28 February 2020
EMA/550514/2019
Committee for Medicinal Products for Human Use (CHMP)

Overview of comments received on 'Draft qualification opinion on Multiple Sclerosis Clinical Outcome Assessment (MSCOA) qualification opinion' (EMA/CHMP/SAWP/336445/2019)

Interested parties (organisations or individuals) that commented on the draft document as released for consultation.

Stakeholder no.	Name of organisation or individual
1	Accelerated Cure Project for Multiple Sclerosis
2	Person living with MS
3	Professor Alan Thompson, University College London Professor Ruth Ann Marrie, University of Manitoba Professor Jeffrey Cohen, Cleveland Clinic Foundation on behalf of the International Advisory Committee on Clinical Trials in Multiple Sclerosis
4	EFPIA
5	F. Hoffmann-La Roche Ltd
6	Biogen

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

1. General comments – overview

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
1	Accelerated Cure Project for Multiple Sclerosis and its iConquerMS People-Powered Research Network (PPRN) support the overall concept that new regulatory outcome measures are needed to evaluate the impact of pharmacological and other interventions on the entire experience of the progression of disability experienced by people living with multiple sclerosis (MS). We further support the proposal by the Multiple Sclerosis Outcomes Assessment Consortium that the evaluation of agents designed to reduce, arrest or reverse disease progression would be improved, relative to evaluation based solely through the use of the Expanded Disability Status Scale (EDSS), if the 4 performance measures (T25FW, 9-HPT, SDMT and LCLA) were employed as outcome measures. Additionally, and most importantly, we strongly support the EMA's comment that patient-reported outcome (PRO) measures, reflecting a broad experience of people living with MS, might ultimately be an alternative to (or better than) the 4 performance measures in evaluating pharmacological and other interventions designed to reduce, arrest or reverse disability progression in MS. Consistent with the work of many other groups (see, for example, references 1-5) the PROs collected since late 2014 from participants in the iConquerMS PPRN, as they register with the online network and twice yearly thereafter, reveal that fatigue, sleep disturbance, anxiety, depression, pain, cognitive impairment, upper limb and dexterity function, bladder and bowel function, stigma, and lack of satisfaction with social roles and activities, in addition to lower extremity functional mobility, are among the symptoms, disabilities and quality of life issues that trouble people living with MS and which worsen as their disease progresses (6-9). While the combination of the T25FW and the 9-HPT capture some aspects of	This comment is acknowledged but has no impact on the current qualification advice on a Clinical Outcome Assessment (COA) and no major amendment is required. See also specific comments below.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	physical function of both upper and lower limbs, PRO measures of physical function are able to capture a wider range of physical activities of daily living that matter to people living with MS and the impairment of these physical functions as MS disability progresses. For example, in the iConquerMS PPRN, using the Neuro-QoL Adult Short Form (10,11) Lower Extremity questionnaire, lower extremity physical activity data is collected in response to questions such as "Are you able to step up and down curbs?" and "Are you able to get in and out of a car?" in addition to the walking assessment question "Are you able to go for a walk of at least 15 minutes?", which is similar in physical assessment to the T25FW. Furthermore, using the Neuro-QoL Adult Short Form Upper Extremity questionnaire, upper extremity physical activity data is collected in response to questions such as "Are you able to turn a key in a lock?" and "Are you able to brush your teeth?" in addition to the question "Are you able to pick up coins from a table top?", which has asessment similarity to the 9-HPT. Our preliminary, unpublished analyses of iConquerMS data have shown that the sensitivities to MS disease progression of PRO measures of lower extremity physical function may be similar to the sensitivity of the Patient-Determined Disease Steps PRO measure, which has a high correlation with EDSS. Additionally, when collected online or through mobile apps at intervals between clinic visits, PROs can also provide a more complete longitudinal picture of the symptoms, disabilities and quality of life issues experienced by a person living with MS. As stated above, the symptoms, disabilities and quality of life issues that matter most to people living with MS cover physical, mental and social domains much broader than those captured either by EDSS or the 4 performance outcome	

