

## **Physical activity monitoring to assess disability progression in multiple sclerosis**

*Charlotte M. Stuart<sup>1</sup>, Aravinthan Varatharaj<sup>1,2</sup>, Janine Domjan<sup>2</sup>, Sheaba Philip<sup>1</sup>, SIMS study group\*,  
Ian Galea<sup>1,2</sup>*

<sup>1</sup> Clinical Neurosciences, Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, UK

<sup>2</sup> Wessex Neurological Centre, University Hospital Southampton NHS Foundation Trust, UK

Key words: Multiple sclerosis, progressive multiple sclerosis, remote physical activity monitoring, accelerometer, wearable electronic devices, teleneurology

\* SIMS study group: Ellen Adams, Raeisa Ali, Sarah Cader, Sheila Chartres, Faye Cornick, Angela Darekar, Janine Domjan, Monica Fenn, Julian Furby, Ian Galea, Remi Guillochon, Chris Halfpenny, Charles Hillier, Roanne Hyde, Carmen Jacob, Elizabeth Jarman, Mayada Nabag, Aimee O'Neill, V Hugh Perry, Sheaba Philip, Chloe Sayce, Aginor Spanoulis, Charlotte M. Stuart, Connie Temple-Brown, Aravinthan Varatharaj and Elina Zotova.

Corresponding Author:

Dr Ian Galea, Associate Professor,

Clinical Neurosciences, Clinical & Experimental Sciences, Faculty of Medicine, University of

Southampton, Mailpoint 806, Level D, Southampton General Hospital, Southampton, SO16 6YD, UK

Email: [I.Galea@soton.ac.uk](mailto:I.Galea@soton.ac.uk)

Twitter: @SIMSstudy

## Abstract

**Background:** Clinical outcome measurement in multiple sclerosis (MS) usually requires a physical visit. Remote activity monitoring (RAM) using wearable technology provides a rational alternative, especially desirable when distance is involved or in a pandemic setting.

**Objective:** To validate RAM in progressive MS using (1) traditional psychometric methods (2) brain atrophy.

**Methods:** 56 people with progressive MS participated in a longitudinal study over 2.5 years. An arm-worn RAM device measured activity over six days, every six months, and incorporated triaxial accelerometry and transcutaneous physiological variable measurement. Five RAM variables were assessed: physical activity duration, step count, active energy expenditure, metabolic equivalents and a composite RAM score incorporating all four variables. Other assessments every six months included EDSS, MSFC, MSIS-29, Chalder Fatigue Scale and Beck's Depression Inventory. Annualized brain atrophy was measured using SIENA.

**Results:** RAM was tolerated well by people with MS; the device was worn 99.4% of the time. RAM had good convergent and divergent validity and was responsive, especially with respect to step count. Measurement of physical activity over one day was as responsive as six days. The composite RAM score positively correlated with brain volume loss.

**Conclusion:** Remote activity monitoring is a valid and acceptable outcome measure in MS.

## Introduction

Monitoring progression in Multiple Sclerosis (MS) is essential for clinical management and research. The most widely used measure is the Expanded Disability Status Scale (EDSS), a 10-point ordinal composite scale that categorises disability based on a number of functional systems, predominantly ambulatory function<sup>1</sup>. It places a lot of emphasis on lower limb physical function, with under-representation of other important functional domains, such as upper limb function, cognition, mood and fatigue. The EDSS struggles to capture progression at the higher end of the scale and has low reproducibility at the lower end<sup>2</sup>. The Multiple Sclerosis Functional Composite (MSFC) is increasingly being used as an outcome measure in MS studies; it provides an improved clinical measure of disability by combining ambulation (the timed 25-foot walk, T25FW), upper limb function (a timed 9-hole peg test, 9HPT) and cognitive function (the Paced Auditory Serial Addition Test, PASAT)<sup>3</sup>.

Clinic-based tests such as the EDSS and MSFC provide a single snapshot of disability at one point in time and in an artificial environment. People with MS may struggle to access clinical services because of travel costs, restricted mobility, or fatigue. Recently the COVID-19 pandemic highlighted the need for remote assessments. A tele-EDSS assessment has been shown to be feasible in people with mild-to-moderate MS capable of guided self-assessment<sup>4</sup>, however assessment strategies are still lacking for more disabled patients. A more convenient way to gain insight into the effects of MS on daily living is to use patient-reported outcomes such as the MS impact scale (MSIS-29)<sup>5</sup>, however questionnaires can suffer from subjective interpretation and recall bias. There is therefore the need for an accessible, objective outcome measure that can monitor disability within a patient's living environment. Remote activity monitoring (RAM) using wearable accelerometer devices has been shown to be feasible and to provide clinically meaningful information in MS and other neurological diseases<sup>6-9</sup>. The growing importance of serial assessments to monitor progression and the rise of telemedicine have accentuated the unmet need to validate the use of RAM in clinical settings. Hence, we set out to perform a robust validation of RAM, using an arm-worn multi-sensor device as

an exemplar. We assess its measurement qualities in progressive MS, alongside three well-established clinical scales (EDSS, MSFC and MSIS-29) using traditional psychometric methods. We also validate RAM against a biological readout for progression, namely brain atrophy on magnetic resonance imaging (MRI). While T2 lesion volume increase is more informative in relapsing-remitting and early progressive MS, brain atrophy is correlated with disability worsening across the entire MS course<sup>10 11</sup>.

## **Patients and Methods**

### **Participants**

This study included adult participants (> 18 years) with primary progressive (PPMS, n=33) and secondary progressive (SPMS, n=24) recruited to a sub-study on remote accelerometry within the Systemic Inflammation in Multiple Sclerosis (SIMS) study at the Wessex Neurological Centre, University Hospital Southampton NHS Foundation Trust. The project was covered by National Research Ethics Service Approval 12 SC 0176 and institutional research ethics approval ERGO 5562. Inclusion criteria were: (1) diagnosis of progressive MS according to the 2010 McDonald criteria<sup>12</sup> (2) age ≤70 (3) EDSS ≤6.5. Exclusion criteria were: (1) relapses in the last year (2) disease-modifying or immunosuppressive treatment in the previous six months (3) comorbidities that could contribute to neurological disability.

### **Clinical measures and protocol**

During the first visit, participants were shown how to correctly apply and remove the device to the back of their upper dominant arm (over the triceps). They were instructed to wear the device for twenty-four hours a day, with the exception of swimming or bathing. Participants started wearing the RAM device at a weekday clinic visit and returned the following week. Previous studies had shown that anything between two and seven days was needed to monitor physical activity reliably with this device<sup>13 14</sup>. Clinics were held on Tuesdays and Wednesdays, so the minimum wear-period

was six full days, and always included the weekend. A RAM measurement week was considered valid if, on average, the device was worn for a minimum of 23 hours a day over a continuous six-day wear period, independent of number of steps per day. A week after the first visit, participants returned to the clinic for data download, height and weight measurement, completion of questionnaires (including the MSIS-29, Chalder Fatigue Scale <sup>15</sup> and Beck Depression Inventory Short Form (BDI) <sup>16</sup>, and MSFC and EDSS assessments performed by trained and certified research staff. Follow-up duration was two and a half years, with a pair of such clinic visits every six months (six assessments in total, Figure 1A).

### **Remote physical activity measurement**

The RAM device (SenseWear Armband, Model MF-SW, Body Media, Pittsburgh, PA), shown in Figure 1B, weighed 45g and is worn on the arm with an adjustable strap. Two stainless steel arrays on the back house multiple sensors that measure heat flux (heat dissipated by the body), galvanic skin response (the conductivity of the skin, varying according to physical and emotional stimuli) and skin temperature. A tri-axial accelerometer measures free-living daily physical activity including upper and lower limb motion. Data from these sensors is processed by an algorithm to provide the four variables studied here, namely daily step count, metabolic equivalents (METs), physical activity duration (PhAD) and active energy expenditure ( $E_A$ ). PhAD and  $E_A$  were quantified for physical activity above 3.0 METs (equivalent to walking). While the Sensewear Armband has been used for physical activity outcome measurement in progressive MS patients<sup>17 18</sup>, it has not been validated for detection of progression in this patient group. Sensewear Armband  $E_A$  measurement has been validated against indirect calorimetry in healthy populations and shown to be particularly accurate for low intensity activity<sup>19 20</sup>.

### **Brain MRI**

MRI was conducted at baseline and study exit on a 3T MR unit (Skyra, Siemens, Erlangen, Germany) at University Hospital Southampton NHS Foundation Trust, using a 20-element phased-array head coil. Due to the long duration of the study, scans at entrance and exit were naturally interleaved. Structural images were acquired isotropically and used a 3D magnetization prepared - rapid gradient echo (MP-RAGE) sequence with the following parameters: TR = 2200 ms, TE = 2.45 ms, TI = 900 ms, flip angle = 8°, GRAPPA under-sampling with parallel imaging factor = 2, field-of-view 250 × 250 × 176 mm<sup>3</sup>, voxel size 1.0 × 1.0 × 1.0 mm<sup>3</sup>. This sequence has excellent grey-white matter contrast, suitable for volume and atrophy estimation and tissue classification<sup>21</sup>. All raw images were visually inspected by an experienced neuroradiologist; subjects with inadequate image quality (e.g. due to movement artefact) were recalled for further imaging. Two time point percentage brain volume change was estimated with SIENA<sup>22</sup>, part of FSL<sup>23</sup>, and then annualised. Standard-space masking was applied to improve removal of non-brain tissue<sup>24</sup>, and brain extraction settings were optimised according to published recommendations<sup>25</sup>. The intermediate results of brain extraction, image registration (and segmentation) were visually examined in all cases, blinded to the result of analysis, and where appropriate manual corrections applied.

#### **RAM variable definitions, analysis and statistics**

Usability was assessed by considering patient feedback, adverse reactions, and tolerance (defined as total wear time as measured by the device, the effect of disability on wear time, and change in wear time within measurement weeks and as the study progressed). SenseWear® software (Version 7.0) was used to analyse RAM data. RAM variables were normalized to wear duration to obtain a single combined metric of physical activity, a RAM composite score was created by combining the four variables: step count, PhAD, METs and E<sub>A</sub>. Since the units of measurement differ across these variables, raw scores for each variable were converted to z-scores. This was after normalization of step count, E<sub>A</sub> and PhAD using the cube root transformation (logarithmic transformation was insufficient for normalising the data). One study participant who was a landscape

gardener with EDSS=1.5, with very high activity and no progression on the study, was excluded from the baseline z-score calculations. The RAM score was constructed by using the simple arithmetic mean of the component variable z-scores i.e. using an equal weight for each. Correlation of individual RAM components with the overall RAM score supported its construct validity (Supplementary Table 1). MSFC component z scores were calculated relative to the mean baseline of the whole study. The physical (questions 1-20) and psychological (questions 21–29) components of the MSIS-29 are referred to here as MSIS-phys and MSIS-psych respectively.

