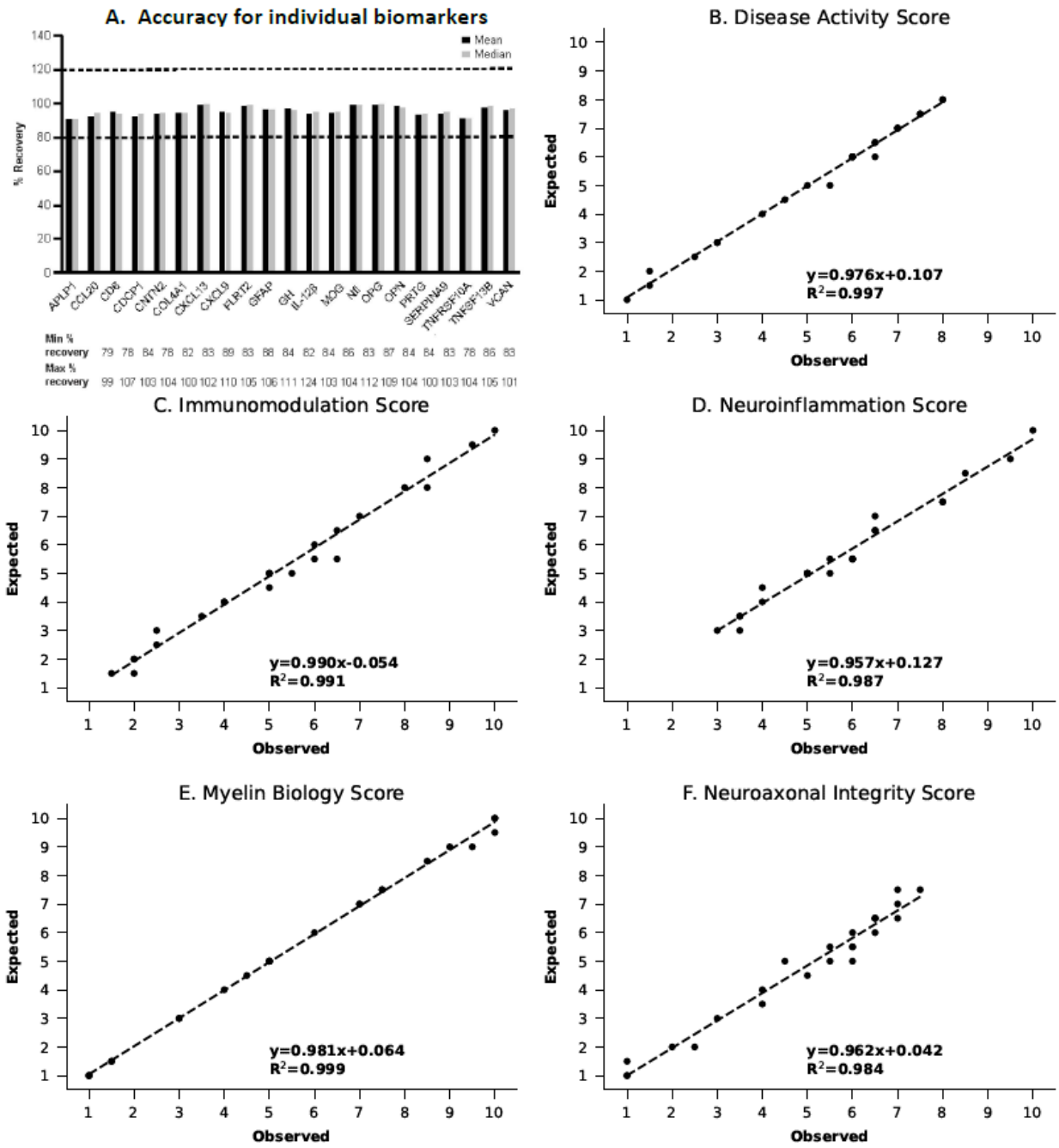
A. Individual biomarkers

Analyte	Mean (%CV)
APLP1	13
CCL20	13
CD6	11
CDCP1	12
CNTN2	11
CCL4/4*	16
CXCL13	10
CXCL9	12
FLRT2	10
GFAP	19
GH	12
IL-12β	11
MOG	9
NN	14
OPG	13
OPN	11
PRTG	8
SERPINA9	11
TNFRSF10A	12
TNFSF13B	11
VCAN	9



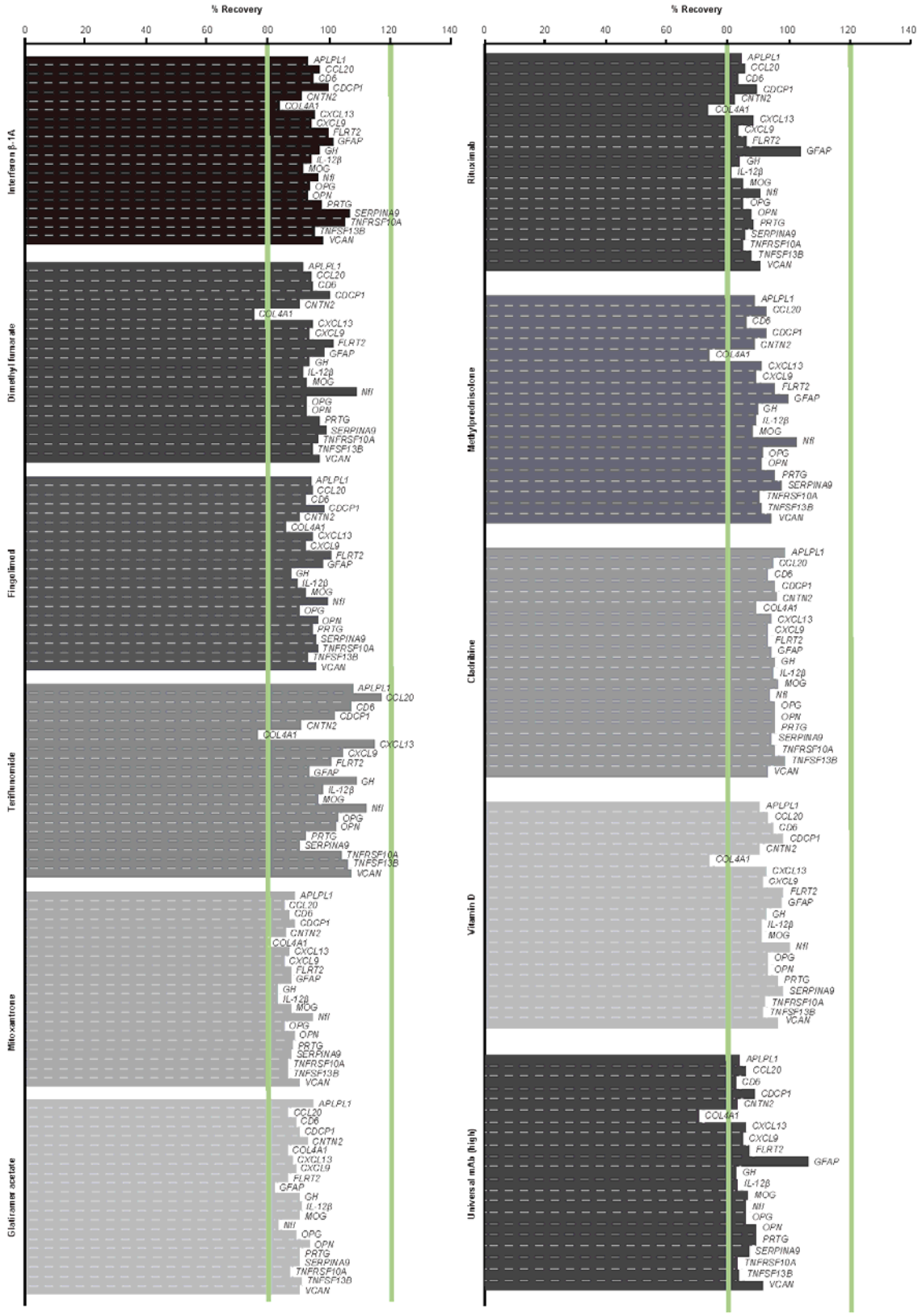
medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIGURE 2



medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIGURE 3



TABLES

TABLE 1 Intra- and inter-assay precision, sensitivity, and reference ranges for biomarkers in the MSDA Test

	Precision												***Sensitivity and reference ranges***					
Analytes	**Shorter MS duration pool[a]**			**Longer MS duration pool[b]**			**RA pool**			**Healthy control pool**								
	Intra %CV	**Inter %CV**	**Conc (pg/mL)**	**Intra %CV**	**Inter %CV**	**Conc (pg/mL)**	**Intra %CV**	**Inter %CV**	**Conc (pg/mL)**	**Intra %CV**	**Inter %CV**	**Conc (pg/mL)**	**LLOQ (pg/mL)**	**ULOQ (pg/mL)**	**Low MS range (pg/mL)[c]**	**High MS range (pg/mL)[d]**	**Samples imputed at LLOQ, % (*N*=1645)**	**Samples imputed at ULOQ, % (*N*=1645)**
APLP1	9	9	10,296	9	8	11,560	4	8	11,868	7	9	11,868	2323.78	142,798.49	5500	22,000	0	0
CCL20	6	9	6.9	8	7	9.2	5	9	13.7	7	9	11.8	0.92	383.49	2.1	52	0	1 (<0.1)
CD6	6	8	89	8	8	108	5	7	137	8	9	112	4.62	3318.60	46	250	0	0
CDCP1	8	9	78	8	9	125	4	9	208	7	10	72	24.22	6795.04	28	230	23 (1.4)	1 (<0.1)
CNTN2	6	7	1120	7	7	1643	4	6	1554	8	8	1256	44.46	12,373.51	650	3300	0	0
COL4A1	7	15	1104	19	20	1334	8	59	1601	47	27	1387	30.65	4573.38	520	3600	0	23 (1.4)
CXCL13	6	8	52.8	8	7	42.9	7	8	65.3	7	9	46.8	1.91	1112.70	22	190	0	0
CXCL9	6	11	31.0	8	10	62.6	5	11	112.3	7	11	27.5	1.89	1832.22	17	250	0	0
FLRT2	7	8	103	9	9	110	5	8	139	8	9	116	35.67	10,107.17	63	180	1 (<0.1)	0
GFAP	7	18	70	10	16	126	8	15	148	9	18	77	12.46	19,582.88	24	220	16 (1.0)	0
GH	7	9	823	8	7	595	5	8	1010	7	9	366	9.63	18,413.83	17	9500	7 (0.4)	9 (0.5)
IL-12β	7	9	109	9	7	122	6	8	118	8	9	71	0.56	3044.33	28	280	0	0
MOG	4	6	21.9	7	6	22.8	5	7	26.0	6	8	17.8	1.75	577.42	12	47	0	0
NfL	10	11	7.6	13	9	15.6	8	8	20.6	11	12	6.5	3.31	599.18	3.5	42	29 (1.8)	0
OPG	6	11	699	9	11	806	6	10	1022	7	12	602	14.58	62,385.21	410	1400	0	0
OPN	6	10	15,733	7	10	15,415	6	12	17,470	7	13	10,450	572.50	157,267.29	9500	39,000	0	0
PRTG	7	6	94	8	7	107	5	6	108	7	7	103	3.90	5920.73	71	180	2 (0.1)	0
SERPINA9	11	8	45.1	11	8	37.9	5	7	60.0	6	14	50.0	5.12	9286.67	12	160	1 (<0.1)	0
TNFRSF10A	9	9	5.1	9	9	5.5	5	8	7.6	9	9	4.9	0.48	1027.28	2.8	9.7	0	0
TNFSF13B	7	10	4075	11	11	4019	5	11	4204	7	13	3003	660.29	130,682.08	2300	10,000	0	0
VCAN	7	8	316	7	7	337	4	7	448	5	8	310	8.54	14,673.95	230	600	0	0

Green shading: intra- (%CV ≤15%) or inter-assay (%CV <20%) are within the acceptability range for the assays. Red shading: intra- (%CV >15%) or inter-assay (%CV ≥20%) precision values are outside the acceptability range for the assays.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[a]Average age of patients with shorter MS duration was 36 (range, 27□43) years.

[b]Average age of patients with longer MS duration was 52 (range, 45□62) years.

[c]Low MS range was defined as the 2.5th percentile.

[d]High MS range was defined as the 97.5th percentile.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TABLE 2 Diurnal variability of eight samples in the MSDA Test across six time points (days 1, 2, 3, 4, 5, and 12)

ANALYTE	MEAN %CV	MEDIAN %CV
APLP1	14	15
CCL20	25	21
CD6	10	7
CDCP1	12	12
CNTN2	11	10
COL4A1	44	39
CXCL13	18	11
CXCL9	13	11
FLRT2	10	8
GFAP	13	12
GH	78	79
IL-12B	10	9
MOG	12	11
NFL	17	17
OPG	11	10
OPN	10	8
PRTG	8	6
SERPINA9	12	12
TNFRSF10A	10	12
TNFSF13B	11	8
VCAN	10	7

Green shading: %CV ≤30%. Red shading: %CV > 30%.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Analytical Validation of a multi-protein, proteomic serum-based assay for disease activity assessments in multiple sclerosis

Ferhan Qureshi,* Wayne Hu,* Louisa Loh, Hemali Patel, Maria DeGuzman, Michael Becich, Fatima Rubio da Costa, Victor Gehman, Fujun Zhang, John Foley, Tanuja Chitnis