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	measures that are the subject of the MSOAC submission. If PROs for physical function alone can be shown robustly to capture a wider range of physical abilities or disabilities than currently used, or proposed, outcome measures and with similar sensitivities to MS disease progression to those outcome measures, the case will be strong for continued work towards improved outcome measures for evaluating the impact of interventions on disability progression across the entire experience of people living with MS. With regard to fatigue, which matters greatly to people living with MS and which iConquerMS data (8,9) and other data show considerable worsening as MS progresses, we note that fatigue was a key secondary outcome for the recent OPTIMUM Phase III clinical trial. In the trial, fatigue data was collected with a PRO measure and preliminary reports of the trial results (for example, 12) include a demonstration of a beneficial effect of the intervention on fatigue in trial participants with relapsing forms of MS. We encourage the Agency to support research and to collaborate closely with other initiatives, such as the Critical Path Institute PRO Consortium's MS PRO Working Group, to explore PRO measures as an alternative to EDSS and the 4 performance measures, when approved, for evaluating the effect of pharmacological and other interventions on disability progression in people living with MS. 1. Kuspinar, A. and Mayo, N.E. (2013) *Health and Quality of Life Outcomes*, 11:71-90. doi:10.1186/1477-7525-11-71. 2. Williams, A.E., *et al.* (2014) *Multiple Sclerosis International*, 2014: Article ID 203183. doi:10.1155/2014/203183. 3. Fox, R.J., *et al.* (2015) *Neurodegenerative Disease Management*, 5(6 Suppl):3-10. doi:10.2217/nmt.15.55. 4. Risson, V., *et al.* (2016) *Journal of Medical Internet Research*, 18(9):e249. doi:10.2196/jmir.5805. 5. Miller, D.M., *et al.* (2019) *Journal of Patient Experience*, doi:10.1177/2374373519864011. 6. McBurney, R.N., *et al.* (2017) ACTRIMS Forum 2017. *Multiple Sclerosis Journal*, 23(Suppl 1):P174, doi.org/10.1177/1352458517693959.	

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	7. McBurney, R.N., *et al.* (2017) CMSC Annual Meeting *International Journal of MS Care*, 19(Suppl 1):QL21, doi.org/10.7224/1537-2073-19.s1.1 8. McBurney, R.N., *et al.* (2018) ACTRIMS Forum 2018. *Multiple Sclerosis Journal*, 24(Suppl 1):P216, doi.org/10.1177/1352458517750967 9. McBurney, R.N., *et al.* (2019) CMSC Annual Meeting *International Journal of MS Care*, 21(Suppl 1):QOL34, doi.org/10.7224/1537-2073-21.s1.1 10. Cella, D., *et al.* (2012) *Neurology*, 78:1860-67. doi:10.1212/WNL.0b013e318258f744 11. Miller, D.M., *et al.* (2015) *Multiple Sclerosis Journal*, 22(6):830-41. doi:10.1177/1352458515599450. https://www.jnj.com/janssen-reports-positive-top-line-phase-3-results-for-ponesimod-in-adults-with-relapsing-multiple-sclerosis	
2	Respectfully submitted: The draft Qualification Opinion is upsetting. As a person living with Relapsing MS, I am surprised that EMA could conclude that a significant worsening on validated tests of cognition, vision, dexterity, or walking speed are not critically important to my daily functioning and well-being. I live with MS every day, and I can assure the EMA reviewers, that these functions are crucially important to me for the activities that constitute my everyday life – working, maintaining my household and family roles, and enjoying recreational and social activities – and the inclusion of these in clinical trials would be an important step forward from today. It is terrifying to me to consider the possibility that any one of these functions would deteriorate by 20%. This would have a direct negative impact on me and my loved ones. As a person with MS, I also don't understand why the EMA states that these tests of critical functions can only be used to supplement the Kurtzke Expanded Disability Status Scale. I realize that the Kurtzke Scale has been used for MS clinical trials in the past, but I don't understand why the EMA cannot approve more modern, validated and sensitive tests for the next generation of MS clinical trials. There is still a very big need for better treatments, and I believe success in the future will also require better testing methods and more sensitive methods. The EDSS does not measure how I am doing well enough, as this does not change for me over time (at least for now). These newer measures seem better suited to measure aspects of my	The comment is acknowledged, however it is respectfully pointed out that it is nowhere stated that cognition, vision, dexterity, or walking speed are not important to patients with MS. The opinion is based on the assessment of the submitted data. Please refer to the summary overall discussion of this Qualification advice and specific comments below. "The concept of interest measuring disability in progressive MS is clear and not at discussion"/…/ No change required.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	MS that are important to me. Please consider the needs and opinions of people like me – people who live every day with MS and concerned about losing abilities related to cognition, vision, dexterity, and walking.	
3	Participants, and collective opinion and expertise of the MSOAC membership. We are particularly disappointed by the findings regarding the Symbol Digit Modalities Test (SDMT) and strongly advocate for it to be considered as a valid component of a multidimensional outcome measure to be used as the primary endpoint in clinical trials of MS DMTs and as a valid secondary outcome measure individually. Cognitive impairment is a critical, clinically relevant issue for persons affected by MS, [5] yet it is well established that EDSS is woefully inadequate in capturing cognition. [6] As a result of reliance on the EDSS in MS clinical trials, 25 years after the first successful trials, the field still lacks an understanding of whether and to what degree DMTs impact the worsening cognitive impairment so common among persons with MS. Incorporation of the SDMT as a clinical trial endpoint is an ideal means to correct this knowledge gap. We agree with the EMA that the SDMT does not assess the full spectrum of cognitive dysfunction in MS and would not be an appropriate measure for trials testing symptomatic benefits of therapies specifically targeting cognition. Nevertheless, the SDMT assesses sustained attention and information processing speed, the domains most commonly affected in persons with MS and ones that plays a fundamental role supporting other cognitive functions such as verbal memory and executive function. Thus, the SDMT is an appropriate endpoint for trials targeting MS-related disability more generally. The SDMT is a valid and highly reliable tool which can be administered by non-clinical personnel after brief training and is highly sensitive at detecting impairment. [3] The SDMT discriminates better than other neuropsychological tests between	The comment is acknowldeged but this qualification procedure was an evaluation of the documentation submitted. No change required as based on the submitted review of the literature, the voice of the patient study and an integrated analysis of aggregated clinical trial data which led to the conclusion that these tests can be accepted as secondary endpoints in clinical trials in comparison to functional scales. See also specific comments below.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	persons with MS and without MS. Impairment detected by the SDMT is strongly associated with brain lesion burden on MRI, whole brain and gray matter atrophy, and whole brain diffusion abnormalities. [3] In the analysis of the MSOAC pooled dataset, the SDMT correlated weakly with the EDSS, [7] emphasizing that it assesses aspects of MS-related disability not captured by the EDSS. Moreover, the SDMT was substantially more sensitive than the EDSS in detecting three-month confirmed disability worsening. The original version of the MS Functional Composite included the Paced Auditory Serial Addition Test (PASAT) as a test of sustained attention and information processing speed. The substantial experience with the PASAT in MS clinical trials support its utility, which is relevant to the SDMT. However, as summarized by Benedict et al, [3] the SDMT has a number of advantages compared to the PASAT, including better measurement properties, less prominent practice effects, and better patient acceptance. Importantly, the aspects of cognition assessed by the SDMT and the four-point change in the SDMT, as proposed in the MSOAC briefing package, are clinically relevant. [3] A four-point decline in the SDMT is associated with worsening of the Physical Component Score of the Short Form 36 by five points – a change in health-related quality of life that is accepted as clinically meaningful. [8] Similarly, SDMT scores differ between individuals who are employed without work issues, employed with work-related challenges, or unemployed due to MS. In a study of 97 persons with definite MS who were employed at baseline, those with a four-point decline in the SDMT had four-fold increased odds of acquiring work disability 3.5 years later (OR 4.2; 1.2-14.8). [9] Moreover, after accounting for age, sex, EDSS at baseline, and clinical course, the change in SDMT accounted for 50% of the variance in change in employment status.	