SPSS v25 was used for statistical analyses and plots were created in GraphPad Prism v8. Results are presented for all progressive MS participants as one group, and for the separate PPMS and SPMS groups. Data distribution was determined graphically and using the Kolmogorov-Smirnov test. Responsiveness was assessed in two ways: (1) statistical significance of the change over time, using the Wilcoxon matched pair signed rank tests, comparing paired measurements separated in time (2) the effect size of the change was calculated as the Wilcoxon Z divided by the square root of the number of observations  $(Z/\sqrt{n})^{26}$ . The progression detection rate was defined as the percentage of epochs with significant responsiveness to change in the variable. The minimum wear period was defined as the smallest number of monitoring days needed to detect significant change. Slopes were calculated as  $dy/dt$ , where  $y$ =RAM or other variable,  $t$ =time. Rate of change in slopes was calculated as  $d^2y/dt^2$ , where  $y$ =RAM or other variable,  $t$ =time. All outcome variables were converted to z-scores for slope analysis. Correlations were assessed using Spearman's or Pearson's tests as appropriate. All hypothesis testing was two-tailed. A p value < 0.05 was considered significant. Mean  $\pm$ SD is quoted unless otherwise stated.

## Results

### Demographics

The CONSORT diagram is shown in Figure 2. Five participants were lost to follow-up because of negative personal life events unrelated to the study, while two could not attend their final visits

due to the COVID-19 pandemic. Baseline demographic and clinical characteristics of the study population are shown in Table 1. There were significantly more males in the PPMS group (Mann Whitney U=184,  $p=0.0001$ ) and a significantly longer disease duration in the SPMS group (Mann Whitney U=154,  $p<0.0001$ ). SPMS study participants took significantly longer to complete a T25FW (Mann-Whitney U=256.5,  $p=0.034$ ) and had significantly higher baseline MSIS-psych scores (Mann-Whitney U=243,  $p=0.019$ ) compared with PPMS participants. No other baseline variables were significantly different between disease groups.

To confirm the internal validity of the overall dataset we performed correlations between baseline characteristics. There was a moderate negative correlation between age and  $E_A$  ( $r=-0.33$ ,  $p=0.014$ ), METs ( $r=-0.40$ ,  $p=0.002$ ), PhAD ( $r=-0.35$ ,  $p=0.009$ ), step count ( $r=-0.30$ ,  $p=0.024$ ), and the RAM score ( $r=-0.32$ ,  $p=0.015$ ) indicating that older study participants were less active. Disease duration correlated positively with baseline EDSS ( $r=0.286$ ,  $p=0.034$ ) and the T25FW ( $r=0.345$ ,  $p=0.009$ ). Body mass index (BMI) did not correlate with step count – the other RAM variables incorporated weight so their correlation with BMI was not considered. Study participants with a baseline EDSS of 6.0 or above (80% of participants) had significantly lower  $E_A$  scores (Mann Whitney U=147,  $p=0.046$ ), PhAD (Mann Whitney U=148,  $p=0.048$ ), step count (Mann Whitney U=32,  $p<0.0001$ ), METs (Mann Whitney U=105.5,  $p=0.003$ ), and RAM score (Mann Whitney U=89,  $p=0.001$ ) compared to those with baseline EDSS lower than 6.0.

## Usability

RAM was well tolerated by the study participants who wore the device most of the time (median: 99.4%, interquartile range: 1.5%, mean:  $99.5 \pm 2.4$  %). Wear-times did not differ for more disabled participants ( $EDSS \geq 6.0$  vs  $EDSS < 6.0$ ,  $n=44$ , Mann-Whitney U=211.5,  $p=0.528$ ). Daily wear-time did not diminish between successive six monthly assessment time points, and while it was constant during the first five days of each assessment time point ( $97.3 \pm 1.3\%$ ), there was a slight decrease on the sixth day, to  $94.9 \pm 3.2\%$  ( $p<0.0001$  for day, in an analysis of covariance of wear-time

versus day and six monthly assessment time point, controlling for between-subject variation). One participant had an allergic reaction to the metal plates of the device after the first trial so did not participate in this sub-study while another participant developed an allergic reaction at the one-year follow-up assessment.

#### **RAM validation: convergent and discriminant validity**

Convergent validity of RAM was assessed using cross-sectional correlations between the RAM variables and other physical disability measures (EDSS, MSFC and MSIS-phys) at each follow-up time point (Table 2, Supplementary Table 2). RAM variables, especially step count, correlated significantly with EDSS, T25FW, 9-HPT and MSIS-phys. Discriminant validity of RAM was assessed using cross-sectional correlations between RAM variables and other scores unrelated to physical disability (PASAT, MSIS-psych, Chalder Fatigue Scale and BDI). RAM variables failed to correlate with any of these measures (Table 2 and Supplementary Table 2).

#### **Responsiveness**

In this cohort progression was sustained, since one-sample Wilcoxon or t-tests showed that the slopes of most variables against time ( $dy/dt$ , where  $y$ =RAM variables, EDSS, MSFC and T25FW,  $t$ =time) were significantly different from zero in the direction of overall progression (Supplementary Table 3). Responsiveness of RAM, alongside the MSFC, EDSS and MSIS-phys, was assessed by the ability to detect significant ( $p<0.05$ ) change between pairs of assessments separated in time, using Wilcoxon matched pair signed rank tests. The rate of detection of progression was calculated for 6-, 12-, 18-, 24- and 30-monthly follow-up epochs, and expressed as a percentage of epochs with significant change out of all epochs (Table 3). All possible combinations of epochs were considered, including overlapping ones (Figure 3). The progression detection rate increased with longer intervals. Out of the RAM variables, step count had the highest rate of responsiveness followed by METs. RAM responsiveness was intermediate between that of the EDSS and MSFC.

Table 4 shows responsiveness for the full 30-month follow-up period. The SIMS study participants experienced a marked increase in disability with changes in all RAM variables, EDSS, MSFC (T25FW and PASAT) and fatigue scores. Participants with PPMS accumulated more disability during this study, compared to SPMS individuals. In order to investigate whether RAM responsiveness varied with disability, we repeated the analysis in participants with baseline EDSS  $\geq 6.0$  (n=44, range: 6.0-6.5) and EDSS < 6.0 (n=11, range: 1.5-5.5) separately. RAM was responsive in both groups (Supplementary Table 4).

### **Change scores**

Although there was good cross-sectional correlation between RAM scores and other physical measures (EDSS, T25FW, 9-HPT and MSIS-phys), there was no meaningful correlation between raw change in RAM variables and raw change in other outcome measures over the 30-month follow-up (Supplementary Table 5). This suggested that RAM measurements did not share the same progression trajectory with the other physical measures. In keeping with this explanation, there was: (1) no correlation between the slopes of the RAM variables ( $dy/dt$ , where  $y$ =RAM variable,  $t$ =time) and the other physical measures (Supplementary Table 6); (2) hardly any correlation between the rate of change of the slopes of the RAM variables ( $d^2y/dt^2$ , where  $y$ =RAM variable,  $t$ =time) and the other physical measures (Supplementary Table 7); (3) considerable difference in the percentage change from baseline between the different outcome measures (Figure 4).

### **Brain atrophy**

In order to validate RAM against an objective structural marker of disease progression, we used annualized percentage brain volume change since brain atrophy<sup>11</sup> has been used as a surrogate marker of disease progression in clinical trials<sup>27</sup>. Alignment of structural brain imaging with RAM assessments within two months of each other was available for 38 study participants. The annual

change in the RAM score (not individual RAM components) significantly correlated with annualized percentage brain volume change ( $r=0.357$ ,  $p=0.028$ ), (Figure 5).

### **Minimum wear-period**

Although participants in this study wore the device for six days, shorter periods may suffice. Since the best performing RAM score was step count, we selected this variable to assess responsiveness when the device was worn for incremental periods of time i.e. 1, 2, 3, 4, 5 and 6 days. A Kruskal-Wallis test showed no significant differences ( $p=0.996$ ) between any of the measurement periods (Supplementary Table 8). Data from the first day was as responsive as data from the entire six days (Table 5).

### **Discussion**

Remote disease monitoring is becoming increasingly desirable to reduce trips to hospitals for patient-specific, public health and environmental reasons, and the availability of patient outcome measures during remote neurology consultations would improve the quality of teleneurology<sup>4 28 29</sup>. While questionnaires can be utilised for remote disease monitoring, RAM was more responsive than the three questionnaires assessed here (MSIS-29, Chalder Fatigue Scale and BDI). This study was unique since it was longitudinal and assessed multiple physical activity variables in parallel in a progressive MS population, with an assessment frequency (six monthly) matching clinical practice, and a follow-up duration sufficient to test acceptability in the long-term. We demonstrated that remote activity monitoring of individuals with MS is feasible and well tolerated including by those with higher disability, and monitoring physical activity over periods shorter than a week may suffice. We trialled posting the device to a number of participants on separate occasions after they had attended for an initial face-to-face visit to instruct them on how to use the device, and this worked well.

The RAM variables displayed very good cross-sectional correlations with all other physical measurements (EDSS, MSFC, T25FW and MSIS-phys). Despite this cross-sectional correlation between the absolute values, several strands of evidence suggest that RAM variable progression trajectories differ from those of the other physical measurements. First, RAM change scores did not correlate with changes in the EDSS, MSFC, T25FW and MSIS-phys. Second, RAM scores had different rates of change compared to the other physical measurements. Third, statistically significant change in RAM and clinically meaningful change in the other variables were not always linked (Supplementary Table 9). Fourth, two participants showed a significant decrease in step count but had no decrease in other outcome measures. In keeping with this, a recent one year study showed that a decrease in step count change may be detectable in the absence of EDSS change<sup>30</sup>.

RAM variables were responsive to change with step count demonstrating the highest responsiveness, and cross-sectional correlation with other outcome measures, across all time intervals assessed. This may suggest that cheaper devices, measuring step alone, may suffice. However, the RAM score incorporating all four variables correlated with brain atrophy, while the individual RAM components did not. Hence, although change in steps may be easier to detect over shorter periods of time, the RAM score which incorporates overall physical activity and changes more slowly, is more closely related to structural brain change.