*These authors contributed equally to this work

Supporting Information

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

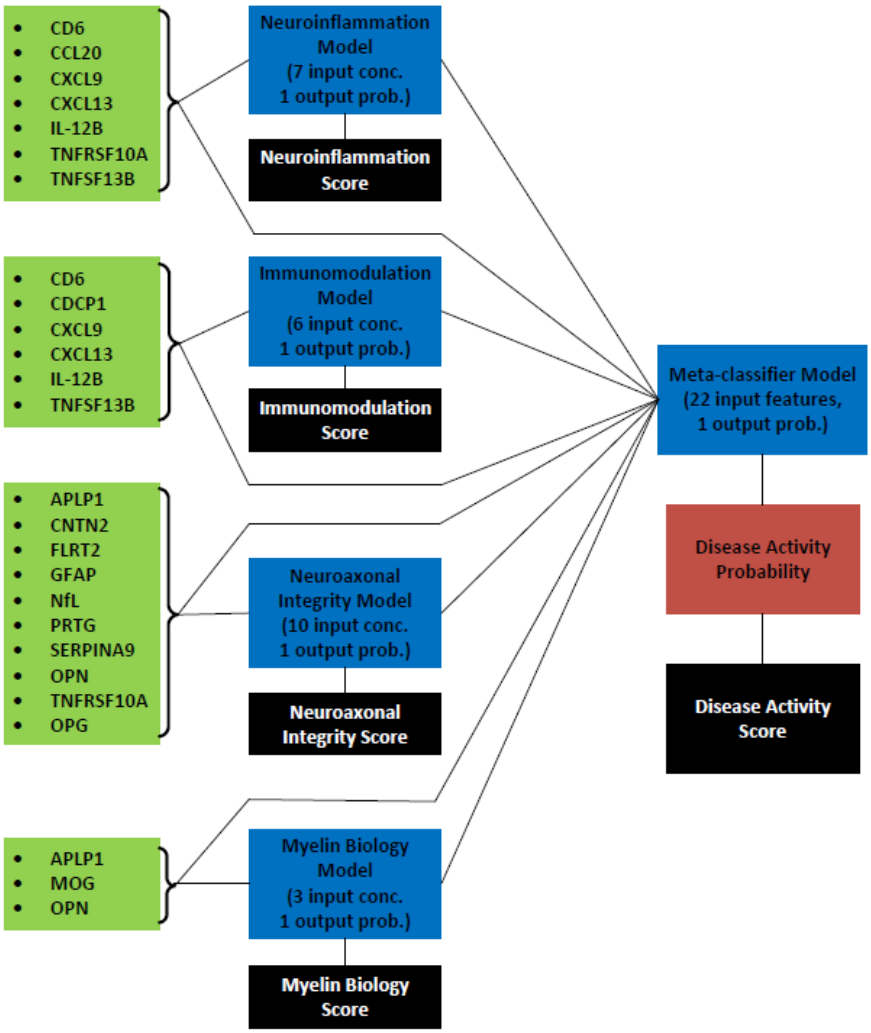
FILE S1. MSDA Test algorithm formula

Key to the MSDA Test was the architecture of a stacked classifier (in the mlxtend [1] framework) where each node (blue boxes in the **Figure**) in both layers are L2-regularized logistic regression models from Scikit Learn.[2] The input layer read in demographically corrected for age and sex, $\log_{10,}$ LOQ-imputed concentrations of proteins associated with four physiological pathways from each sample and generated probabilities that the sample in question was positive for the presence of Gd+lesions. The meta-classifier (output) layer of the model ingested the probabilities from each pathway model along with all of the protein concentrations that fed the first layer and generated an overall probability that the sample in question was Gd+ lesion-positive. This probability was then mapped into a Disease Activity score by being shifted, scaled, rounded, and clipped to the nearest half point over a range from 1 to 10 (Eq. 1). A similar process converted the pathway probabilities into pathway scores after the pathway probabilities were adjusted so that for each sample, the mean over the adjusted pathway probabilities was equal to the Disease Activity probability for each sample (Eq. 2). The datasets used to develop and evaluate the demographic correction, four Disease Pathway models, and Disease Activity model, as well as the performance of these models, have been described previously.[2]

The process for calculating the Disease Activity score is shown in the **Figure**.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIGURE. Schematic representation of the Disease Activity score calculation from protein concentrations (green) to both layers of the stacked classifier (blue), the output Disease Activity probability (red), and the final Disease Activity and Disease Pathway score (black).



The overall Disease Activity score had the mathematical form:

$$S_{DA} = \text{Clip}_0^9\left(\frac{1}{2}\text{Round}\left(18 \times \frac{P_{DA} - P_{DA}^{Low}}{P_{DA}^{High} - P_{DA}^{Low}}\right)\right) + 1 \quad (1)$$

The “Clip” function limited the output of its argument to between 0 and 9. The “Round” function rounded to the nearest whole number. P_{DA} was the Disease Activity probability output from the final stage of the stacked classifier model. P^{High}_{DA} and

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

P^{Low}_{DA} were probability clip values chosen for the overall Disease Activity (and each pathway) to maximize the range of the score over the entire training data set. Their values are tabulated below.

Similar to the Disease Activity score, each pathway score was computed from the pathway probabilities with the additional step of recentering the ensemble of pathway probabilities around the Disease Activity probability for each sample:

$$S_{pth} = \mathrm{Clip}_0^9\left(\frac{1}{2}\mathrm{Round}\left(18 \times \left(\frac{P_{pth} - P_{pth}^{Low}}{P_{pth}^{High} - P_{pth}^{Low}} + \frac{P_{DA} - P_{DA}^{Low}}{P_{DA}^{High} - P_{DA}^{Low}} - \frac{1}{N_{pth}}\sum_{pth'}\frac{P_{pth'} - P_{pth'}^{Low}}{P_{pth'}^{High} - P_{pth'}^{Low}}\right)\right)\right) + 1$$

(2)

"Clip" and "Round" served the same purpose in Eq. 2 as in Eq.1. P_{pth} was the probability for pathway model *pth*. P_{DA} was again the probability output by the final stage of the stacked classifier. The "High" and "Low" *P*s for each *pth* served a similar purpose to those for the Disease Activity: rescaling the model probabilities so that they had a larger dynamic range. Lastly, N_{pth} was the number of pathway models that fed into the final layer of the stacked classifier model. In this case, N_{pth}=4.

The probabilities associated with each layer of the stacked classifier had a closed analytic form as well. Working backward from the final Disease Activity score, the Disease Activity probability had the form:

$$P_{DA} = \frac{1}{1 + e^{-\left(C^0_{DA} + \sum_{feat} C_{feat} V_{feat}\right)}}$$

(3)

C^0_{DA} was the intercept (or bias) value for the Disease Activity meta-classifier. C_{feat} was the coefficient associated with V_{feat}. V_{feat} was the value of the feature being summed, either the probability for one of the pathways from the first layer of the model (written out explicitly in Eq. 4) or the age- and sex- corrected, $\log_{10}$, LOQ-imputed concentrations that fed into the first-layer pathway models. These meta-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

classifier coefficients were fit using L2-penalized logistic regression with Gd+ lesion presence as the dependent variable, an intercept, and the stacked classifier features as the independent variables, with inverse regularization strength C=1.0, balanced class weight, and tolerance =0.0001, with the 'lbfgs' solver.

