Multimodal predictors of functional and cognitive decline in relapsing-remitting multiple sclerosis

Max Korbmacher[1-3]*, Ingrid Anne Lie[4], Kristin Wesnes[5], Eric Westman[6,7], Thomas Espeseth[8,9], Karsten Specht[210] , Ole A. Andreassen[11,12], Lars T. Westlye[8,11, 13], Stig Wergeland[1,13-14], Kjell-Morten Myhr[1,13], Øivind Torkildsen[1,13,†], Einar A. Høgestøl[4,8,15,†]

[†]These authors contributed equally to this work.

Author affiliations:

[1]Neuro-SysMed, Department of Neurology, Haukeland University Hospital, Norway

[2]Mohn Medical Imaging and Visualisation center, Department of Radiology, Haukeland University Hospital, Norway

[3]Department for Radiography, Western Norway University of Applied Sciences, Norway

[4]Department of Neurology, Oslo University Hospital, Norway

[5]Department of Neurology, St. Olav's Hospital, Trondheim University Hospital, Norway

[6]Department of Neurobiology, Karolinska Instituttet, Sweden

[7]Department of Psychiatry, King's College London, UK

[8]Department of Psychology University of Oslo, Norway

[9]Faculty of Psychology, Oslo New University College, Norway

[10]Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway

[11]Center for Precision Psychiatry, sUniversity of Oslo and Oslo University Hospital, Norway

[12]K.G. Jebsen-Centre for Neurodevelopmental disorders, University of Oslo, Norway

[13]Department of Clinical Medicine, University of Bergen, Norway

[14]Norwegian MS-Registry and Biobank, Helse Bergen, Haukeland University Hospital, Norway

[15]Institute for Clinical Medicine, University of Oslo, Norway

Correspondence to: Max Korbmacher

Jonas Lies vei 71, 5053 Bergen, Norway

max.korbmacher@helse-bergen.no

Running title: Multiverse multiple sclerosis stratification

Abstract

The underlying mechanisms for neurodegeneration in multiple sclerosis are complex and incompletely understood. Multivariate and multimodal investigations integrating demographic, clinical, multi-omics, and neuroimaging data provide opportunities for nuanced analyses, aimed to define disease progression markers. We used data from a 12-year longitudinal cohort of 88 people with multiple sclerosis, to test the predictive value of multi-omics, MRI, clinical examinations, self-reports on quality of life, demographics, and general health-related variables for future functional and cognitive disability. Progressive functional loss beyond an Expanded Disability Status Scale score≥4 was used to define a functional loss group. A cognitive decline group was defined by a ≥25% decrease from the maximum (cognitive) Paced Auditory Serial Addition Test score. We used a multiverse approach to identify which baseline variables were most predictive for functional and cognitive loss group memberships, independent of analysis bias.

We identified several factors predicting an increased risk of future functional loss (FLG) and cognitive decline groups (CDG) within the next 12 years from baseline: functional score (0-10, median Odds Ratio per baseline unit increase [mOR_{FLG}=2.15±0.51; mOR_{CDG}=2.46±1.60]), cognitive scores (1-60 [mOR_{FLG}=0.98±0.03; mOR_{CDG}=0.91±0.06]), the number of previous relapses [mOR_{FLG}=1.56±0.26; mOR_{CDG}=1.44±0.60], serum vitamin A levels (umol/l [mOR_{FLG}=0.92±0.06; mOR_{CDG}=0.33±0.36]), self-reported mental health (1-100 [mOR_{FLG}=0.96±0.02; mOR_{CDG}=0.91±0.09]) and physical functioning (1-100 [mOR_{FLG}=0.99±0.01; mOR_{CDG}=0.97±0.03]). Our results suggest that clinical assessment of physical function and cognition, self-reported mental health, and potentially vitamin A levels are the best predictors for risk-group stratifications of people with MS at baseline. While these findings are promising, we also want to underscore the observed analysis-choice induced

variability which necessitates both an increase in transparency when reporting study findings as well as strategies which are robust to the many researcher degrees of freedom.

Keywords: Disease course prediction, multiverse, specification curve, patient-reported outcome measures

Introduction

Multiple sclerosis (MS) is a chronic immune-mediated disease of the central nervous system. Globally, there are more than 2.8 million people with MS (pwMS), with a higher prevalence in regions such as North America and Europe compared to Africa and East Asia.[1] Despite significant advancements in disease-modifying therapies over the last two decades, pwMS exhibit considerable heterogeneity in their disease progression and response to treatment.[2,3] While some pwMS develop severe disability despite high-efficacy disease-modifying therapies, others remain stable under similar regimens.[4] This variability underscores the importance of stratifying pwMS to identify distinct disease trajectories and improve treatment precision.

Disease trajectories can be defined by functional or cognitive decline, which can appear independent of each other.[5] A clinically meaningful definition of disease trajectories requires a multimodal approach that integrates clinical, imaging, and molecular markers.[6]

An advantage of a multivariate approach is the possibility of highlighting multiple variables associated with MS, both those assessed in routine clinical assessments and beyond.[7] For example, various brain imaging markers such as lesion count[8], lesion volume[9,10], and brain age[11] have been linked to disability. Brain age is an age estimate based on a comparison of a person's brain MRI with a large dataset from a healthy population.[12] Furthermore, several serum markers have been explored as prognostic markers in MS, including serum neurofilament light chain (NfL)[13], chitinase-3-like protein1 (CHI3L1)[14,15], vitamin D[16], and potentially other vitamins[17,18], and HLA-DRB1 carriership[16]. General health risk markers such as smoking[19] and high body mass index (BMI)[20] also contribute to MS-related outcomes. Clinical variables such as the relapse rate[21], Expanded Disability Status Scale (EDSS)[22], and cognitive performance, for example measured by the Paced Auditory Serial Addition Test (PASAT)[23], further enhance the understanding of disease-related impairment. Additionally, patient-reported outcome measures (PROMs)[24] can offer valuable insights for a more comprehensive assessment of disease impact.

We have established a unique cohort, containing longitudinal recordings for all the above variables. While the richness of such data provides insights from many different perspectives, it also presents analytical challenges, particularly the potential for bias introduced by researchers' subjective decisions in data (pre)processing, modelling, and interpretation. Study designs often highlight one or few predictors of MS outcomes, not accounting for multiple other predictors and their possible combinations. However, due to statistical dependencies, the selection of the variables included in such models influences the observed results, in addition to many other decision researchers make when processing and analysing data.[25]

To enhance reproducibility and minimize bias, robust and systematic analytical frameworks are essential.[25] Multiverse analyses, which explore multiple plausible analytical paths and evaluate their impact on results, offer a promising approach to address analysis bias.[26]

Relevant not only to MS research, variable selection remains an unsolved statistical issue[27] for which multiverse analyses provide an opportunity to decrease analysis-specific bias when presenting which variables are important for disease trajectories. In our study, we combine commonly used data from clinical follow-up and experimental biomarkers of MS derived from MRI, blood samples, clinical assessments and quality of life (QoL). We stratify pwMS into groups of cognitive and functional impairment and showcase variables predicting future trajectories already at baseline.