When designing this study, we planned for six days of monitoring since there was evidence that one needs this length of time for reliable measurements in healthy individuals using the device we used<sup>13</sup>. Hence we were surprised to find that one day (in our case the first day was a weekday) was as responsive as six days in terms of detecting progression. Our interpretation is that people with progressive MS may have less day-to-day variability than healthy individuals, so that one day of monitoring may suffice to capture their level of disability. Although we do not have data from age and sex matched individuals to prove this, studies in MS patients using the same<sup>14</sup> and other<sup>31</sup>

devices have shown that 2-3 days may suffice, compared to longer periods in healthy adults and children<sup>13 32</sup>.

This study had a number of important limitations. The use of intermittent assessment of physical activity, as opposed to continuous measurement, has the risk of bias from reactivity, i.e. a temporary increase in physical activity due to knowledge of being observed. On the other hand, intermittent assessment is more likely to be adopted in routine clinical practice. Seasonal variation may be important, since people may be more active during the summer months. Limited data was collected on the socioeconomic background or normal ambulation behaviours and environment, so activity differences related to participant lifestyles could not be controlled for. In this cohort, the mean baseline EDSS was 5.7 (range 1.5 to 6.5). While this is not unusual in progressive MS studies, it limits generalizability to early stages in MS and more study is therefore needed at lower levels of disability. Finally, brain atrophy may be influenced by factors other than MS progression (such as age and cerebrovascular disease risk factors). This study was not designed to investigate whether RAM is able to establish the relative contribution of these factors.

In conclusion, the RAM score seems a feasible and valid tool to measure progression and a potential longitudinal outcome measure in MS. Remote monitoring would reduce the need to travel for clinical appointments, reduce costs and carbon footprint, empower people with MS to participate in assessment of their condition, provide their healthcare professional with clinical outcome data during remote and/or face-to-face consultations, and facilitate longer term follow-up in research studies.

**Acknowledgements**

We would like to acknowledge APC Cardiovascular for their support of the study.

**Conflict of interest**

The authors declare that there is no conflict of interest.

**Funding**

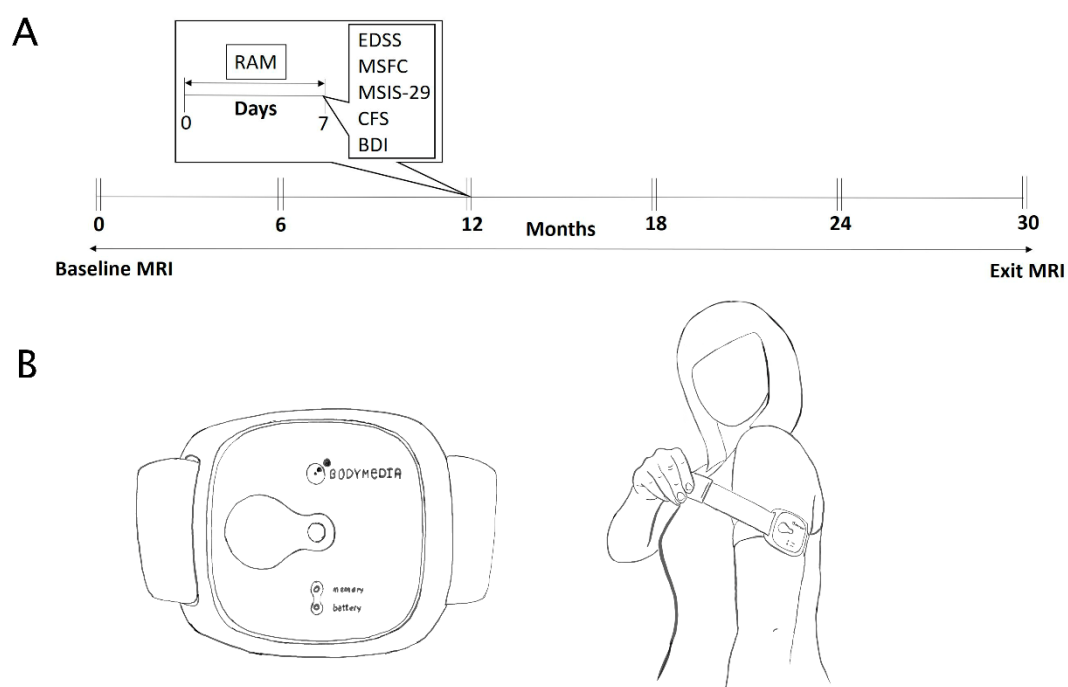
This work was supported by Wessex Medical Research and MS Society grant number 996 and Medical Research Council (award reference MR/R017352/1).

## References

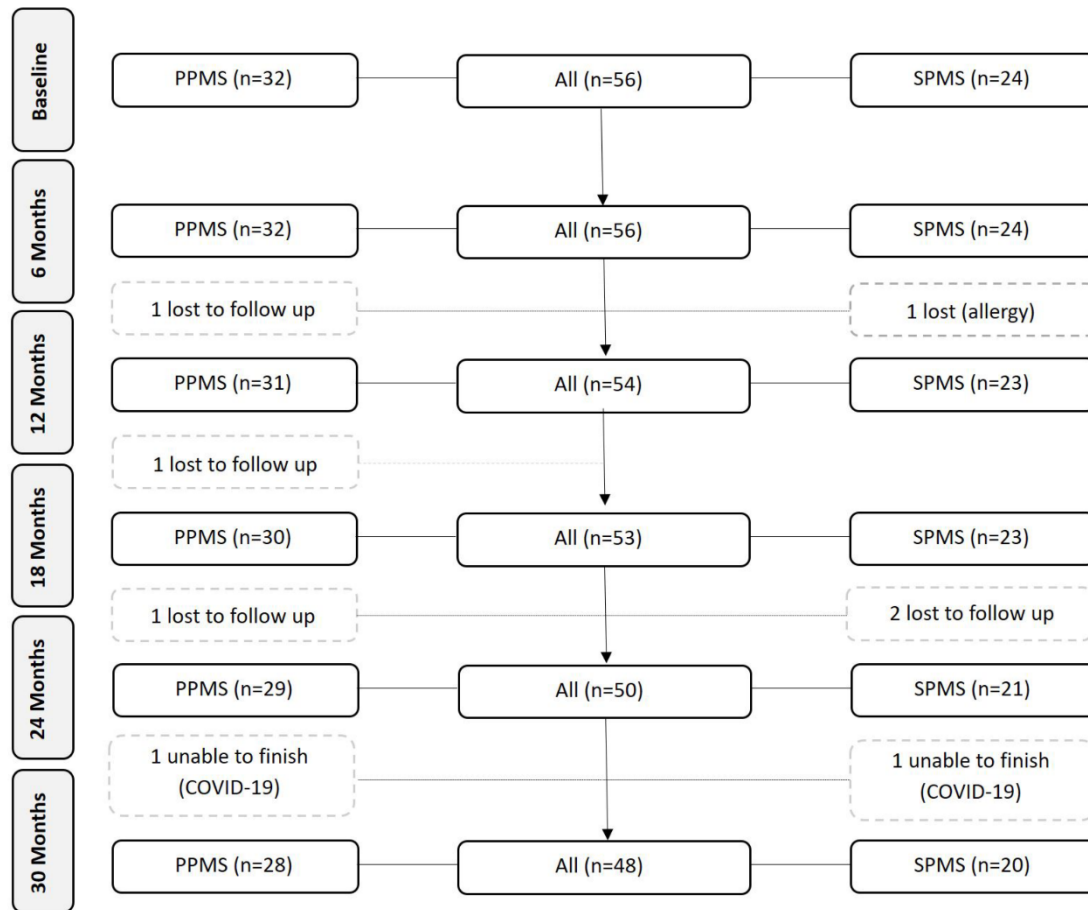
1. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;33(11):1444-52. doi: 10.1212/wnl.33.11.1444
2. Whitaker JN, McFarland HF, Rudge P, et al. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. *Multiple sclerosis (Houndmills, Basingstoke, England)* 1995;1(1):37-47. doi: 10.1177/135245859500100107
3. Fischer JS, Rudick RA, Cutter GR, et al. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force. *Multiple sclerosis (Houndmills, Basingstoke, England)* 1999;5(4):244-50. doi: 10.1177/135245859900500409
4. Bove R, Bevan C, Crabtree E, et al. Toward a low-cost, in-home, telemedicine-enabled assessment of disability in multiple sclerosis. *Multiple Sclerosis Journal* 2019;25(11):1526-34. doi: 10.1177/1352458518793527
5. Hobart J, Lamping D, Fitzpatrick R, et al. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. *Brain : a journal of neurology* 2001;124(Pt 5):962-73. doi: 10.1093/brain/124.5.962
6. Block VJ, Lizée A, Crabtree-Hartman E, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. *J Neurol* 2017;264(2):316-26. doi: 10.1007/s00415-016-8334-6
7. Block VA, Pitsch E, Tahir P, et al. Remote Physical Activity Monitoring in Neurological Disease: A Systematic Review. *PloS one* 2016;11(4):e0154335. doi: 10.1371/journal.pone.0154335
8. Motl RW, McAuley E, Sandroff BM. Longitudinal change in physical activity and its correlates in relapsing-remitting multiple sclerosis. *Physical therapy* 2013;93(8):1037-48. doi: 10.2522/ptj.20120479
9. McIninch J, Datta S, DasMahapatra P, et al. Remote Tracking of Walking Activity in MS Patients in a Real-World Setting (P3.209). *Neurology* 2015;84(14 Supplement):P3.209.
10. Bermel RA, Bakshi R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. *The Lancet Neurology* 2006;5(2):158-70. doi: [https://doi.org/10.1016/S1474-4422\(06\)70349-0](https://doi.org/10.1016/S1474-4422(06)70349-0)
11. Miller DH, Lublin FD, Sormani MP, et al. Brain atrophy and disability worsening in primary progressive multiple sclerosis: insights from the INFORMS study. *Ann Clin Transl Neurol* 2018;5(3):346-56. doi: 10.1002/acn3.534
12. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of neurology* 2011;69(2):292-302. doi: 10.1002/ana.22366
13. Scheers T, Philippaerts R, Lefevre J. Variability in physical activity patterns as measured by the SenseWear Armband: how many days are needed? *European journal of applied physiology* 2012;112(5):1653-62. doi: 10.1007/s00421-011-2131-9 [published Online First: 2011/08/30]
14. Norris M, Anderson R, Motl RW, et al. Minimum number of days required for a reliable estimate of daily step count and energy expenditure, in people with MS who walk unaided. *Gait & Posture* 2017;53:201-06. doi: <https://doi.org/10.1016/j.gaitpost.2017.02.005>
15. Chalder T, Berelowitz G, Pawlikowska T, et al. Development of a fatigue scale. *Journal of psychosomatic research* 1993;37(2):147-53. doi: 10.1016/0022-3999(93)90081-p
16. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. *Archives of general psychiatry* 1961;4:561-71. doi: 10.1001/archpsyc.1961.01710120031004
17. Coote S, O'Dwyer C. Comparative Validity of Accelerometer-Based Measures of Physical Activity for People With Multiple Sclerosis. *Archives of Physical Medicine and Rehabilitation* 2012;93(11):2022-28. doi: <https://doi.org/10.1016/j.apmr.2012.05.010>
18. Coote S, Uszynski M, Herring MP, et al. Effect of exercising at minimum recommendations of the multiple sclerosis exercise guideline combined with structured education or attention control