The individual pathway probabilities had the form:

$$P_{pth} = \frac{1}{1 + e^{-(C^0_{pth} + \sum_{prot} C_{prot} V_{prot})}} \quad (4)$$

C^0_{pth} was the intercept value for the pathway model in question. C_{prot} was the coefficient for a protein in pathway *pth*. V_{prot} was the demographically corrected, $\log_{10}$, LOQ-imputed concentration of protein *prot*. These coefficients were fit using L2-penalized logistic regression with Gd+ lesion presence as the dependent variable, an intercept, and the pathway's demographically corrected (for age, sex), $\log_{10}$, LOQ-imputed protein concentrations as the independent variables. The final pathway coefficients were the result of a stratified, 10-fold cross-validation with 3 repeats, grid-searched over inverse regularization strength C=[100, 10, 1.0, 0.1, 0.01], class weights=[None, balanced], solvers=["saga," "lbfgs," "liblinear"], and tolerances=[0.01, 0.001, 0.0001]. The demographic correction for the $\log_{10}$ of the LOQ-imputed protein concentration had the form:

$$V_{prot} = log_{10}\left(A_{prot}\left[\frac{pg}{mL}\right]\right) - [D_{prot,0} + (D_{prot,Sex} \times Sex) + (D_{prot,Age} \times Age[years])] \quad (5)$$

A_{prot} was the LOQ-imputed absolute concentration of protein *prot* in pg/mL (the units of these direct measurements have been written explicitly for added clarity). $D_{prot,\,0}$, $D_{prot,\,Sex}$, and $D_{prot,\,Age}$ were the intercept, sex, and age coefficients for demographic correction, respectively. These coefficients were fit using the following procedure on three datasets not reported here. The dataset used to train the stacked classifier and pathway models provided the final non-zero coefficients for the

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

demographic correction. The demographic correction coefficients were fit independently on each dataset using linear regression using StatsModels OLS[4] with $\log_{10}$, LOQ-imputed protein concentration as the dependent variable, an intercept, and age and sex as the independent variables, after outliers were removed. Outliers were defined for each dataset as ≥96th percentile or ≤4th percentile for $\log_{10}$, LOQ-imputed protein concentration. After fitting the linear regression, the age and sex coefficients were set to zero if the 95% confidence interval contained zero in any of the three datasets or if the sign of the coefficient was not consistent across three datasets. Sex was 0 for women and 1 for men, and age was the patient's age in years at time of blood draw.

Finally, see below for tables of the coefficients used in each step of the scoring process. First, we tabulated the demographic correction coefficients. These were the *D*s in Eq. 5.

Protein	**Sex**	**Age**	**Intercept**
APLP1	−0.052	0	4.065
CCL20	0	0	1.088
CD6	0	0	1.959
CDCP1	0	0.006	1.64
CNTN2	0	0.002	3.128
CXCL13	0	0	1.819
CXCL9	0	0.005	1.419
FLRT2	0	0	1.968
GFAP	−0.05	0.005	1.692
IL-12B	0	0	2.062
MOG	−0.04	0.003	1.267
NfL	0	0.008	0.5
OPG	−0.014	0.002	2.743

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

OPN	0.045	0	4.172
PRTG	0	0	2.052
SERPINA9	0	0	1.56
TNFRSF10A	0	0.003	0.565
TNFSF13B	0	0	3.75

Next, we tabulated the protein coefficients for each pathway. These were the *C*s in Eq. 4.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

- Neuroinflammation:

Feature	Coefficient
CCL20	0.397
CD6	0.243
CXCL13	0.668
CXCL9	0.778
IL-12B	−1.515
TNFRSF10A	0.089
TNFSF13B	−2.705
Intercept	−0.283

- Immunomodulation:

Feature	Coefficient
CD6	0.825
CDCP1	−0.633
CXCL13	1.075
CXCL9	1.178
IL-12B	−1.657
TNFSF13B	−4.232
Intercept	−0.483

- Neuroaxonal Integrity:

Feature	Coefficient
APLP1	−0.149
CNTN2	−0.434
FLRT2	−0.321
GFAP	0.024
NfL	2.998
OPG	0.384

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

OPN	0.026
PRTG	−0.621
SERPINA9	−0.779
TNFRSF10A	−0.284
Intercept	−0.744

- Myelin Biology:

Feature	**Coefficient**
APLP1	−1.206
MOG	2.653
OPN	−0.037
Intercept	−0.552

Then we tabulated the meta-classifier coefficients for the second stage of the stacked classifier. These were the *C*s in Eq. 3.

Feature	**Coefficient**
APLP1	−0.363
CCL20	0.222
CD6	0.119
CDCP1	−0.524
CNTN2	−0.21
CXCL13	0.1
CXCL9	0.252
FLRT2	−0.111
GFAP	0.042
IL-12B	−0.432
MOG	0.485
NfL	2.038
OPG	0.056
OPN	−0.072
PRTG	0.768
SERPINA9	0.142
TNFRSF10A	−0.178
TNFSF13B	−1.3
Neuroinflammation pathway	1.14

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

classifier	
Immunomodulation pathway classifier	1.737
Neuroaxonal integrity pathway classifier	1.525
Myelin biology pathway classifier	0.435
Intercept	−2.51

Finally, we tabulated the low and high values for scaling the probabilities in Eq.1 and Eq.2.

Disease Pathway/Disease Activity score	$P_{Low\ pth}$	$P_{High\ pth}$
Disease Pathway		
Immunomodulation	0.05	1.0
Myelin Biology	0.2	0.6
Neuroaxonal Integrity	0.1	1.0
Neuroinflammation	0.15	0.95
Disease Activity	0.05	1.00

Abbreviations: $P_{High\ pth}$, high values for scaling probabilities; $P_{Low\ pth}$, low values for scaling probabilities.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FILE S2. Assay interferents

Samples were analyzed for potential interference with endogenous substances found naturally in patients' samples (eg, lipemia, hemolysis, high protein count, and high bilirubin levels) and heterophilic antibodies such as HAMA and RF. Serum pools (*n*=4) were spiked with bilirubin, hemolysate, and lipids at typical concentrations using the ASSURANCE™ Interference Test Kit (Sun Diagnostics).