Materials and methods

Sample

The sample included 88 pwMS with relapsing-remitting MS (RRMS) participating in the omega-3 fatty acid in MS (OFAMS) multicentre clinical trial[28]. The trial entailed data collection every 6 months over 2 years, followed by a single follow-up visit 10 years after the original trial concluded. A total of 85 (97%) of the original clinical trial participants were included and considered in the present analyses. In addition to age, sex, BMI, smoking status, and ω-3 supplementation (original clinical study intervention), the collected data used in this study entailed assessments of clinical (EDSS)[22] and cognitive (PASAT)[23] performance, together with QoL, evaluated by the short form quality of life inventory (SF-36)[29], T_1- and T_2-weighted MRI and blood samples (see Supplemental Table 13 for overview over variables; Supplemental Figure 3 for variability of raw measures).

We also used a reference MRI sample (see Supplemental Table 1) to establish brain age models ($n_{training}$=58,317, $n_{validation}$=6,608, age span = 5-90), non-clinical repeatedly sampled data of three individuals across one year (minimum of 25 scans each, Bergen Breakfast Scanning Club [BBSC])[30], and a cross-sectional MS sample (n=748, mean age = 38.63±9.46 years) to validate the brain age model in a relevant clinical cohort, matched with healthy controls (n=751, mean age = 38.73±9.61).

Demographic, clinical, self-reported, multi-omics, and neuroimaging assessments

Age, sex, BMI, and smoking status (smoker/non-smoker assessed by both serum cotinine levels and self-reports) were general variables available for all participants. Clinical and cognitive scores included EDSS[22] and PASAT[23], and the number of relapses 12 months prior to study start. SF-36 indicated QoL in eight categories [physical functioning, role limitations due to physical health, role limitations due to emotional problems, energy/fatigue, emotional well-being, social functioning, pain, and general health] based on the National Multiple Sclerosis Society's scoring recommendations[31,32] with scores ranging from 0-100 (worst to best outcome) except for general health (1-5). Serum analyses for the current sample[13,33], and MRI acquisition and post-processing (cortical reconstruction and lesion segmentation) have been described previously[19]. MRI markers of interest included baseline brain age gap (BAG) in years (see next section) from T_1-weighted MRI data, as well as lesion count and volume (mm^3) from T_2-weighted MRI data. Serum markers included CHI3L1 (mg/ml), NfL (pg/ml), and vitamin A and E levels (in µmol/L) and vitamin D levels (in nmol/L), as well as HLA−DRB1 carriership (yes/no). Intra-class correlation coefficients were good (ICC > 0.82), apart from CHI3L1 (ICC ≈ 0), vitamin D levels (ICC = 0.56), and zero-variance variables (see Statistical Analyses).

Sample stratification

EDSS-based stratification

We first stratified pwMS into a functional loss group (FLG) and the rest of the participants into a functionally stable or improving group (FSIG). FLG (n=19, 22.4%) was defined by reaching significant disability (EDSS≥4) after the baseline assessment and latest at the 144-month (range: 128-156, M=141.93±5.83 months) follow-up, in addition to either a confirmed disability progression during the original study's 24 months with 6-month follow-ups (11/19), or an EDSS≥1 increase between 24-month and 144-month visits (16/19). EDSS ≥ 4 was previously defined as severe disability[22] and has been used as a threshold to identify pwMS who exhibit limitations in their activities of daily living[34]. For further quality control of the

grouping, we assessed whether there were clinical relapses present in a six-month proximity to each EDSS measurement, potentially explaining short-term EDSS increases at the time of measurement. This was only found for 3 out of 19 (15.8%) FLG members, for whom disability progression were confirmed independent of the event. Disability progression could be confirmed for 11 out of 19 (57.9%) FLG members during the first 24 months of evenly spaced 6-month visits, and 18 out of 19 FLG members presented an EDSS increase between visits at month 24 and month 144. This disability progression could not be confirmed at 144-month follow-up, as only one EDSS measurement was available (see Supplementary Note 1).

PASAT-based stratification

In a second, independent stratification, we defined a cognitive decline group (CDG) based on a 15-point decrease in PASAT score (corresponding to 25% of the maximum score of 60 points) either between baseline and 144-month follow-up or 24-month session and 144-month follow-up, and with the scores decreasing over time (n_{CDG}=11). In contrast, a cognitively stable or improving group (CSIG) was defined by smaller cognitive decline than in the CDG, within expected effects of aging, or even stable or improving PASAT scores. Additional quality assurance can be found in Supplementary Note 1.

Brain age

We developed a machine-learning based generalized additive model (allowing polynomials up to the 4th order) trained on 58,317 healthy controls' cortical thickness, volume and surface area averages across Desikan-Killany atlas[35] parcels to predict age. Such models allow sufficient flexibility to approximate non-linear age-dependencies, while still being simple enough to be interpretable.

Bias corrections for brain age prediction models are important as brain ages are otherwise systematically overestimated for lowest ages and underestimated for highest ages, respectively. While this is a know problem, previous studies do usually only apply post-hoc corrections, which do however not remove all of the bias. Hence, we used a combined approach: The original training sample (n=58,317) was up-sampled (to n=150,517) using SMOGN[36] to approximate a unimodal distribution to minimize training sample age-distribution bias. To account for remaining age-distribution bias, we applied a post-hoc correction using the training sample's linear age and brain age (BA) association's intercept (a) and slopes (b): $Corrected\ BA = BA + (Age - (a + b * BA))$. The correction procedure improved model performance (Supplemental Table 2). Finally, the brain age gap (BAG) was calculated as the

difference between corrected BA and chronological age: $BAG = Corrected\ BA - Age$. Our model predicted lower corrected BAGs in healthy controls compared to pwMS in the utilized cross-sectional sample ($t_{(1462)}$=-19.40, $p<2.2*10^{-16}$, Cohen's d=-1.00, [-1.11,-0.89]). Moreover, we validated our models in BBSC data, showing within-subject correlations of Pearson's r_{sub1}=0.25, r_{sub2}=0.23, r_{sub3}=0.53 between age and brain age, yet with only the last correlation being significant (p_{sub3}=0.006). Note that intra-class correlation coefficients could not be computed for these N=3 subjects due to the n-to-p ratio (small sample size, large number of observations). Additional information on the model validation procedure can be found at https://github.com/MaxKorbmacher/OFAMS_Brain_Age.

Statistical Analysis

Descriptives

We first examined the validity of the stratification using linear mixed effects regression models with random intercept at the level of the individual to assess interaction effects between age and risk grouping, while controlling for age and sex. Specifically, age-group interaction effects on EDSS were used to evaluate functional loss stratification, and age-group interaction effects on PASAT for cognitive decline stratification (Fig.1, Supplemental Figures 1-2).

To describe apparent differences at baseline, we analyzed univariate baseline differences between the respective risk vs stable groups using simple linear regression models, correcting for age and sex (Tables 1-2); and for differences in frequencies using Fisher's exact tests (Supplemental Tables 7-8).

Multiverse analyses

We used a multiverse approach to identify which baseline variables were most predictive of risk-membership using binary logistic regression (Fig. 2). Multiverse analyses have the aim to test the robustness of effects using different analytical choices. Multiple models are run, leading to distributions of coefficients, instead of a single estimate. Hence, instead of presenting single coefficients, we present the central tendency of coefficents (median) and variability (median absolute deviation).