- education – secondary results of the step it up randomised controlled trial. *BMC Neurology* 2017;17(1):119. doi: 10.1186/s12883-017-0898-y
19. Reece JD, Barry V, Fuller DK, et al. Validation of the SenseWear Armband as a Measure of Sedentary Behavior and Light Activity. 2015;12(9):1229. doi: 10.1123/jpah.2014-0136
  20. Santos-Lozano A, Hernández-Vicente A, Pérez-Isaac R, et al. Is the SenseWear Armband accurate enough to quantify and estimate energy expenditure in healthy adults? *Ann Transl Med* 2017;5(5):97-97. doi: 10.21037/atm.2017.02.31
  21. Vrenken H, Jenkinson M, Horsfield MA, et al. Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis. *J Neurol* 2013;260(10):2458-71. doi: 10.1007/s00415-012-6762-5
  22. Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage* 2002;17(1):479-89. doi: 10.1006/nimg.2002.1040
  23. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 2004;23 Suppl 1:S208-19. doi: 10.1016/j.neuroimage.2004.07.051
  24. Battaglini M, Smith SM, Brogi S, et al. Enhanced brain extraction improves the accuracy of brain atrophy estimation. *NeuroImage* 2008;40(2):583-89. doi: <https://doi.org/10.1016/j.neuroimage.2007.10.067>
  25. Popescu V, Battaglini M, Hoogstrate WS, et al. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. *NeuroImage* 2012;61(4):1484-94. doi: <https://doi.org/10.1016/j.neuroimage.2012.03.074>
  26. Rosenthal R. Parametric measures of effect size. The handbook of research synthesis. New York, NY, US: Russell Sage Foundation 1994:231-44.
  27. Altmann DR, Jasperse B, Barkhof F, et al. Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. *Neurology* 2009;72(7):595-601. doi: 10.1212/01.wnl.0000335765.55346.fc
  28. Dorsey ER, Glidden AM, Holloway MR, et al. Teleneurology and mobile technologies: the future of neurological care. *Nature Reviews Neurology* 2018;14(5):285-97. doi: 10.1038/nrneurol.2018.31
  29. Marziniak M, Brichetto G, Feys P, et al. The Use of Digital and Remote Communication Technologies as a Tool for Multiple Sclerosis Management: Narrative Review. *JMIR Rehabil Assist Technol* 2018;5(1):e5-e5. doi: 10.2196/rehab.7805
  30. Block VJ, Bove R, Zhao C, et al. Association of Continuous Assessment of Step Count by Remote Monitoring With Disability Progression Among Adults With Multiple Sclerosis. *JAMA Netw Open* 2019;2(3):e190570-e70. doi: 10.1001/jamanetworkopen.2019.0570
  31. Motl RW, Zhu W, Park Y, et al. Reliability of scores from physical activity monitors in adults with multiple sclerosis. *Adapted physical activity quarterly : APAQ* 2007;24(3):245-53. doi: 10.1123/apaq.24.3.245 [published Online First: 2007/10/06]
  32. Ridgers ND, Hnatiuk JA, Vincent GE, et al. How many days of monitoring are needed to reliably assess SenseWear Armband outcomes in primary school-aged children? *Journal of science and medicine in sport* 2016;19(12):999-1003. doi: 10.1016/j.jsams.2016.02.009 [published Online First: 2016/03/14]

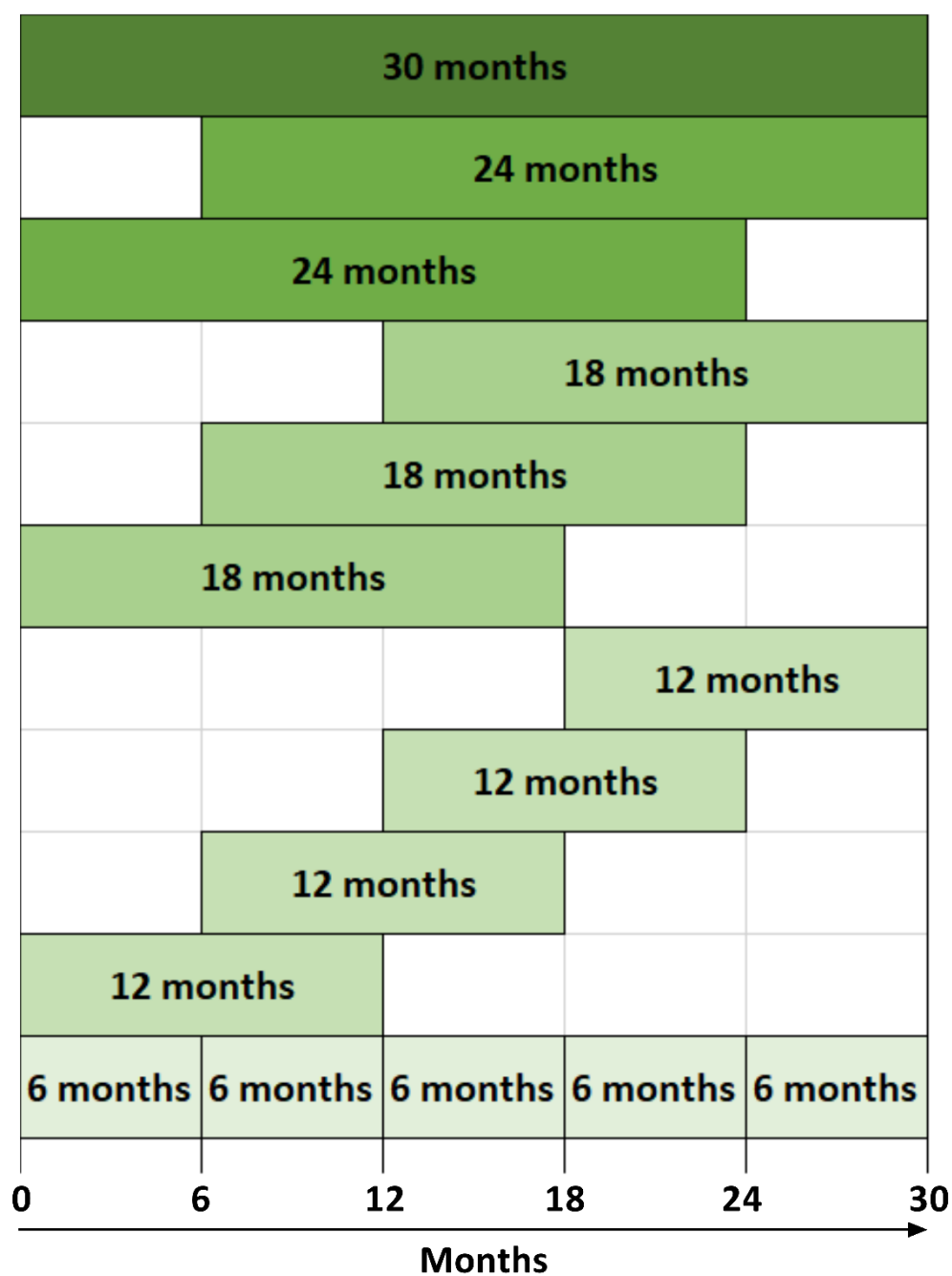
**Figure 1:** (A) Planned timeline for a study participant. At each assessment time point, participants attended two clinic visits, one week apart. The RAM device was worn between the first and second visits. The EDSS, MSFC, MSIS-29, CFS and BDI were conducted on the second visit. (B) The RAM device, an armband, was worn on the non-dominant arm.