Heterophilic antibodies, including RF and HAMA, are established sources of potential interference in immunoassays. RF concentrate (Lee BioSolutions, Maryland Heights, MO, USA) was used to spike six serum samples at low (150 IU/mL) and high (2000 IU/mL) concentrations to determine the effect of RF interference on the analysis. For HAMA interference, five HAMA-positive serum samples with known established HAMA levels (ASSURANCE™ Interference Test Kit) were mixed at different ratios (10:90; 50:50; 90:10) with four MS samples from an internal cohort. Two HAMA-positive samples had a titer level >240 and three samples had a titer level >480, indicating that the positive samples had either >240 or >480 times more activity than a known negative. The mixed samples were compared with expected ratios of the individually measured samples to calculate the percent recovery.

Except for heterophilic antibody interference, the mean percentage recovery was calculated relative to the corresponding spike control (representing the same alteration of the serum sample without the addition of the interferent). For heterophilic antibody interference, the mean percentage recovery was calculated by comparing the RF spiked (low and high concentrations) samples with a corresponding spike control.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TABLE S1 Protein biomarkers included in assay panel to perform the MSDA Test

Biomarker	Full name	UniProt identifier
APLP1	Amyloid beta precursor-like protein 1	P51693
CCL20	C-C motif chemokine ligand 20 (MIP 3-alpha)	P78556
CD6	Cluster of differentiation 6	P30203
CDCP1	CUB domain-containing protein 1	Q9H5V8
CNTN2	Contactin 2	Q02246
CXCL13	Chemokine (C-X-C motif) ligand 13	P02462
CXCL9	Chemokine (C-X-C motif) ligand 9 (MIG)	O43927
COL4A1 [a]	Collagen type IV alpha 1	Q07325
FLRT2	Fibronectin leucine-rich repeat transmembrane protein	O43155
GFAP	Glial fibrillary acidic protein	P14136
GH [a]	Growth hormone (somatotropin)	P01241
IL-12β	Interleukin-12 subunit beta	P29460
MOG	Myelin oligodendrocyte glycoprotein	Q16653
NfL	Neurofilament light polypeptide chain	P07196
OPG	Osteoprotegerin	O00300
OPN	Osteopontin	P10451
PRTG	Protogenin	Q2VWP7
SERPINA9	Serpin family A member 9	Q86WD7
TNFRSF10A	Tumor necrosis factor receptor superfamily member 10A (TRAIL-R1)	O00220
TNFSF13B	Tumor necrosis factor superfamily member 13B (BAFF)	Q9Y275
VCAN [a]	Versican core protein	P13611

[a]These biomarkers were not used in the algorithm to determine Disease Pathway and Disease Activity scores.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TABLE S2 Common over-the-counter and prescription drugs, reference materials and standards, and mAbs evaluated for interference in the MSDA Test

Assay interferents	Concentration
Common drugs	
Acetaminophen	15.6 mg/dL
Ampicillin Na	7.5 mg/dL
Acetylsalicylic acid	3 mg/dL
Doxycycline HCl	1.8 mg/dL
Cefoxitin Na	660 mg/dL
Heparin	3300 U/L
Cyclosporine	0.18 mg/dL
Levodopa	0.75 mg/dL
Ibuprofen	21.9 mg/dL
Methyldopa	2.25 mg/dL
Phenylbutazone	32.1 mg/dL
Metronidazole	12.3 mg/dL
Rifampicin	4.8 mg/dL
Theophylline	6 mg/dL
Acetylcysteine	15 mg/dL
Reference materials and standards to evaluate DMTs	
Interferon β-1A (Avonex, Rebif)	0.56 ng/mL
Dimethyl fumarate (Tecfidera)	3.74 µg/mL
Fingolimod (Gilenya)	8.8 ng/mL
Teriflunomide (Aubagio)	90.6 µg/mL
Mitoxantrone (Novantrone)	0.674 µg/mL
Rituximab (Rituxan)	370 µg/mL
Methylprednisolone	0.141 µg/mL
Cladribine (Mavenclad)	0.058 µg/mL
Glatiramer acetate (Copaxone, Glatopa)	1.0 µg/mL
Vitamin D2 (Ergocalciferol)	0.2 µg/mL

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Universal mAb surrogates[a]	
High concentration	424 µg/mL
Low concentration	7.93 µg/mL

[a]Surrogates for natalizumab, ocrelizumab, alemtuzumab, and ofatumumab.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TABLE S3 Lot to lot (batch 1 vs. batch 2) and laboratory to laboratory (development vs. clinical laboratory) consistency of the MSDA Test

	Mean % difference		**R^2 correlation of samples**	
	Batch 1 vs. batch 2	**Development vs. clinical laboratory**	**Batch 1 vs. batch 2**	**Development vs. clinical laboratory**
APLP1	0	−1	0.93	0.95
CCL20	−14	−12	0.99	1.00
CD6	−9	−9	0.97	0.97
CDCP1	−9	−10	0.99	1.00
CNTN2	8	4	0.96	0.97
COL4A1	−27	−17	0.97	0.98
CXCL13	−7	−5	0.99	1.00
CXCL9	−16	−11	1.00	1.00
FLRT2	−3	−6	0.95	0.98
GFAP	−13	−9	0.96	0.97
GH	−3	−6	0.98	0.99
IL-12β	−5	−5	0.97	0.98
MOG	2	0	0.98	0.98
NfL	−13	−10	0.98	0.99
OPG	−17	−9	0.97	0.98
OPN	−14	−7	0.97	0.99
PRTG	0	−5	0.97	0.98
SERPINA9	−4	−5	0.99	0.99
TNFRSF10A	−7	−11	0.98	0.99
TNFSF13B	−6	−5	0.96	0.98

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

VCAN	−8	−9	0.91	0.95

TABLE S4 Storage, processing, and stability of the individual biomarkers at various temperatures (room temperature, 4°C, and −20°C) over time during the initial and follow-up studies

	Mean % difference vs. experimental control conditions																				
	Initial study																	Follow-up study			
	Room temperature						4°C						−20°C					4°C			
Analyte	**4 h**	**Day 1**	**Day 3**	**Day 7**	**Day 14**	**Day 28**	**4 h**	**Day 1**	**Day 3**	**Day 7**	**Day 14**	**Day 28**	**Day 1**	**Day 3**	**Day 7**	**Day 14**	**Day 28**	**Day 1**	**Day 2**	**Day 3**	**Day 7**
APLP1	4	17	7	19	48	134	−6	−8	12	31	12	25	5	9	0	5	10	10	5	7	11
CCL20	4	2	−6	−7	8	47	1	−3	11	24	15	23	12	13	6	5	7	-1	-2	4	12
CD6	−2	−4	−14	−5	19	93	−10	−11	−7	7	−2	10	1	5	−1	1	1	2	-1	2	5
CDCP1	−5	−1	−13	−2	26	114	−10	−10	−2	6	1	10	0	7	−3	1	1	4	1	3	4
CNTN2	0	−3	−13	−3	24	71	−8	−10	−9	2	−3	7	2	6	1	2	3	3	0	1	1
COL4A1	−5	−3	−14	−1	22	91	−10	−11	−12	2	−2	10	0	4	9	3	2	9	-3	11	14
CXCL13	−1	−7	−25	−27	−24	7	−5	−6	−8	−6	−4	−3	6	12	4	7	5	2	0	0	1