We included models with all possible combinations of variables of interest up to the 8th order with age as a covariate included in all models. All models were run with and without sex as a covariate, leading to a total of 1,761,938 models. The QoL "role limitations due to emotional

problems" and "role limitations due to physical problems" were constant (mean=0±0) across participants and therefore excluded from the analyses.

McFadden's[37] pseudo R^2 (pR^2) was estimated for each regression model as a model fit proxy. Wald-based power/sensitivity analysis[38] was used to estimate the power of each model at $\alpha=0.05$ by testing the assumption of a greater probability of functional or cognitive decline group-membership when predictors≠0. For the EDSS-based FLG, model fit was good[1] (pR^2=11.35±5.00%; 60.69%>pR^2>10%) when considering all 803,390 converged models. The median power for these models was 32.90±45.98%. For the PASAT-based CDG the median model fit was R^2=37.19±19.00% (99.69%>pR^2>10%) considering the same converged models, presenting a medium power of 16.87±25.01%. Considering the low median power, we present various versions of multiverse analysis results, considering only well-powered or well-fitting models in the supplement. Multiverse analyses ought to inform about all models, whether well specified or not. This is useful when assessing the literature while assuming analytic flexibility and the possibility of mis-specified models, both of which cannot be excluded. Hence, we reported on all models in the main text but reported only models selected for high power and/or excellent model fit in the supplement. We report raw/uncorrected p-values and 95% Confidence Intervals (CI) in square brackets when describing group differences for the stratifications. For multiverse analyses, median and median absolute deviation (MAD) are reported for the estimated ORs and p-values of each variable of interest across the specified models in the multiverse to a) belong to FLG vs FSIG or b) CDG vs CSIG. Additionally, the proportion of the OR's directionality (PORSD), such as the number of models including EDSS, with EDSS OR>1 for risk group membership, as share of all models was reported. When effects presented consistent directionality in 75% or more of the models they were used in, they were presented in the main text, in addition to clinically meaningful effect sizes. Clinically meaningful was defined here as plausible and possible changes based on the assessed scales and distributions of values and their changes, as well as the magnitude of the effect. For example, a 0.001% increase in the odds of belonging to one of the defined risk-groups per year of age at baseline might be a significant effect but accumulates to less than 1% higher odds even when being older than 100 years of age at baseline. Hence, in this example, the effect of age would not be counted as clinically meaningful. Finally, due to the relatively limited sample size and yet limited missingness, to avoid biasing the results, we did not conduct imputations

[1] Note that pseudo variance explained is generally lower than conventional variance explained.[37]

of missing values. For baseline missingness of all utilized variables see Supplemental Figure 5, being relevant for the main analyses. EDSS and PASAT score missingness was relevant for stratification (see Supplementa tables 14-15).

Results

Study participants

The 85 pwMS with a 12-year follow-up were aged 38.9±8.3 (range: 19-58) years at baseline and 49.6±8.6 at the final follow-up, with 65.9% being females. Median baseline EDSS±MAD was 2.00±0.74, baseline PASAT was 47.9±9.4, and time since diagnosis was 1.9±3.2 years, and since the first symptom 5.5±5.5 years. Missingness was minimal ($n \leq 6$) for the baseline measures, which informed the main analyses (Supplemental Figure 5). Additional demographics and an overview of all variables can be found in the Supplemental Material.

Functional loss (disability) grouping

We show that, while not possible to stratify at baseline, FLG presented a stronger increase in EDSS than FSIG (Fig. 1), shown by a significant interaction effect between group membership and age on EDSS ($p=1.03*10^{-17}$), indicating a 0.14 unit faster annual increase in EDSS in the FLG group. At the same time, the groups were overall similar at baseline, with slightly worse physical function among FLG members (Table 1).

-- Figure 1 around here –

At baseline, FLG showed higher corrected EDSS=0.45 [0.02, 0.88] points ($t=2.07$, $p=0.0416$; $M_{FLG}=2.25\pm0.78$, $M_{FSIG}=1.65\pm0.61$). There were no other significant differences ($p>.05$, Table 1).

-- Table 1 around here –

Multiverse analysis of functional loss group membership

While no multiverse analysis derived *median* p-values were below 0.05, the median area under the curve (mAUC) was mAUC=0.72±0.05, and there were multiple effects which were on average relatively large, as well as unaffected in their directionality, independent of the analytic choices (Fig. 2, Supplemental Table 3). Clinically relevant and least analysis-choice dependent predictors of FLG included higher EDSS (0-10, mOR_{FLG}=2.15±0.51, p_{median}=.113, PORSD=100%), and QoL PROM physical functioning (mOR_{FLG}=0.99±0.01, p_{median}=0.483, PORSD=84.45%), HLA-DRB1 carriership (yes/no, mOR_{FLG}=1.57±0.72, p_{median}=0.508, PORSD=82.67%) and smoking (yes/no, mOR_{FLG}=1.35±0.28, p_{median}=0.670, PORSD=94.22%), which increased the odds of belonging to the functional loss group. Higher Vitamin E (umol/l, mOR_{FLG}=0.92±0.06, p_{median}=0.259, PORSD=99.99%) and A levels (umol/l, mOR_{FLG}=0.58±0.26, p_{median}=0.512, PORSD=90.26%) as well as higher CHI3L1-levels (mg/ml, mOR_{FLG}=0.96±0.01, p_{median}=0.126, PORSD=100%), NfL levels (pg/ml, mOR_{FLG}=0.99±0.01, p_{median}=0.629, PORSD=87.2%), PASAT score (1-60, mOR_{FLG}=0.98±0.03, p_{median}=0.565, PORSD=80.52%), relapses at baseline dating 1 year back in time (mOR_{FLG}=1.56±0.26, p_{median}=0.324, PORSD=99.99%), and the QoL PROMs (1-100) social functioning (mOR_{FLG}=1.02±0.03, p_{median}=0.430, PORSD=80.78%; overview of all effects: Supplemental Table 3). Importantly, these results did only change marginally when only including models with power≥80% (Supplemental Table 4), models with excellent model fit (pR^2≥0.20, Supplemental Table 5), or models with both excellent model fit and power≥80% (Supplemental Table 6).

-- Figure 2 around here --

Cognitive decline grouping

CDG presented a stronger decrease in PASAT than CSIG (Fig. 1, Supplemental Figure 2), shown by a significant ($p=3.83*10^{-5}$) interaction effect between group membership and age on PASAT, indicating a 0.80-unit faster annual decrease in PASAT score in CDG compared to CSIG. Moreover, CDG were on average 6.4±2.2 years older than CSIG, had a 3.8±1.6 points higher BMI, and 4.6±2.3 umpl/l higher vitamin E level at baseline after correcting for age and sex. There were no other significant differences (p>0.05, Table 2).

-- Table 2 around here ---

Cognitive decline multiverse analysis

Similar to the results for FLG, predictors of CDG membership were on average non-significant (median p>0.05), and mAUC=0.56±0.06. Yet multiple predictors had large median effects which were consistently of the same directionality (Fig. 2, Supplemental Table 10). Among those effects, higher CHI3L1 (mOR_{CDG}=1.26±0.17, p_{median}=0.077, PORSD=99.59%), higher EDSS (mOR_{CDG}=2.76±1.96, p_{median}=0.258, PORSD=94.56%), and a higher number of relapses (mOR_{CDG}=1.34±0.47, p_{median}=0.657, PORSD=79.50%) increased the odds of CDG membership, as also observed for FLG. Higher QoL PROMs for mental health (mOR_{CDG}=0.91±0.09, p_{median}=0.315, PORSD=94.79%) and physical functioning (mOR_{CDG}=0.97±0.03, p_{median}=0.293, PORSD=95.33%), higher PASAT score (mOR_{CDG}=0.93±0.06, p_{median}=0.301, PORSD=91.11%), and higher vitamin A (mOR_{CDG}=0.39±0.42, p_{median}=0.516, PORSD=82.09%) decreased the odds of CDG membership, again, similar to FLG membership odds.