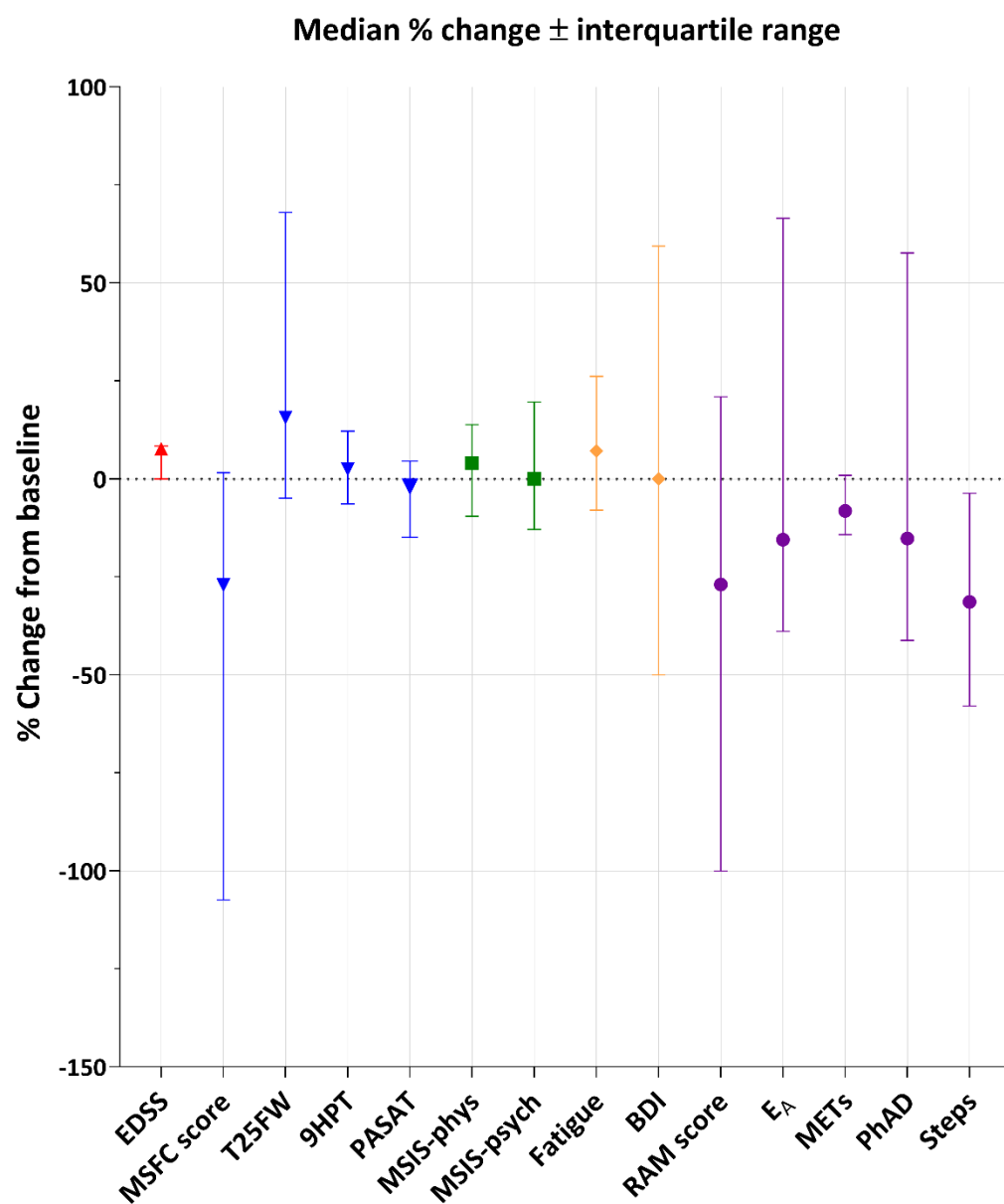
**Figure 2: CONSORT diagram**



**Figure 3:** Epochs considered in the responsiveness analyses (Tables 3 and 5). All possible epochs of each duration were considered, including overlapping epochs, i.e. 15 epochs in total.

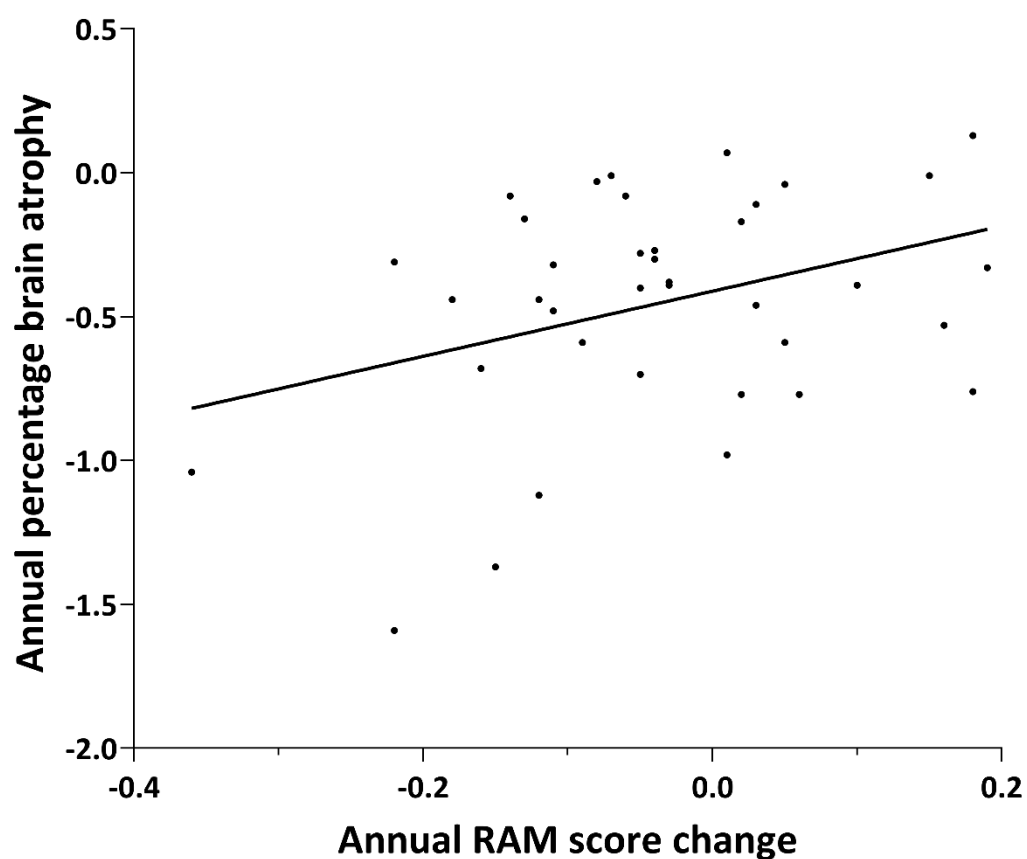


**Figure 4:** Percentage (%) change from baseline to 2.5 years, median  $\pm$  interquartile range. Positive change indicates worsening for EDSS, T25FW, 9HPT, MSIS-phys, MSIS-psych, fatigue and BDI while negative change indicates worsening for the RAM variables, RAM score, MSFC score and PASAT.



**Figure 5:** (A) Annual change in RAM score correlated with annualized percentage brain volume change while (B) individual RAM components did not correlate. EA and PhAD were normalised using the cube root transformation.

**A**



**B**

Annual RAM Change	RAM Score	E <sub>A</sub>	METs	PhAD	Steps
Pearson's r (p)	<b>0.357 (0.028)</b>	0.174 (0.296)	0.193 (0.245)	0.181 (0.277)	0.270 (0.101)

Table 1: Baseline demographics for the study cohort. Values represent mean (SD). The differences between PPMS and SPMS baseline scores were determined using a Mann-Whitney U test.

Demographic	All (n=56)	PPMS (n=32)	SPMS (n=24)	PPMS vs SPMS (significance, p)
Male/Female	26/30	22/10	4/20	<b>0.0001</b>
Age, years	53.6 (8.0)	53.88 (9.98)	53.13 (4.48)	0.362
Disease duration, years	12.2 (8.60)	8.82 (8.0)	16.8 (7.20)	<b>&lt;0.0001</b>
BMI	27.9 (5.90)	28.06 (5.09)	27.74 (6.94)	0.696
EDSS	5.7 (1.30)	5.42 (1.54)	6.0 (0.84)	0.352
MSFC score	-0.017 (0.75)	-0.06 (0.71)	-0.11 (0.80)	0.540
T25FW, seconds	13.8 (11.10)	12.36 (9.54)	15.79 (12.82)	<b>0.034</b>
9-HPT, seconds	34.6 (50.50)	39.21 (66.50)	28.39 (7.80)	0.889
PASAT	43.2 (12.60)	43.75 (13.15)	42.38 (12.02)	0.551
MSIS-phys	49.0 (13.10)	47.66 (13.90)	50.75 (12.11)	0.422
MSIS-psych	18.0 (6.10)	16.31 (5.88)	20.17 (5.85)	<b>0.019</b>
Fatigue score	15.50 (6.30)	14.28 (6.63)	17.13 (5.60)	0.103
BDI	5.20 (4.70)	4.52 (4.60)	6.17 (4.79)	0.180
E <sub>A</sub> , joules	1401.8 (1510.6)	1460.79 (1783.87)	1323.13 (1075.34)	0.954
METs	1.30 (0.30)	1.29 (0.26)	1.28 (0.26)	0.997
PhAD, hours	1.20 (1.20)	1.19 (1.32)	1.28 (1.14)	0.663
Step count	3411.50 (2913.6)	3607.31 (3330.32)	3150.43 (2286.05)	0.85
RAM score	0.007 (0.50)	-0.01 (0.62)	0.03 (0.42)	0.889

Table 2: Spearman's rho correlation matrix for baseline scores, separated by those showing convergent validity (ambulatory/physical scores) and those showing discriminant validity (cognitive/psychological scores). Significant correlations are shown in bold. Shading reflects strength of significant correlations.

	Baseline correlations	E <sub>A</sub>	METs	PhAD		Steps	RAM score	Key	
Convergent validity	EDSS (n=54)	<b>-0.343</b> (-0.011)	<b>-0.411</b> (-0.002)	<b>-0.325</b> (-0.016)		<b>-0.589</b> ( <b>&lt;0.0001</b> )	<b>-0.419</b> (-0.002)		p > 0.05
	MSFC (n=55)	<b>0.316</b> (0.019)	<b>0.365</b> (-0.006)	<b>0.305</b> (-0.024)		<b>0.493</b> (-0.0001)	<b>0.376</b> (-0.005)		p ≤ 0.05
	T25FW (n=55)	<b>-0.331</b> (0.014)	<b>-0.412</b> (-0.002)	<b>-0.317</b> (-0.018)		<b>-0.64</b> ( <b>&lt;0.0001</b> )	<b>-0.444</b> (-0.001)		p ≤ 0.01
	9-HPT (n=55)	-0.212 (0.121)	-0.254 (-0.061)	-0.22 (-0.106)		<b>-0.436</b> (-0.001)	<b>-0.321</b> (-0.017)		p ≤ 0.001
	MSIS-phys (n= 55)	<b>-0.299</b> (0.027)	<b>-0.287</b> (-0.033)	<b>-0.278</b> (-0.04)		<b>-0.447</b> (-0.001)	<b>-0.356</b> (-0.008)		p ≤ 0.0001
Discriminant validity	PASAT (n=55)	0.137 (0.318)	0.154 (-0.261)	0.126 (-0.359)		0.134 (-0.329)	0.105 (-0.444)		
	MSIS-psych (n=55)	-0.142 (0.303)	-0.185 (-0.177)		-0.119 (-0.385)	-0.234 (-0.085)	-0.143 (-0.297)		
	Fatigue score (n=55)	-0.163 (0.235)	-0.231 (-0.09)	-0.153 (-0.265)		-0.203 (-0.138)	-0.14 (-0.307)		
	BDI (n=54)	-0.108 (0.437)	-0.052 (-0.707)	-0.043 (-0.758)		-0.179 (-0.195)	-0.112 (-0.422)		

Table 3: Percentage of epochs with significant responsiveness for EDSS, MSFC variables, MSIS-phys and RAM variables. The n-numbers in the first column refer to the number of possible epochs of a certain duration – see Figure 3. The mean progression detection rate in the bottom row represents the mean of all possible epochs of any duration i.e. 15 epochs.