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

CXCL9	−2	−4	−14	−7	16	65	−6	−1 0	−11	−4	−6	−2	2	6	1	3	4	4	0	1	0
FLRT2	−4	−1	−11	0	26	97	−10	−1 0	−2	11	2	13	−1	8	−1	1	2	3	0	2	3
GFAP	0	2	−5	9	39	142	−7	−9	−5	20	1	18	3	11	−1	3	9	1	3	4	8
GH	1	0	−10	−6	9	52	−6	−9	2	10	7	12	4	9	1	5	7	2	3	2	10
IL-12β	−2	−13	−34	−40	−3 2	−24	−7	−1 1	−14	−18	−9	−15	3	8	0	3	4	3	0	1	-4
MOG	−2	0	−10	1	23	83	−7	−9	−1	13	2	12	1	6	1	3	3	1	2	1	3
NfL	−3	−6	−11	2	29	110	−11	−9	1	4	3	12	−4	4	0	2	2	-3	6	1	6
OPG	−2	−4	−13	−5	18	94	−6	−9	−8	1	−1	5	1	6	−1	2	2	2	0	1	1
OPN	−4	−5	−18	−13	1	47	−6	−1 6	−27	−24	−26	−26	0	5	−1	3	−1	2	-2	-2	-8
PRTG	−1	−6	−19	−20	−1 3	−1	−6	−8	−7	−3	−4	−1	3	6	0	2	2	2	0	1	0
SERPINA9	−1	−3	−15	−8	10	43	−7	−1 1	−4	2	0	3	1	5	−1	1	2	9	10	7	9
TNFRSF10A	0	−2	−11	0	31	108	−7	−1 1	1	11	3	10	0	9	−2	0	2	5	0	2	5
TNFSF13B	−4	−7	−20	−10	19	108	−10	−1 4	−12	0	−6	1	1	7	1	3	2	3	-1	3	-1

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

VCAN	−3	−3	−10	−1	23	72	−6	−1 1	−1	9	0	7	0	4	−2	2	1	5	2	4	7

Green shading: %CV within ±20%. Red shading: %CV < −20% or >20%.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

TABLE S5 Storage, processing, and stability of the MSDA Test during various freeze-thaw cycles for individual biomarkers

	Freeze-thaw cycle				
Analyte	**Cycle 1**	**Cycle 2**	**Cycle 3**	**Cycle 4**	**Cycle 5**
APLP1	−6	**6**	−13	−5	−7
CCL20	−7	−2	−13	−10	−11
CD6	−1	**0**	−9	−8	−7
CDCP1	−1	−1	−11	−8	−7
CNTN2	**1**	**3**	−6	−1	−3
COL4A1	**2**	**3**	−11	−5	−7
CXCL13	−1	−5	−11	−10	−9
CXCL9	−1	**1**	−11	−5	−7
FLRT2	−2	−3	−10	−10	−8
GFAP	−9	−4	−20	−23	−28
GH	−7	**1**	−12	−15	−13
IL-12β	−4	**0**	−13	−7	−9
MOG	−3	**1**	−11	−7	−9
NfL	**1**	−2	−12	−12	−1
OPG	−1	−3	−11	−6	−8
OPN	−3	−2	−11	−7	−9
PRTG	−3	−1	−10	−9	−8
SERPINA9	**13**	−2	−13	−11	−12
TNFRSF10A	−8	**2**	−12	−8	−9
TNFSF13B	−1	**0**	−9	−4	−7
VCAN	−1	**1**	−8	−4	−6

Green shading: average % difference within ±20%. Red shading: average % difference < −20%.

TABLE S6 Initial study of the storage, processing, and stability of the MSDA Test at various temperatures (room temperature, 4°C, and −20°C) over time and during freeze-thaw at the Disease Activity score level

ample ID	−80° C	Room temperature						−80° C	4°C						−80°C	−20 C					Sample ID	Freeze-thaw					
		4 h	Day 1	Day 3	Day 7	Day 14	Day 28		4 h	Day 1	Day 3	Day 7	Day 14	Day 28		Day 1	Day 3	Day 7	Day 14	Day 28		Fresh	1	2	3	4	5
A	3	3	3	3	3	3	2.5	3	3.5	2.5	3	3.5	3.5	3	3	2.5	3	3.5	2.5	2.5	E	5	5.5	5	5	4	5.5
B	4	4.5	4.5	4.5	4.5	4.5	3.5	4	4	5	4.5	4.5	5	5	4	4.5	4.5	4.5	5	5	F	6.5	7	6.5	7	6.5	7
C	7	6.5	6	7	7	6.5	6.5	7	7	7.5	7	7	7.5	7	7	6	6.5	7	6.5	7	G	3	3.5	3	3.5	3.5	3.5
D	9	8.5	8.5	9	9	9	8.5	9	9	8.5	9	9	9	9	9	8.5	9	9	8.5	8.5	H	2.5	3	2.5	2.5	2.5	3

Low (1.0□4.0) Disease Activity scores are noted in green, moderate (4.5□7.0) Disease Activity scores are noted in blue, and high (7.5□10.0) Disease Activity scores are noted in orange.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

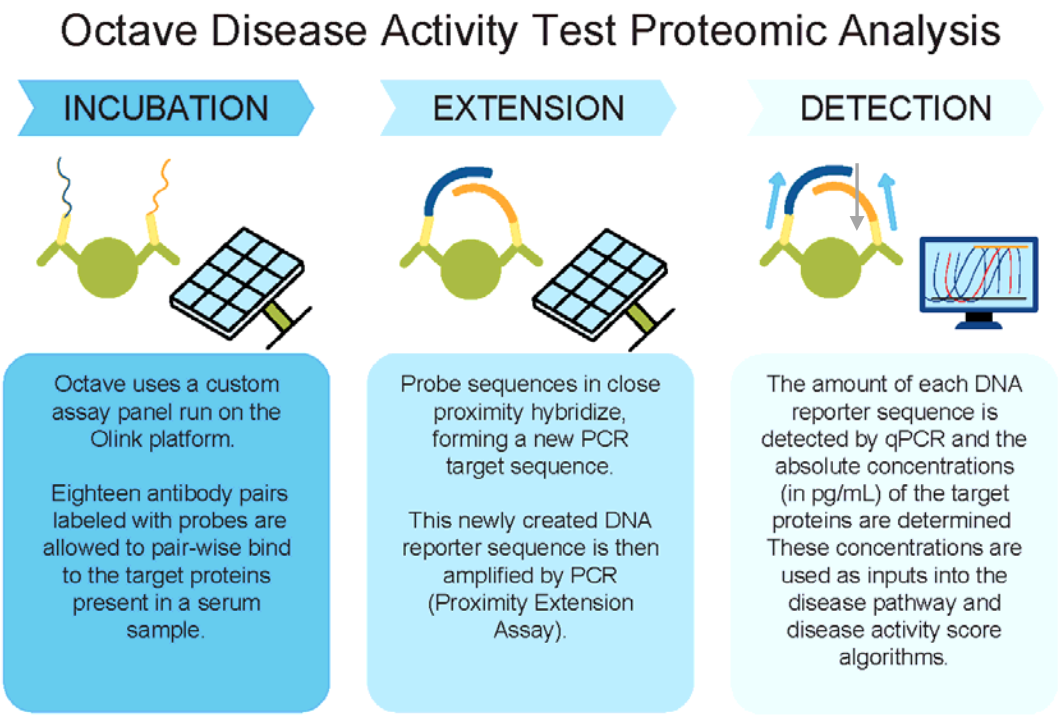
TABLE S7 Follow-up study of the storage, processing, and stability of the MSDA Test at 4°C over time at the Disease Activity score level