Results unique to increased odds of CDG and not FLG membership included older age (mOR_{CDG}=1.10±0.05, p_{median}=0.219, PORSD=92.58%), higher BMI (mOR_{CDG}=1.62±0.51, p_{median}=0.052, PORSD=99.70%), larger lesion volume (cm^3, mORCDG=1.12±0.13, p_{median}=0.381, PORSD=92.15%). Female sex (mOR_{CDG}=0.43±0.43, p_{median}=0.535, PORSD=84.71%), omega-3 supplementation (mOR_{CDG}=0.17±0.19, p_{median}=0.220, PORSD=99.13%), higher vitamin D (mOR_{CDG}=1.26±0.17, p_{median}=0.077, PORSD=99.59%) and E levels (mOR_{CDG}=1.07±0.04, p_{median}=0.075, PORSD=99.77%) indicated lower odds of CDG membership. The results were robust to including only models with power≥80% (Supplemental Table 11), models with excellent model fit (pR^2≥0.20, Supplemental Table 12), or models with both excellent model fit and power≥80% (Supplemental Table 13).

Discussion

This study provides a comprehensive analysis of baseline variables associated with the risk and outcome of MS to assess their combined predictive ability for stratification of functional loss (EDSS) and cognitive decline (PASAT). Our results are presented in the light of multiple possible analytical pathways, supporting conclusions independent of more than a million specified analytical choices.

While our results highlight that multiple clinically relevant factors contribute to risk-group allocations, their individual significance, including p-value, consistency and magnitude of

effects, generally depended on the modelling choices. This highlights the usefulness of testing different analysis pipelines which includes varying control variables, variables of interest, and boundary conditions when assessing MS progression markers.

We identify several predictors of risk-group allocation which were the least dependent on analysis-choices. Such risk predictors of functional and cognitive loss included a higher EDSS, number of prior relapses, lower vitamin A levels, PASAT and self-reported mental health and physical functioning. Higher CHI3L1 at baseline decreased the odds of belonging to a risk-group.

Our analyses are based on the identification of favourable compared to unfavourable disease trajectories. Such trajectories can be evaluated in different ways, often consulting EDSS scores which focus on physical function. However, previous evidence showcases that pwMS with functional disability progression do not necessarily correspond with those with comparably accelerated cognitive decline.[39] The small overlap between the identified risk groups in our study confirmed this assumption. Moreover, quality controls ensured that the presented stratifications were robust to temporal fluctuations such as new relapses, new lesions or lesion activity, and we found clear differences in the trajectories when examining the variables which were stratified for.

Multiple studies have examined single or a small selection of MS markers.[14–18,22,24] Yet only few studies have focussed on multivariate assessments and particularly on longitudinal data.[40–43] Multimodal approaches have the advantage of enhancing predictions, for example as previously shown by combining electronic health records, clinical notes and neuroimaging data to predict MS severity[44], conversion from clinically isolated syndrome to MS,[45] or, similar to this study, to predict EDSS development.[46] However, just as conventional statistical approaches, machine and deep learning techniques are influenced by the input variable selection and hyperparameter tuning suffers from generally small samples in the MS field.

The utilized input variables in our multivariate approaches have previously, often univariately, been associated with different MS outcomes. For example, higher EDSS scores[22], HLA-DRB1 carriership[16], lower vitamin levels[16–18], higher lesion count[8], cigarette smoking[19], a lower PASAT test score[23], several PROMs[24], and a higher number of baseline relapses[21] have previously been associated with functional loss in MS, similar to elevated CHI3L1 and NfL levels[13–15]. While we found that higher CHI3L1 levels indicated higher odds of disability progression, serum CHI3L1 were reported to be influenced by third variables such as exercise

intensity[47] and CHI3L1 serum measures were unstable in our sample (ICC≈0, Supplemental Figure 3). Hence, a meaningful interpretation of our finding of higher CHI3L1 levels at baseline predicting less disability-progression at the 12-year follow-up not involving large uncertainty was not possible. Cerebrospinal fluid-derived CHI3L1[14,15] might be a more reliable measure. Additionally, our sample was relatively small and difficult to compare to more recent studies as patients were not treated during the first 6 months of the clinical trial. Hence, before drawing further inference, the influence of the characteristics of the examined limited sample and individual differences in disease trajectories need to be ruled out first by replicating of the findings reported here. Moreover, the statistical significance of the mentioned effects might depend on the selection of covariates. For example, the negative effect of cigarette smoking on health outcomes is generally known and has previously been documented in MS.[19] However, to contribute to an acceleration of the disease or, for instance, functional or cognitive decline, other cumulative effects might need to be considered.

PROMs and vitamin levels are yet underrepresented among variables informing MS assessments. Yet, PROMs have been suggested as a cost-effective and valid tool to inform on the long-term impact of MS.[48] Considering multiple inventories and variability in study designs, systematic investigations of the value of PROMs, specifically certain dimensions, to inform and establish predictive models are necessary. Similarly, vitamin levels beyond vitamin D are not commonly measured in MS care. Yet, as vitamins A and D present a broad influence on the immune system, baseline vitamin levels might indirectly reflect immune-related processes, where for example a lower serum level might indicate a larger metabolic need and hence an activated immune system.[49,50] The predictive value of vitamin levels on risk group membership might be confounded by nutritional health. While vitamin D deficiency has been linked with a higher risk of developing MS,[16,33] for example, vitamin D supplementation does not seem to significantly affect clinical outcomes in pwMS.[51] Such associations are not established for other vitamin serum concentrations.

The main strengths of the study were the unique dataset, with near complete follow-up (95%) for over 12 years, in addition to a large brain age training set of more than 60,000 healthy subjects' MRI data, and multiple validation sets, including hundreds of cross-sectionally scanned individuals. Combining these data allowed for more comprehensive analyses than previously conducted. Moreover, our analyses represent a good example for the potential of data-reusage. We were able to apply a new multiverse approach to the data, showcasing the influence of analysis pipelines on research outcomes.

Our study faces some limitations. First, several data-related issues limit the generalizability of the results. The size of the risk groups was relatively small, and treatment adjusted to the individual, including medication changes over time reduce the comparability of pwMS exacerbating the observation of true population estimates. Moreover, as common with automatic segmentation tools of MRI data, the segmentations were imperfect. MRI data processing challenges beyond the data processing software are scanner hardware and software changes over the acquisition period which could not be controlled for in the present national multi-centre study.

Machine learning implementations such as the utilized brain age model can "learn" and account for variability from unwanted covariates. Hence, the brain age predictions, which were trained on diverse, multi-site, software and field strength data, were meaningfully related to the examined developmental trajectories. The utilized brain age model is freely available together with a detailed user guide to enhance dissemination (https://github.com/MaxKorbmacher/OFAMS_Brain_Age).