Follow up period	EDSS	MSFC score	T25FW	9HPT	MSIS-phys	E <sub>A</sub>	METs	Step count	PhAD	RAM score	Key	
6 months (n=5)	0	40	20	20	20	0	0	20	0	0		0% -20%
12 months (n=4)	0	25	75	25	50	50	50	50	50	25		≥20% <40%
18 months (n=3)	33.3	66.7	66.7	33.3	33.3	33.3	33.3	66.7	33.3	33.3		≥40% <60%
24 months (n=2)	50	100	100	0	0	0	100	100	0	50		≥60% <80%
30 months (n=1)	100	100	100	0	0	0	100	100	100	100		≥80% ≤100%
Mean progression detection rate (%)	36.7	66.3	72.3	15.7	20.7	36.7	56.7	67.3	36.7	41.7		

Table 4: Responsiveness over the 30 month follow-up period for all participants. Wilcoxon signed rank tests were used to assess if a significant change had occurred. Significant changes are in bold. Effect sizes were calculated by dividing the Wilcoxon Z by the square root of the number of observations ( $Z/\sqrt{n}$ ).

Disability measure	Baseline median (IQR)			End median (IQR)			Significance (exact, 2-tailed)			Effect size		
	All	PPMS	SPMS	All	PPMS	SPMS	All	PPMS	SPMS	All	PPMS	SPMS
EDSS	6.0 (0.5)	6.0 (1.75)	6.0 (0.5)	6.5 (0.5)	6.5 (0.5)	6.5 (0.5)	<b>0.0005</b>	<b>0.004</b>	0.092	0.49	0.53	0.45
MSFC score	0.06 (1.1)	0.07 (1.09)	0.04 (1.2)	0.05 (1.27)	0.04 (1.72)	0.14 (1.13)	<b>0.0003</b>	<b>0.001</b>	0.111	0.51	0.6	0.35
T25FW, seconds	9.95 (7.61)	8.82 (5.79)	11.66 (6.48)	11.99 (10.71)	12.46 (13.66)	11.99 (7.92)	<b>0.0002</b>	<b>&lt;0.0001</b>	0.393	0.51	0.77	0.19
9-HPT, seconds	26.92 (9.63)	26.15 (10.68)	27.44 (9.14)	25.68 (12.13)	27.46 (16.25)	24.02 (8.75)	0.067	<b>0.014</b>	0.919	0.26	0.46	0.03
PASAT	46.5 (21.75)	49.0 (21.75)	44.5 (21.75)	46.0 (24.0)	47.0 (30.25)	46.0 (15.5)	<b>0.035</b>	<b>0.023</b>	0.821	0.3	0.43	0.05
MSIS-phys	50.0 (20.5)	48.0 (20.75)	50.5 (19.0)	49.0 (17.0)	48.5 (26.25)	50.0 (15)	0.336	0.383	0.693	0.14	0.17	0.09
MSIS-psych	16.5 (9.75)	15.5 (6.75)	20.5 (11.25)	19.0 (10.0)	16.0 (10.75)	21.0 (11.0)	0.332	0.17	0.908	0.14	0.26	0.03
Fatigue score	14.0 (8.75)	13.0 (7.5)	16.5 (9.25)	16.0 (8.5)	15.0 (11.0)	18.0 (6.0)	<b>0.033</b>	<b>0.032</b>	0.499	0.3	0.4	0.15
BDI	3.0 (7.0)	3.0 (5.0)	5.0 (9.75)	5.0 (7.5)	5.0 (7.0)	4.0 (9.0)	0.854	0.244	0.372	0.03	0.22	0.2
E <sub>A</sub> , joules	1169.72 (1454.91)	1117.92 (1406.36)	1389.78 (1691.41)	881.31 (1321.33)	951.2 (1384.5)	837.78 (1343.38)	0.078	<b>0.014</b>	0.927	0.25	0.46	0.03
METs	1.29 (0.32)	1.25 (0.26)	1.31 (0.42)	1.22 (0.42)	1.22 (0.37)	1.23 (0.57)	<b>0.001</b>	<b>0.001</b>	0.258	0.45	0.58	0.26
PhAD, hours	0.93 (1.41)	0.9 (1.23)	0.99 (2.06)	0.73 (1.29)	0.72 (1.13)	0.73 (1.66)	<b>0.044</b>	<b>0.009</b>	0.985	0.29	0.49	0.01
Step count	2503.4 (3137.42)	2897.99 (3372.76)	2365.16 (2965.13)	1841.9 (2205.04)	1881.8 (2547.62)	1671.32 (2076)	<b>0.0002</b>	<b>0.001</b>	0.097	0.52	0.61	0.38
RAM score	0.18 (0.81)	0.18 (0.88)	0.17 (0.81)	-0.07 (0.75)	-0.03 (0.84)	-0.15 (0.69)	<b>0.02</b>	<b>0.003</b>	0.784	0.33	0.54	0.07

Table 5: Percentage of epochs with significant responsiveness (rows) to change in step count across different wear time durations (columns). The n-numbers in the first column refer to the number of possible epochs of a certain duration – see Figure 3. The mean progression detection rate (bottom row) is the average across all 15 epochs.

Follow up period	1 day	2 days	3 days	4 days	5 days	6 days	Key	
6 months (n=5)	0	20	40	20	20	20		0% -20%
12 months (n=4)	50	75	75	50	75	50		≥20% <40%
18 months (n=3)	100	66.7	100	66.7	100	66.7		≥40% <60%
24 months (n=2)	100	100	100	100	100	100		≥60% <80%
30 months (n=1)	100	100	100	100	100	100		≥80% ≤100%
Mean progression detection rate (%)	70.0	72.3	83.0	67.3	79.0	67.3		

## Supplementary materials

*Supplementary Table 1: Spearman's rho correlation coefficients (p-value) between RAM components and overall RAM scores at each assessment time point.*

Assessment time point	E <sub>A</sub>	METs	PhAD	Steps
<b>Baseline (n=56)</b>	0.875 (<0.0001)	0.796 (<0.0001)	0.844 (<0.0001)	0.817 (<0.0001)
<b>6 months (n=55)</b>	0.897 (<0.0001)	0.730 (<0.0001)	0.877 (<0.0001)	0.777 (<0.0001)
<b>12 months (n=54)</b>	0.865 (<0.0001)	0.694 (<0.0001)	0.80 (<0.0001)	0.850 (<0.0001)
<b>18 months (n=52)</b>	0.871 (<0.0001)	0.691 (<0.0001)	0.837 (<0.0001)	0.762 (<0.0001)
<b>24 months (n=50)</b>	0.889 (<0.0001)	0.772 (<0.0001)	0.837 (<0.0001)	0.788 (<0.0001)
<b>30 months (n=48)</b>	0.878 (<0.0001)	0.820 (<0.0001)	0.861 (<0.0001)	0.819 (<0.0001)

Supplementary Table 2: Spearman's rho correlation coefficients (p-value) between scores at all assessment time points. Significant correlations are highlighted in bold

Baseline correlations	E <sub>A</sub>	METs	PhAD	Steps	RAM score
EDSS (n=54)	<b>-0.343</b> (0.011)	<b>-0.411</b> (0.002)	<b>-0.325</b> (0.016)	<b>-0.589</b> ( <b>&lt;0.0001</b> )	<b>-0.419</b> (0.002)
MSFC (n=55)	<b>0.316</b> (0.019)	<b>0.365</b> (0.006)	<b>0.305</b> (0.024)	<b>0.493</b> (0.0001)	<b>0.376</b> (0.005)
T25FW (n=55)	<b>-0.331</b> (0.014)	<b>-0.412</b> (0.002)	<b>-0.317</b> (0.018)	<b>-0.640</b> ( <b>&lt;0.0001</b> )	<b>-0.444</b> (0.001)
9-HPT (n=55)	-0.212 (0.121)	-0.254 (0.061)	-0.220 (0.106)	<b>-0.436</b> (0.001)	<b>-0.321</b> (0.017)
MSIS-phys (n= 55)	<b>-0.299</b> (0.027)	<b>-0.287</b> (0.033)	<b>-0.278</b> (0.04)	<b>-0.447</b> (0.001)	<b>-0.356</b> (0.008)
PASAT (n=55)	0.137 (0.318)	0.154 (0.261)	0.126 (0.359)	0.134 (0.329)	0.105 (0.444)
MSIS-psych (n=55)	-0.142 (0.303)	-0.185 (0.177)	-0.119 (0.385)	-0.234 (0.085)	-0.143 (0.297)
Fatigue score (n=55)	-0.163 (0.235)	-0.231 (0.09)	-0.153 (0.265)	-0.203 (0.138)	-0.14 (0.307)
BDI (n=54)	-0.108 (0.437)	-0.052 (0.707)	-0.043 (0.758)	-0.179 (0.195)	-0.112 (0.422)

6 month correlations	E <sub>A</sub>	METs	PhAD	Steps	RAM score
EDSS (n=39)	-0.140 (0.396)	-0.270 (0.097)	-0.125 (0.45)	<b>-0.355</b> (0.027)	-0.236 (0.147)
MSFC (n=55)	0.265 (0.050)	0.225 (0.099)	0.249 (0.066)	<b>0.316</b> (0.019)	<b>0.342</b> (0.011)
T25FW (n=55)	<b>-0.30</b> (0.026)	<b>-0.282</b> (0.037)	<b>-0.282</b> (0.037)	<b>-0.497</b> (0.0001)	<b>-0.421</b> (0.001)
9-HPT (n=55)	-0.147 (0.283)	-0.181 (0.187)	-0.161 (0.240)	<b>-0.333</b> (0.013)	<b>-0.284</b> (0.035)
MSIS-phys (n=54)	-0.260 (0.057)	-0.288 (0.034)	-0.246 (0.073)	<b>-0.346</b> (0.010)	<b>-0.323</b> (0.017)
PASAT (n=55)	0.143 (0.298)	0.005 (0.974)	0.099 (0.472)	-0.039 (0.780)	0.102 (0.457)
MSIS-psych (n=54)	-0.071 (0.609)	-0.129 (0.352)	-0.075 (0.592)	-0.155 (0.263)	-0.074 (0.593)

Fatigue score (n=55)	-0.184 (0.18)	-0.239 (0.079)	-0.228 (0.095)	-0.215 (0.114)	-0.181 (0.187)
BDI (n=55)	-0.013 (0.923)	-0.098 (0.475)	-0.003 (0.985)	-0.153 (0.264)	-0.041 (0.768)