Sample ID	−80°C	4°C			
		Day 1	Day 2	Day 3	Day 7
A	8.5	8.5	8.5	8.5	8.5
B	5	4.5	4.5	4.5	4.5
C	7.5	6	7.5	7.5	7
D	6.5	6.5	6.5	5.5	7
E	5	5	5	5	5.5
F	7	7	7	6.5	7.5
G	5.5	5	5	5	5.5
H	6.5	6.5	7.5	7.5	7.5
I	6	6	6.5	6	7
J	3.5	3.5	3	3.5	3
K	1.5	1	1.5	1.5	1.5
L	2.5	2.5	2.5	2.5	2
M	2.5	2.5	3	2	3
N	7.5	7.5	8.5	8	8

Low (1.0□4.0) Disease Activity scores are noted in green, moderate (4.5□7.0) Disease Activity scores are noted in blue, and high (7.5□10.0) Disease Activity scores are noted in orange.

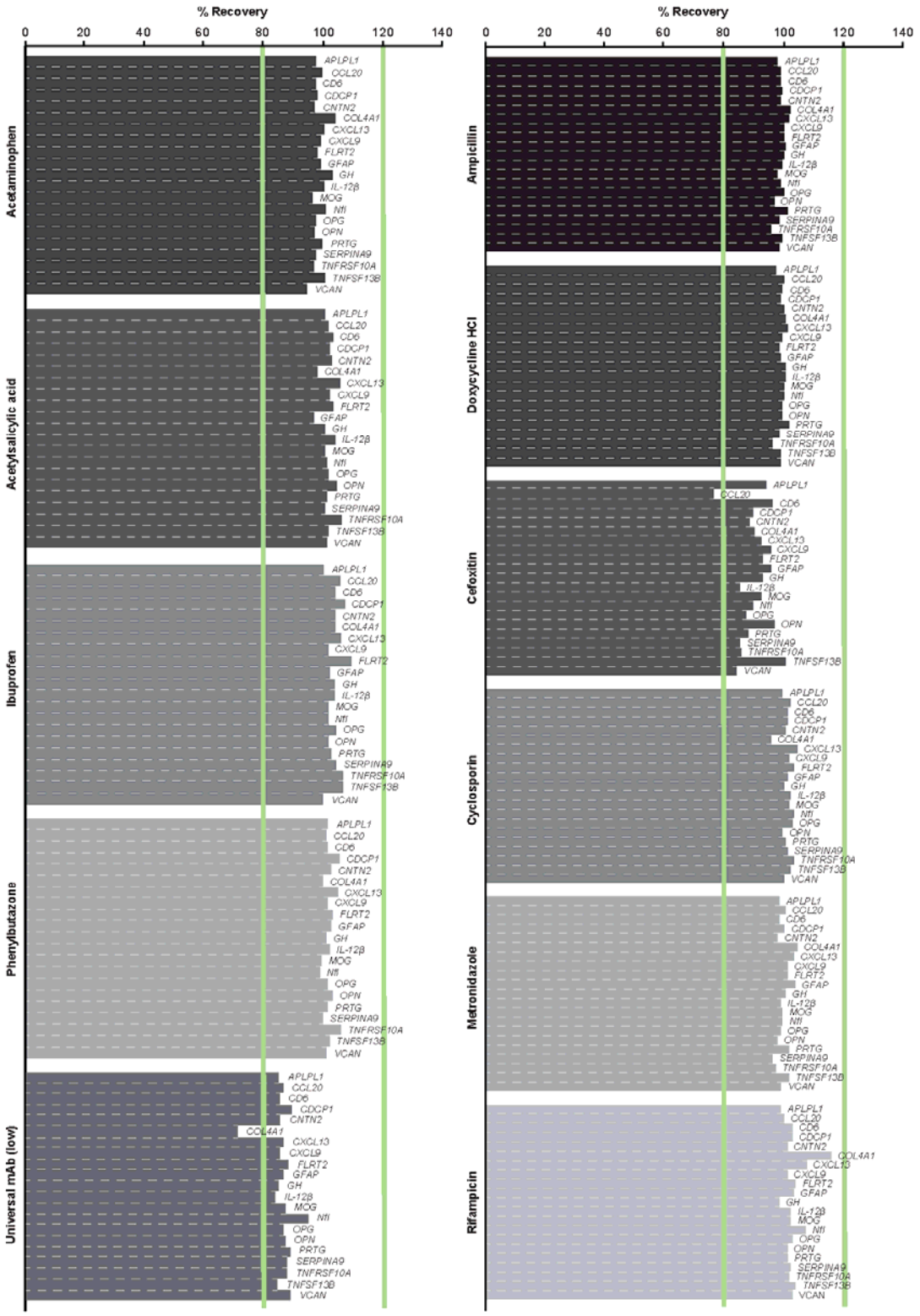
medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIGURE S1. Overview of PEA technology and MSDA Test format. Eighteen antibody pairs, labeled with DNA oligonucleotides, bind to target antigen in solution. Oligonucleotides that are brought into proximity will hybridize and are extended by a DNA polymerase. This newly created piece of DNA barcode is amplified by PCR. The amount of each DNA barcode is quantified by microfluidic qPCR (Biomark™ HD, Fluidigm, Maryland Heights, MO, USA), with results reported in cycle threshold values. Data processing is then performed in the Olink® NPX Manager (Olink Proteomics, Uppsala, Sweden) software to convert the Ct values to a Normalized Protein eXpression value. The signal obtained from the assay (Normalized Protein eXpression) is converted to absolute concentration (pg/mL) using three calibrators that cover the range of sample response in the MS population (calibrators: high, middle, low) and then referenced back to the standard curve. These concentrations are then used as inputs into algorithms corresponding to disease activity and biological pathway scores.



medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIGURE S2. Assay interference using common drugs.



medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIGURE S3. Assay interference using routine endogenous interferents and heterophilic antibodies.