The examined results were stable, even when selecting for only for sufficiently powered or well-fitting models, despite the lack of control over model misspecifications. Misspecifications might have resulted from thousands of probed modeling choices, including, for instance, not accounting for multicollinearity and non-linear relationships. On the other hand, overall model performance might have been weak, as in the case of low area under the curve for cognitive decline predictions, with yet single predictors containing useful information, indicated by consistent and strong effect sizes which compare with functional loss group classifications. Additionally, perfect model specifications were not the goal. Instead, we aimed to account for researcher bias, such as executing multiple but only reporting the significant analyses, while also testing multiple variables of interest in the same sample. This approach allows to closer approximate the reality of the literature than only models showing perfect fit.

We analysed a broad range of established and experimental MS markers. However, the selection of variables of interest in this study is not exhaustive and there might be more recently established markers which are more suitable for risk group stratifications, such as novel, strong risk loci as DNM3–PIGC or DYSF–ZNF638[52], giving additional inside into the heritability of MS.

The utilized brain age model focusses on cortical grey matter. While such approach improves the interpretability of the model, other approaches including subcortical grey matter[12] or

information on white matter microstructure[53] might be more sensitive to MS-specific degeneration.

Overall, we show that pwMS can be stratified into functional and cognitive decline risk groups. Predictors for such group stratifications, are dependent on analytic choices. The strongest and most consistent effects were found for the number of prior relapses, EDSS, self-reported mental health, PASAT and potentially vitamin A. These findings offer not only additional insight into MS disease trajectories, but also a new perspective on analysing multiple MS markers. We encourage the MS research community to adopt the proposed multiverse approach to mitigate the influence of analysis pipelines on research outcomes and underline the potential for multivariate analyses. Additionally, we underscore the importance of incorporating disease stratifications, particularly using longitudinal data, as critical steps for advancing future studies.

Data availability

Brain age model training data are available from the respective websites of the databases either openly or after application. OFAMS data can be shared after receiving a new ethics approval. Brain age models and analysis code are freely available at https://github.com/MaxKorbmacher/OFAMS_Brain_Age.

Acknowledgements

We want to thank Nils Inge Landrø for the helpful comments and recommendations to improve this study.

We extend our sincere gratitude to the people with MS who participated in this study, whose contributions are invaluable. This research originated as a national intervention study (OFAMS) and ten years follow-up examinations. We thank the OFAMS study group at numerous hospitals across Norway.

Moreover, we want to extend our gratitude to the thousands of scanned participants in the numerous studies which served for establishing the brain age models, and the study facilitators.

Funding

This project was funded by the Norwegian MS-union (no reference). Model training was performed on the Service for Sensitive Data (TSD) platform, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo IT-Department (USIT). Computations were performed using resources provided by UNINETT Sigma2

(#NS9666S) – the National Infrastructure for High Performance Computing and Data Storage in Norway, supported by the Norwegian Research Council (#223273).

Competing interests

OAA has received a speaker's honorarium from Lundbeck, Janssen, Otsuka and Lilly, and is a consultant to Coretechs.ai and Precision Health.

LTW is a minor shareholder of baba.vision.

KMM has served on scientific advisory board for Alexion, received speaker honoraria from Biogen, Novartis, Roche and Sanofi, and has participated in clinical trials organized by Biogen, Merck, Novartis, Otivio, Roche and Sanofi.

EAH received honoraria for advisory board activity from Sanofi-Genzyme, and his department has received honoraria for lecturing from Biogen and Merck.

ØT received speaker honoraria from and served on scientific advisory boards of Biogen, Sanofi-Aventis, Merck, and Novartis, and has participated in clinical trials organized by Merck, Novartis, Roche and Sanofi.

The remaining authors declare no other competing interests.

Contributors

Conceptualisation and design of the study: MK. Drafting the manuscript: MK. Statistical analyses: MK. Acquisition and analysis of data: All authors contributed in different capacities to data acquisition, administration, or analyses. Interpretation of data: MK. Critical revision of the article: all authors. Responsible for the overall content: MK. All authors approved the final version of the manuscript.

Ethics Approvals

The study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REK, 814351). The OFAMS-study and the 10-year follow-up were previously approved by REK (2016/1906) and registered as clinical trial (clinicaltrials.gov identifier: NCT00360906).

Ethical approval for the different brain age training datasets was obtained (REK 567301, PVO 17/21624), as well as for the longitudinal validation set (Bergen Breakfast Scanning Club, REK 238310), and the cross-sectional MS data (REK 2011/1846, REK 2016/102).

All participants gave their written informed consent according to the Declaration of Helsinki.

References

1. Walton, C. *et al.* Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Mult Scler* **26**, 1816–1821 (2020).
2. Jakimovski, D. *et al.* Multiple sclerosis. *The Lancet* **403**, 183–202 (2024).
3. Kuhlmann, T. *et al.* Multiple sclerosis progression: time for a new mechanism-driven framework. *The Lancet Neurology* **22**, 78–88 (2023).
4. Figueira, G. M. A., Soares, P. V., da Silveira, R. C. & Figueira, F. F. A. "Stable" vs. "silent progressive multiple sclerosis": a real-world retrospective clinical imaging Brazilian study. *Arq Neuropsiquiatr* **80**, 405–409.
5. Morrow, S. A. *et al.* Detecting isolated cognitive relapses in persons with MS. *Mult Scler* **29**, 1786–1794 (2023).
6. Andorra, M. *et al.* Predicting disease severity in multiple sclerosis using multimodal data and machine learning. *J Neurol* **271**, 1133–1149 (2024).
7. Thompson, A. J. *et al.* Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology* **17**, 162–173 (2018).
8. Tedeschi, G. *et al.* Brain atrophy and lesion load in a large population of patients with multiple sclerosis. *Neurology* **65**, 280–285 (2005).
9. Dwyer, M. G. *et al.* Atrophied Brain Lesion Volume: A New Imaging Biomarker in Multiple Sclerosis. *Journal of Neuroimaging* **28**, 490–495 (2018).
10. Genovese, A. V. *et al.* Atrophied Brain T2 Lesion Volume at MRI Is Associated with Disability Progression and Conversion to Secondary Progressive Multiple Sclerosis. *Radiology* **293**, 424–433 (2019).
11. Høgestøl, E. A. *et al.* Cross-Sectional and Longitudinal MRI Brain Scans Reveal Accelerated Brain Aging in Multiple Sclerosis. *Front Neurol* **10**, 450 (2019).