12 month correlations	E <sub>A</sub>	METs	PhAD	Steps	RAM score
EDSS (n=44)	<b>-0.381</b> (0.011)	<b>-0.370</b> (0.013)	<b>-0.387</b> (0.009)	<b>-0.514</b> (0.003)	<b>-0.433</b> (0.003)
MSFC (n=54)	<b>0.330</b> (0.015)	<b>0.276</b> (0.043)	<b>0.314</b> (0.021)	<b>0.504</b> (0.0001)	<b>0.395</b> (0.003)
T25FW (n=54)	<b>-0.432</b> (0.001)	<b>-0.399</b> (0.003)	<b>-0.405</b> (0.002)	<b>-0.635</b> ( <b>&lt;0.0001</b> )	<b>-0.512</b> ( <b>&lt;0.0001</b> )
9-HPT (n=53)	-0.235 (0.09)	-0.184 (0.188)	-0.230 (0.097)	<b>-0.492</b> (0.0002)	<b>-0.399</b> (0.003)
MSIS-phys (n=54)	-0.235 (0.088)	-0.233 (0.089)	-0.206 (0.135)	<b>-0.456</b> (0.001)	<b>-0.341</b> (0.012)
PASAT (n=54)	0.116 (0.404)	0.08 (0.566)	0.122 (0.378)	0.032 (0.817)	0.027 (0.844)
MSIS-psych (n=54)	-0.002 (0.986)	-0.021 (0.88)	0.014 (0.922)	-0.175 (0.205)	-0.096 (0.491)
Fatigue score (n=54)	-0.111 (0.426)	-0.11 (0.429)	-0.105 (0.45)	-0.159 (0.25)	-0.136 (0.328)
BDI (n=54)	-0.126 (0.363)	-0.099 (0.477)	-0.104 (0.452)	<b>-0.273</b> (0.046)	-0.174 (0.208)

18 month correlations	E <sub>A</sub>	METs	PhAD	Steps	RAM score
EDSS (n=28)	-0.284 (0.136)	-0.244 (0.203)	-0.253 (0.185)	<b>-0.585</b> (0.001)	<b>-0.498</b> (0.006)
MSFC (n=49)	0.241 (0.085)	0.199 (0.158)	0.246 (0.079)	<b>0.495</b> (0.0002)	<b>0.476</b> (0.0004)
T25FW (n=49)	<b>-0.328</b> (0.018)	<b>-0.304</b> (0.028)	<b>-0.329</b> (0.017)	<b>-0.543</b> ( <b>&lt;0.0001</b> )	<b>-0.490</b> (0.0002)
9-HPT (n=49)	<b>-0.347</b> (0.013)	<b>-0.357</b> (0.011)	<b>-0.371</b> (0.008)	<b>-0.510</b> (0.0002)	<b>-0.534</b> ( <b>&lt;0.0001</b> )
MSIS-phys (n=48)	<b>-0.334</b> (0.019)	-0.273 (0.057)	<b>-0.353</b> (0.013)	<b>-0.406</b> (0.004)	<b>-0.462</b> (0.001)
PASAT (n=49)	0.085 (0.559)	-0.064 (0.661)	0.052 (0.72)	0.043 (0.766)	0.176 (0.221)
MSIS-psych (n=48)	0.009 (0.951)	0.047 (0.746)	-0.007 (0.964)	-0.082 (0.574)	-0.067 (0.647)
Fatigue score (n=48)	0.024 (0.868)	0.052 (0.722)	0.003 (0.984)	0.014 (0.926)	0.023 (0.876)

BDI (n=48)	-0.020 (0.892)	-0.037 (0.8)	-0.026 (0.86)	-0.163 (0.263)	-0.110 (0.453)
------------	-------------------	-----------------	------------------	-------------------	-------------------

24 month correlations	E <sub>A</sub>	METs	PhAD	Steps	RAM score
EDSS (n=42)	-0.30 (0.053)	<b>-0.320</b> <b>(0.039)</b>	-0.275 (0.077)	<b>-0.679</b> <b>(&lt;0.0001)</b>	<b>-0.520</b> <b>(0.0004)</b>
MSFC (n=50)	<b>0.304</b> <b>(0.032)</b>	<b>0.352</b> <b>(0.012)</b>	<b>0.301</b> <b>(0.033)</b>	<b>0.629</b> <b>(&lt;0.0001)</b>	<b>0.507</b> <b>(0.0002)</b>
T25FW (n=50)	<b>-0.347</b> <b>(0.014)</b>	<b>-0.374</b> <b>(0.008)</b>	<b>-0.310</b> <b>(0.029)</b>	<b>-0.668</b> <b>(&lt;0.0001)</b>	<b>-0.514</b> <b>(0.0001)</b>
9-HPT (n=50)	-0.205 (0.153)	<b>-0.332</b> <b>(0.018)</b>	-0.216 (0.131)	<b>-0.550</b> <b>(&lt;0.0001)</b>	<b>-0.403</b> <b>(0.004)</b>
MSIS-phys (n=50)	<b>-0.353</b> <b>(0.012)</b>	<b>-0.383</b> <b>(0.006)</b>	<b>-0.373</b> <b>(0.008)</b>	<b>-0.433</b> <b>(0.002)</b>	<b>-0.440</b> <b>(0.001)</b>
PASAT (n=50)	0.162 (0.26)	0.098 (0.498)	0.144 (0.317)	0.165 (0.252)	0.240 (0.093)
MSIS-psych (n=50)	-0.211 (0.141)	-0.217 (0.130)	-0.206 (0.152)	-0.148 (0.306)	-0.168 (0.244)
Fatigue score (n=50)	-0.059 (0.684)	-0.149 (0.303)	-0.063 (0.665)	-0.156 (0.28)	-0.085 (0.557)
BDI (n=49)	-0.156 (0.284)	-0.196 (0.178)	-0.175 (0.229)	-0.235 (0.104)	-0.186 (0.20)

30 month correlations	E <sub>A</sub>	METs	PhAD	Steps	RAM score
EDSS (n=43)	<b>-0.347</b> <b>(0.019)</b>	<b>-0.375</b> <b>(0.011)</b>	<b>-0.363</b> <b>(0.014)</b>	<b>-0.665</b> <b>(&lt;0.0001)</b>	<b>-0.609</b> <b>(&lt;0.0001)</b>
MSFC (n=46)	<b>0.376</b> <b>(0.008)</b>	<b>0.421</b> <b>(0.003)</b>	<b>0.399</b> <b>(0.005)</b>	<b>0.599</b> <b>(&lt;0.0001)</b>	<b>0.588</b> <b>(&lt;0.0001)</b>
T25FW (n=46)	<b>-0.484</b> <b>(0.0004)</b>	<b>-0.510</b> <b>(0.0002)</b>	<b>-0.487</b> <b>(0.0004)</b>	<b>-0.708</b> <b>(&lt;0.0001)</b>	<b>-0.705</b> <b>(&lt;0.0001)</b>
9-HPT (n=46)	<b>-0.390</b> <b>(0.006)</b>	<b>-0.373</b> <b>(0.009)</b>	<b>-0.401</b> <b>(0.005)</b>	<b>-0.443</b> <b>(0.002)</b>	<b>-0.496</b> <b>(0.0003)</b>
MSIS-phys (n=46)	<b>-0.295</b> <b>(0.042)</b>	<b>-0.346</b> <b>(0.016)</b>	<b>-0.304</b> <b>(0.036)</b>	<b>-0.378</b> <b>(0.008)</b>	<b>-0.437</b> <b>(0.002)</b>
PASAT (n=46)	0.167 (0.257)	0.135 (0.361)	0.170 (0.247)	0.154 (0.295)	0.250 (0.086)
MSIS-psych (n=46)	-0.138 (0.35)	-0.222 (0.13)	-0.127 (0.39)	-0.103 (0.487)	-0.166 (0.260)
Fatigue score (n=46)	-0.207 (0.157)	-0.239 (0.102)	-0.222 (0.130)	-0.122 (0.407)	-0.216 (0.140)
BDI (n=46)	-0.124 (0.401)	-0.139 (0.345)	-0.108 (0.463)	-0.088 (0.550)	-0.159 (0.280)

*Supplementary Table 3: To assess if overall progression was significantly different from zero, one sample t-tests (for parametric data) or one sample Wilcoxon tests (for non-parametric data, denoted by \*) were employed. Mean (\*or median) slopes of variables against time (dy/dt, where y=variables of disability measures, t=time) and standard deviation (\*or interquartile range, IQR) are shown with significant changes (p values) highlighted in bold. All outcome variables were converted to z-scores for this analysis, therefore a negative value represents worsening of scores for all outcome variables.*

Disability measure	N	Mean (or Median*) slope	SD (or IQR*)	Significance (2-tailed)
EDSS*	54	-0.08*	0.18*	<b>0.007</b>
MSFC score*	55	-0.05*	0.19*	<b>0.002</b>
T25FW, seconds*	50	-0.07*	0.21*	<b>&lt;0.0001</b>
9-HPT, seconds*	55	-0.05*	0.23*	0.074
PASAT*	55	-0.01*	0.21*	0.257
MSIS-phys	55	-0.04	0.25	0.254
MSIS-psych*	55	0.0*	0.33*	0.527
Fatigue score	55	-0.05	0.36	0.267
BDI*	55	0.01*	0.39*	0.912
E <sub>A</sub> , joules	54	-0.09	0.27	<b>0.023</b>
METs	54	0.11	0.29	<b>0.006</b>
PhAD, hours	54	-0.09	0.23	<b>0.008</b>
Step count	54	-0.15	0.24	<b>&lt;0.0001</b>
RAM score*	54	-0.05*	0.15*	<b>0.026</b>

Supplementary Table 4: Responsiveness over the 30 month follow-up period: sensitivity analysis by EDSS <6.0 and EDSS ≥6.0. Wilcoxon signed rank tests were used to assess if a significant change had occurred. Significant changes are in bold. Effect sizes were calculated by dividing the Wilcoxon Z by the square root of the number of observations ( $Z/\sqrt{n}$ )