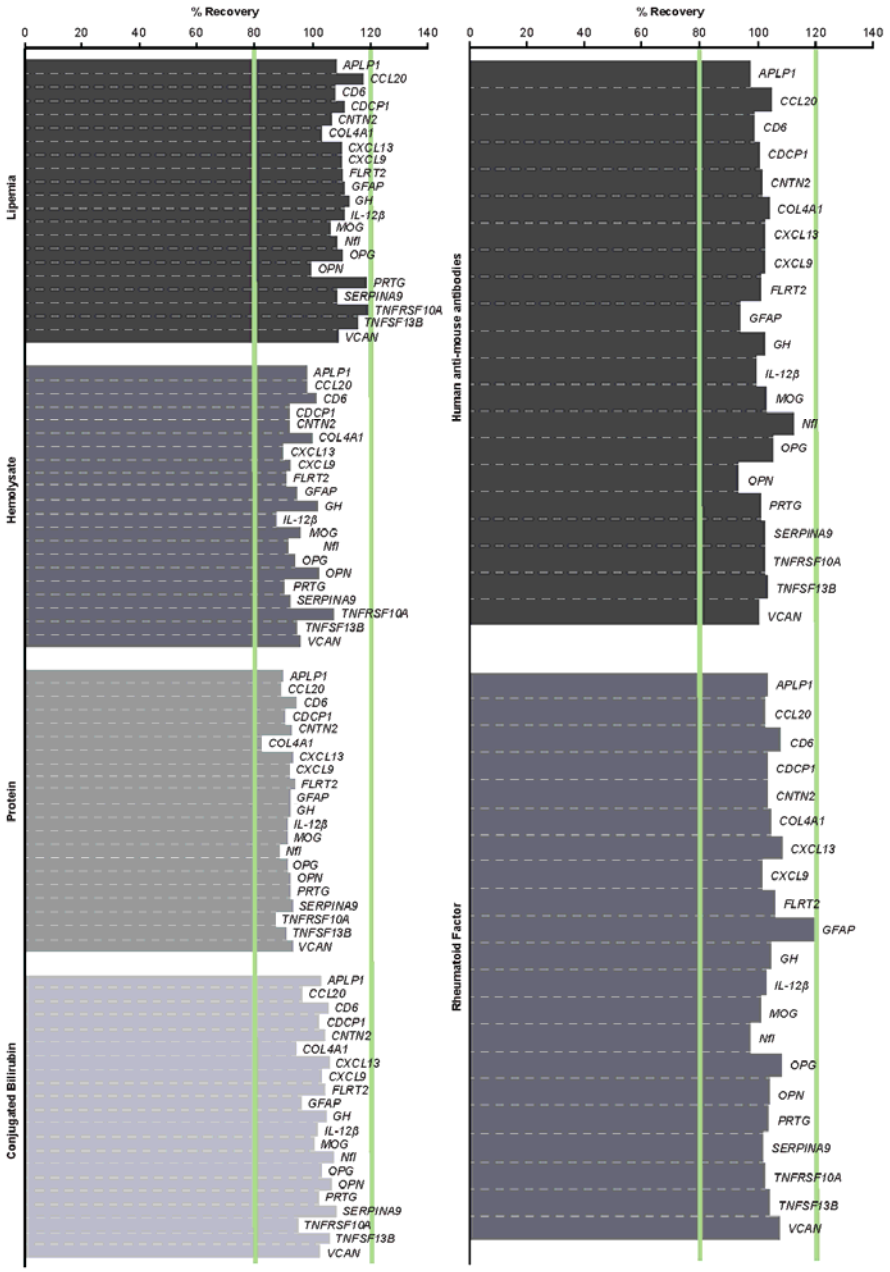
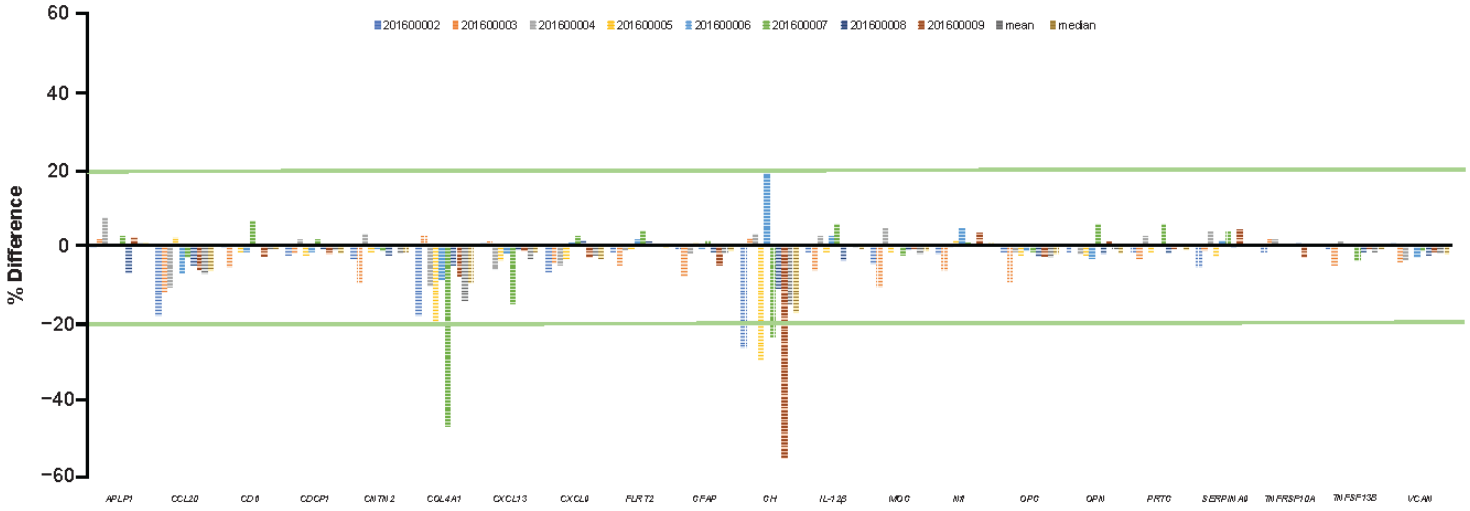


FIGURE S4. Percent difference of the observed protein concentration relative to the average concentration as determined from six time points (days 1, 2, 3, 4, 5, and 12) in eight samples assayed in the MSDA Test.



medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.23.22275201; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

REFERENCES

[1] Raschka, S. (2018). MLxtend: providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J Open Source Softw*, *3*, 638.

[2] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

[3] Chitnis, T., Foley, J., Ionete, C., El Ayoubi, N., Saxena, S., Gaitan-Walsh, P., … Khoury, S. J. Clinical validation study results of a multivariate proteomic serum based assay for disease activity assessments in multiple sclerosis. Presented at: 37th Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS); October 13-15, 2021; Virtual.

[4] Seabold, S., Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference,* 92–96.