12. Korbmacher, M. *et al.* Brain asymmetries from mid-to late life and hemispheric brain age. *Nature Communications* **15**, 956 (2024).
13. Varhaug, K. N. *et al.* Neurofilament light chain predicts disease activity in relapsing-remitting MS. *Neurology Neuroimmunology & Neuroinflammation* **5**, e422 (2018).
14. Lucchini, M. *et al.* CSF CXCL13 and Chitinase 3-like-1 Levels Predict Disease Course in Relapsing Multiple Sclerosis. *Mol Neurobiol* **60**, 36–50 (2023).
15. Talaat, F., Abdelatty, S., Ragaie, C. & Dahshan, A. Chitinase-3-like 1-protein in CSF: a novel biomarker for progression in patients with multiple sclerosis. *Neurol Sci* **44**, 3243–3252 (2023).
16. Wergeland, S. *et al.* Vitamin D, HLA-DRB1 and Epstein-Barr virus antibody levels in a prospective cohort of multiple sclerosis patients. *Eur J Neurol* **23**, 1064–1070 (2016).
17. Zhang, S. M. *et al.* Intakes of carotenoids, vitamin C, and vitamin E and MS risk among two large cohorts of women. *Neurology* **57**, 75–80 (2001).
18. Fragoso, Y. D., Stoney, P. N. & McCaffery, P. J. The Evidence for a Beneficial Role of Vitamin A in Multiple Sclerosis. *CNS Drugs* **28**, 291–299 (2014).
19. Lie, I. A. *et al.* The Effect of Smoking on Long-term Gray Matter Atrophy and Clinical Disability in Patients with Relapsing-Remitting Multiple Sclerosis. *Neurology Neuroimmunology & Neuroinflammation* **9**, e200008 (2022).
20. Manuel Escobar, J. *et al.* Body mass index as a predictor of MS activity and progression among participants in BENEFIT. *Mult Scler* **28**, 1277–1285 (2022).
21. Bennetto, L. *et al.* The relationship between relapse, impairment and disability in multiple sclerosis. *Mult Scler* **17**, 1218–1224 (2011).
22. Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis. *Neurology* **33**, 1444–1444 (1983).

23. Rao, S. A manual for the brief repeatable battery of neuropsychological tests in multiple sclerosis. (1990).
24. D'Amico, E., Haase, R. & Ziemssen, T. Review: Patient-reported outcomes in multiple sclerosis care. *Multiple Sclerosis and Related Disorders* **33**, 61–66 (2019).
25. Korbmacher, M. *et al.* The replication crisis has led to positive structural, procedural, and community changes. *Commun Psychol* **1**, 1–13 (2023).
26. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. *Perspect Psychol Sci* **11**, 702–712 (2016).
27. George, E. I. The Variable Selection Problem. *Journal of the American Statistical Association* **95**, 1304–1308 (2000).
28. Torkildsen, O. *et al.* ω-3 fatty acid treatment in multiple sclerosis (OFAMS Study): a randomized, double-blind, placebo-controlled trial. *Arch Neurol* **69**, 1044–1051 (2012).
29. Ware, J. E. & Sherbourne, C. D. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* **30**, 473–483 (1992).
30. Wang, M.-Y., Korbmacher, M., Eikeland, R. & Specht, K. Deep brain imaging of three participants across 1 year: The Bergen breakfast scanning club project. *Front. Hum. Neurosci.* **16**, (2022).
31. National Multiple Sclerosis Society. Multiple Sclerosis Quality of Life Inventory: A User's Manual. (1997).
32. Stewart, A. L., Hays, R. D. & Ware, J. E. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* **26**, 724–735 (1988).
33. Wesnes, K. *et al.* Low vitamin D, but not tobacco use or high BMI, is associated with long-term disability progression in multiple sclerosis. *Multiple Sclerosis and Related Disorders* **50**, (2021).

34. Ruggieri, R. M. *et al.* Cognitive impairment in patients suffering from relapsing-remitting multiple sclerosis with EDSS $\leq$ 3.5. *Acta Neurologica Scandinavica* **108**, 323–326 (2003).
35. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
36. Branco, P., Torgo, L. & Ribeiro, R. P. SMOGN: a Pre-processing Approach for Imbalanced Regression.
37. McFadden, D. Conditional logit analysis of qualitative choice behavior. (1972).
38. Demidenko, E. Sample size determination for logistic regression revisited. *Statistics in Medicine* **26**, 3385–3397 (2007).
39. Fuchs, T. A. *et al.* Cognitive progression independent of relapse in multiple sclerosis. *Mult Scler* **30**, 1468–1478 (2024).
40. Labiano-Fontcuberta, A. *et al.* Predictive models of multiple sclerosis-related cognitive performance using routine clinical practice predictors. *Multiple Sclerosis and Related Disorders* **76**, (2023).
41. Lopez-Soley, E. *et al.* Dynamics and Predictors of Cognitive Impairment along the Disease Course in Multiple Sclerosis. *Journal of Personalized Medicine* **11**, 1107 (2021).
42. Al-iedani, O. *et al.* Multi-modal neuroimaging signatures predict cognitive decline in multiple sclerosis: A 5-year longitudinal study. *Multiple Sclerosis and Related Disorders* **81**, (2024).
43. Bsteh, G. *et al.* Long Term Clinical Prognostic Factors in Relapsing-Remitting Multiple Sclerosis: Insights from a 10-Year Observational Study. *PLOS ONE* **11**, e0158978 (2016).

44. Zhang, K., Lincoln, J. A., Jiang, X., Bernstam, E. V. & Shams, S. Predicting multiple sclerosis severity with multimodal deep neural networks. *BMC Medical Informatics and Decision Making* **23**, 255 (2023).
45. Rasouli, S. *et al.* Predicting the conversion from clinically isolated syndrome to multiple sclerosis: An explainable machine learning approach. *Multiple Sclerosis and Related Disorders* **86**, 105614 (2024).
46. Campanioni, S. *et al.* Explainable machine learning on baseline MRI predicts multiple sclerosis trajectory descriptors. *PLoS One* **19**, e0306999 (2024).
47. Bilek, F., Ercan, Z., Deniz, G., Ozgul, S. & Demir, C. F. High-intensity intermittent exercise increases serum levels of chitinase 3-like protein-1 and matrix metalloproteinase-9 in persons with multiple sclerosis. *J Neuroimmunol* **395**, 578434 (2024).
48. Lerede, A., Rodgers, J., Middleton, R. M., Hampshire, A. & Nicholas, R. Patient-reported outcomes in multiple sclerosis: a prospective registry cohort study. *Brain Commun* **5**, fcad199 (2023).
49. Mora, J. R., Iwata, M. & von Andrian, U. H. Vitamin effects on the immune system: vitamins A and D take centre stage. *Nat Rev Immunol* **8**, 685–698 (2008).
50. Newsholme, P. Cellular and metabolic mechanisms of nutrient actions in immune function. *Eur J Clin Nutr* **75**, 1328–1331 (2021).
51. Mahler, J. V. *et al.* Vitamin D3 as an add-on treatment for multiple sclerosis: A systematic review and meta-analysis of randomized controlled trials. *Mult Scler Relat Disord* **82**, 105433 (2024).
52. Harroud, A. *et al.* Locus for severity implicates CNS resilience in progression of multiple sclerosis. *Nature* **619**, 323–331 (2023).

53. Korbmacher, M. *et al.* Brain-wide associations between white matter and age highlight the role of fornix microstructure in brain ageing. *HBM* **44**, (2023).