Disability measure	Baseline median (IQR)		End median (IQR)		Significance (exact, 2-tailed)		Effect size	
	EDSS <6.0	EDSS ≥6.0	EDSS <6.0	EDSS ≥6.0	EDSS <6.0	EDSS ≥6.0	EDSS <6.0	EDSS ≥6.0
EDSS	3.5 (2.5)	6.0 (0.5)	4.0 (4.0)	6.5 (0.5)	0.15	<b>0.0004</b>	0.48	0.59
MSFC score	0.73 (0.71)	-0.11 (1.06)	0.53 (0.78)	-0.23 (1.55)	0.495	<b>0.0004</b>	0.24	0.56
T25FW, seconds	5.95 (2.24)	11.62 (9.27)	5.94 (2.66)	14.75 (14.37)	0.10	<b>0.0003</b>	0.53	0.56
9-HPT, seconds	22.21 (7.58)	27.58 (10.34)	23.98 (8.79)	26.56 (14.21)	0.32	0.14	0.34	0.24
PASAT	50.0 (8.0)	44.5 (23.0)	51.5 (13.0)	45.5 (30.25)	0.90	<b>0.008</b>	0.05	0.42
MSIS-phys	36.0 (22)	53.0 (13.75)	34.5 (19.25)	53.5 (19.5)	0.54	0.48	0.21	0.12
MSIS-psych	14.0 (6.0)	18.0 (10.5)	15.0 (10.25)	20.0 (10.75)	0.63	0.50	0.23	0.11
Fatigue score	11.0 (9.0)	14.0 (9.0)	13.5 (6.5)	16.5 (8.0)	<b>0.02</b>	0.24	0.75	0.19
BDI	2.0 (2.0)	4.0 (7.0)	2.5 (6.0)	5.0 (8.0)	0.59	0.89	0.2	0.02
E <sub>A</sub> , joules	1434.97 (545.53)	1018.41 (1627.61)	1153.77 (1292.15)	769.27 (1383.13)	0.06	0.31	0.6	0.17
METs	1.44 (0.13)	1.21 (0.36)	1.32 (0.17)	1.18 (0.42)	<b>0.03</b>	<b>0.01</b>	0.69	0.42
PhAD, hours	1.36 (1.11)	0.84 (1.57)	1.05 (0.91)	0.65 (1.27)	<b>0.05</b>	0.23	0.63	0.2
Step count	6636.75 (2926.67)	2190.23 (1767.83)	4273.75 (3704.14)	1480.58 (1790.39)	<b>0.009</b>	<b>0.003</b>	0.79	0.47
RAM score	0.42 (0.27)	-0.07 (0.79)	0.22 (0.48)	-0.25 (0.76)	0.08	0.06	0.56	0.31

Supplementary Table 5: Spearman's rho correlation coefficients (p-value) between change scores.  
Significant correlations are highlighted in bold.

Change correlations	E <sub>A</sub>	METs	PhAD	Steps	RAM score
EDSS (n=47)	0.065 (0.663)	-0.196 (0.187)	0.005 (0.975)	-0.128 (0.390)	-0.032 (0.833)
MSFC (n=50)	-0.191 (0.184)	-0.09 (0.532)	-0.17 (0.238)	-0.091 (0.531)	-0.11 (0.449)
T25FW (n=50)	0.155 (0.282)	0.024 (0.867)	0.137 (0.343)	0.076 (0.6)	0.007 (0.964)
9-HPT (n=50)	0.178 (0.216)	0.088 (0.543)	0.203 (0.157)	0.093 (0.519)	0.124 (0.391)
MSIS-phys (n=50)	0.197 (0.171)	0.168 (0.243)	0.182 (0.207)	0.014 (0.922)	0.129 (0.374)
PASAT (n=50)	0.140 (0.333)	0.140 (0.332)	0.165 (0.253)	0.178 (0.215)	0.176 (0.221)
MSIS-psych (n=50)	0.226 (0.114)	0.154 (0.284)	0.186 (0.195)	0.272 (0.056)	<b>0.306</b> <b>(0.031)</b>
Fatigue Score (n=50)	0.046 (0.749)	0.098 (0.498)	0.057 (0.693)	0.013 (0.931)	0.108 (0.456)
BDI (n=49)	<b>0.295</b> <b>(0.039)</b>	0.075 (0.608)	0.250 (0.083)	0.246 (0.088)	<b>0.356</b> <b>(0.012)</b>

*Supplementary Table 6: Spearman's rho correlation coefficients (p-value) between the slopes of the outcome variables over the 30-month follow-up period. All outcome variables were converted to z-scores for this analysis.*

Slope correlations	E <sub>A</sub>	METs	PhAD	Steps	RAM score
<b>EDSS (n=53)</b>	0.111 (0.427)	0.009 (0.949)	0.118 (0.399)	0.163 (0.244)	0.150 (0.282)
<b>MSFC (n=54)</b>	-0.023 (0.87)	0.103 (0.46)	-0.041 (0.771)	-0.049 (0.727)	-0.094 (0.498)
<b>T25FW (n=54)</b>	0.062 (0.656)	0.064 (0.648)	0.057 (0.683)	0.221 (0.109)	0.110 (0.428)
<b>9-HPT (n=54)</b>	0.174 (0.209)	-0.080 (0.567)	0.125 (0.37)	0.125 (0.368)	0.054 (0.701)
<b>MSIS-phys (n=54)</b>	-0.007 (0.962)	0.079 (0.572)	-0.011 (0.936)	0.026 (0.853)	-0.020 (0.885)
<b>PASAT (n=54)</b>	-0.130 (0.351)	0.092 (0.507)	-0.145 (0.296)	0.125 (0.368)	-0.024 (0.861)
<b>MSIS-psych (n=54)</b>	0.168 (0.23)	-0.034 (0.807)	0.179 (0.2)	-0.002 (0.989)	0.023 (0.871)
<b>Fatigue Score (n=54)</b>	0.101 (0.466)	-0.029 (0.836)	0.102 (0.461)	-0.053 (0.702)	-0.013 (0.925)
<b>BDI (n=54)</b>	0.002 (0.988)	-0.082 (0.554)	0.011 (0.935)	-0.117 (0.399)	-0.166 (0.231)

Supplementary Table 7: Spearman's rho correlation coefficients (p-value) between the rate of change of slopes of the outcome variables. Slopes were calculated for every six monthly epoch except for EDSS which was calculated using two 18 month epochs. Significant correlations are highlighted in bold. All outcome variables were converted to z-scores for this analysis.

Rate of change of slopes correlations	E <sub>A</sub>	METs	PhAD	Steps	RAM score
EDSS (n=43)	0.044 (0.779)	<b>-0.399</b> <b>(0.008)</b>	0.043 (0.783)	0.150 (0.337)	-0.085 (0.589)
MSFC (n=50)	-0.128 (0.367)	-0.106 (0.456)	-0.136 (0.337)	-0.109 (0.441)	-0.17 (0.228)
T25FW (n=50)	0.009 (0.951)	-0.031 (0.83)	-0.007 (0.959)	0.253 (0.071)	-0.02 (0.889)
9-HPT (n=50)	-0.209 (0.137)	-0.129 (0.361)	-0.207 (0.141)	-0.133 (0.346)	-0.271 (0.052)
MSIS-phys (n=50)	<b>0.279</b> <b>(0.047)</b>	-0.189 (0.184)	0.273 (0.052)	<b>0.343</b> <b>(0.014)</b>	0.245 (0.084)
PASAT (n=50)	-0.128 (0.365)	0.136 (0.335)	-0.129 (0.361)	0.01 (0.946)	-0.089 (0.529)
MSIS-psych (n=50)	0.244 (0.082)	<b>-0.332</b> <b>(0.016)</b>	0.255 (0.068)	<b>0.274</b> <b>(0.05)</b>	0.206 (0.143)
Fatigue Score (n=50)	0.102 (0.472)	0.046 (0.747)	0.066 (0.64)	-0.06 (0.674)	0.141 (0.319)
BDI (n=50)	0.183 (0.193)	0.003 (0.984)	0.195 (0.165)	0.222 (0.114)	0.211 (0.132)

*Supplementary Table 8: A Kruskal-Wallis test were performed to assess whether there was a difference in the mean step count between RAM device wear duration of six days versus shorter periods ( $p=0.996$ ).*

<b>Averaged over:</b>	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Mean rank</b>	<b>Mean rank difference</b>	<b>Z</b>	<b>Significance, (approximate, 2-tailed)</b>
<b>One Day</b>	326	2890.1	2674.0	966.1	11.2	0.253	>0.9999
<b>Two Days</b>	326	3005.9	2731.3	988.8	-11.5	0.260	>0.9999
<b>Three Days</b>	326	3017.3	2771.8	987.6	-10.3	0.234	>0.9999
<b>Four Days</b>	326	2957.5	2745.8	977.4	-0.1	0.002	>0.9999
<b>Five Days</b>	326	2925.5	2683.8	973.7	3.6	0.082	>0.9999
<b>Six Days</b>	326	2912.9	2604.8	977.3	-	-	-

*Supplementary Table 9: RAM variable responsiveness was assessed using Wilcoxon signed rank tests, at each time point corresponding to a clinic visit (i.e. at 6, 12, 18, 24 and 30 months), compared to baseline. If the RAM variable was significantly responsive at that time point ( $p < 0.05$ ), a “+” is shown, otherwise a “-” is shown. For the other variables the median value at each time point was compared to baseline medians. If the variable had exhibited a minimum clinically important difference (MID) <sup>1-4</sup>, a “+” is shown, otherwise a “-” is shown.*

Time points (months)	6	12	18	24	30
E <sub>A</sub>	-	+	+	-	-
METs	-	+	+	+	+
PhAD	-	+	+	-	+
Steps	-	-	+	+	+
RAM score	-	-	+	+	+
EDSS <sup>1</sup>	-	+	+	-	+
MSFC <sup>2</sup>	-	-	-	-	+
T25FW <sup>2</sup>	-	-	-	-	+
9HPT <sup>2</sup>	-	-	-	-	-
PASAT <sup>2</sup>	-	-	-	-	-
MSIS-phys <sup>3</sup>	-	-	-	-	-
MSIS-psych <sup>4</sup>	-	-	-	-	-

## References

1. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;33(11):1444-52. doi: 10.1212/wnl.33.11.1444
2. Rudick RA, Polman CH, Cohen JA, et al. Assessing disability progression with the Multiple Sclerosis Functional Composite. *Multiple sclerosis (Houndmills, Basingstoke, England)* 2009;15(8):984-97. doi: 10.1177/1352458509106212 [published Online First: 2009/08/12]
3. Costelloe L, O'Rourke K, Kearney H, et al. The patient knows best: significant change in the physical component of the Multiple Sclerosis Impact Scale (MSIS-29 physical). *Journal of Neurology, Neurosurgery & Psychiatry* 2007;78(8):841-44. doi: 10.1136/jnnp.2006.105759
4. McGuigan C, Hutchinson M. The multiple sclerosis impact scale (MSIS-29) is a reliable and sensitive measure. *J Neurol Neurosurg Psychiatry* 2004;75(2):266-69.