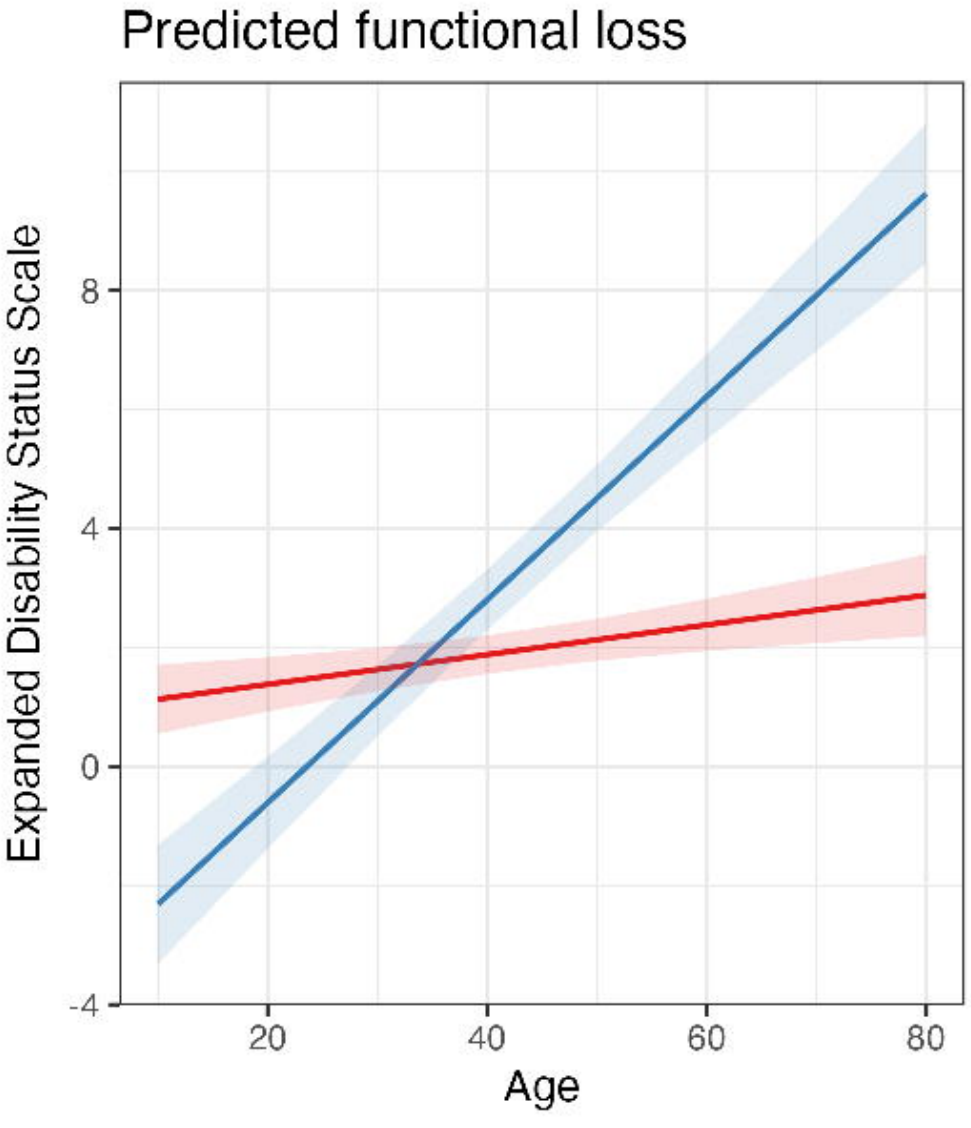
Figure legends

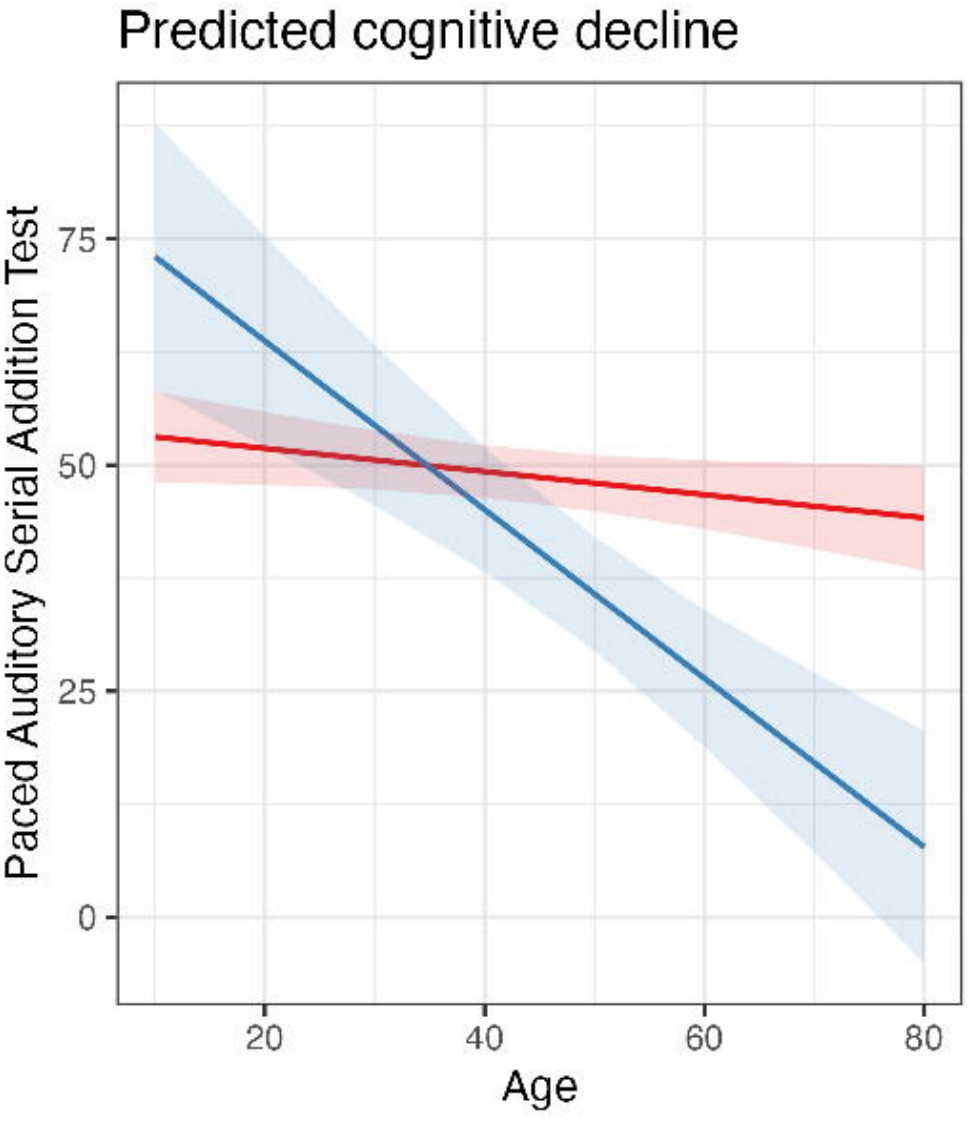
Figure 1. Sample stratification into functional and cognitive decline groups. Marginal interaction effects between group and age in random intercept models at the level of the individual indicate that risk groups (blue) show a faster functional and cognitive loss than functionally and cognitively stable or improving groups (red). Note that the stratifications were done independent from each other, resulting in n_{FLG}=19 and n_{CDG}=11, with only n=3 participants being allocated in both groups.

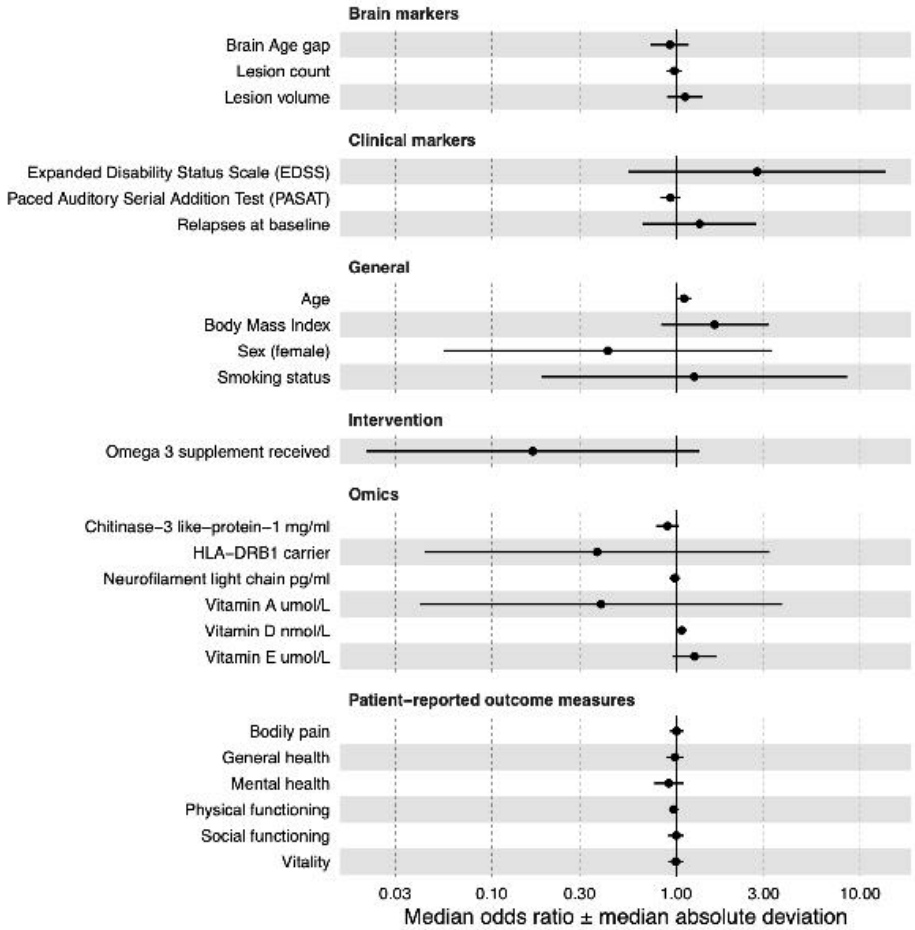
Figure 2. Multiverse estimates of predictors of functional (left) and cognitive decline (right). Note that while all median p-values>0.05, multiple effects were consistently unidirectional. For log OR distributions see Supplemental Figure 4.

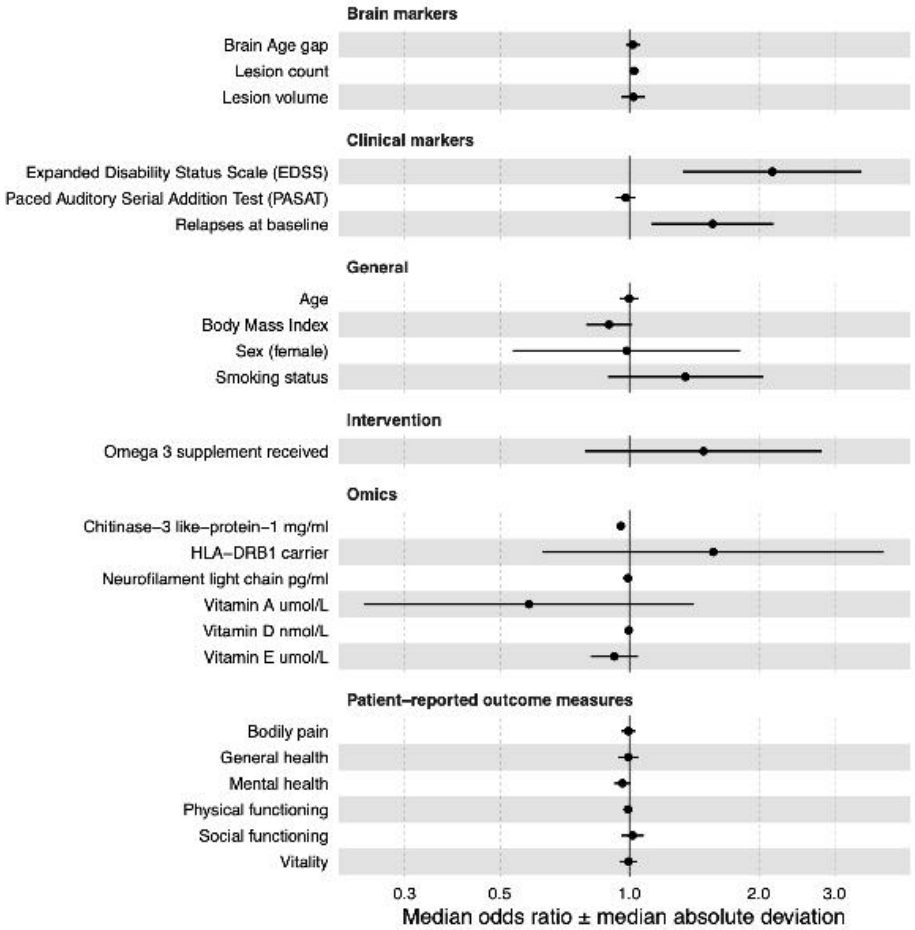
Supplemental Material

'Supplementary material is available at *Brain* online'


Predicted functional loss
Expanded Disability Status Scale
8
4
0
-4
20
40
60
80
Age


Predicted cognitive decline
Paced Auditory Serial Addition Test
75
50
25
0
20
40
60
80
Age


Brain markers
Brain Age gap
Lesion count
Lesion volume
Clinical markers
Expanded Disability Status Scale (EDSS)
Paced Auditory Serial Addition Test (PASAT)
Relapses at baseline
General
Age
Body Mass Index
Sex (female)
Smoking status
Intervention
Omega 3 supplement received
Omics
Chitinase−3 like−protein−1 mg/ml
HLA−DRB1 carrier
Neurofilament light chain pg/ml
Vitamin A umol/L
Vitamin D nmol/L
Vitamin E umol/L
Patient−reported outcome measures
Bodily pain
General health
Mental health
Physical functioning
Social functioning
Vitality
0.03
0.10
0.30
1.00
3.00
10.00
Median odds ratio ± median absolute deviation


Brain markers
Brain Age gap
Lesion count
Lesion volume
Clinical markers
Expanded Disability Status Scale (EDSS)
Paced Auditory Serial Addition Test (PASAT)
Relapses at baseline
General
Age
Body Mass Index
Sex (female)
Smoking status
Intervention
Omega 3 supplement received
Omics
Chitinase−3 like−protein−1 mg/ml
HLA−DRB1 carrier
Neurofilament light chain pg/ml
Vitamin A umol/L
Vitamin D nmol/L
Vitamin E umol/L
Patient−reported outcome measures
Bodily pain
General health
Mental health
Physical functioning
Social functioning
Vitality
0.3
0.5
1.0
2.0
3.0
Median odds ratio ± median absolute deviation