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	We urge the EMA to reconsider its assessment of the multidimensional outcome measure proposed by MSOAC and, particularly, the agency's assessment of the SDMT. References 1. Motl RW, Cohen JA, Benedict R, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. Mult Scler J 2017; 23(5): 704-10. 2. Feys P, Lamers I, Francis G, et al. The nine-hole peg test as a manual dexterity performance measure for multiple sclerosis. Mult Scler J 2017; 23(5): 711-20. 3. Benedict RHB, DeLuca J, Phillips G, et al. Validity of the symbol digit modalities test as a cognition performance outcome measure for multiple sclerosis. Mult Scler J 2017; 23(5): 721-33. 4. Balcer LJ, Raynowska J, Nolan R, et al. Validity of low-contrast letter acuity as a visual performance outcome measure for multiple sclerosis. Mult Scler J 2017; 23(5): 734-47. 5. Sumowski JF, Benedict R, Enzinger C, et al. Cognition in multiple sclerosis: state of the field and priorities for the future. Neurology 2018; 90(6): 278-88. 6. Cohen JA, Reingold SC, Polman CH, Wolinsky JS, for the International Advisory Committee on Clinical Trials in Multiple Sclerosis. Disability outcome measures in multiple sclerosis trials: current status and future prospects. Lancet Neurol 2012; 11(5): 467-76. 7. Goldman MD, LaRocca NG, Rudick RA, et al. Evaluation of multiple sclerosis disability outcome measures using pooled clinical trial data. Neurology 2019 (in press). 8. Strober L, DeLuca J, Benedict RHB, et al. Symbol Digit Modalities Test: A valid clinical trial endpoint for measuring cognition in multiple sclerosis. Mult Scler J 2018; Epub ahead of print 18 October 2018 doi: 10.1177/1352458518808204. 9. Morrow SA, Drake A, Zivadinov R, Munschauer F, Weinstock-Guttman B, Benedict RH. Predicting loss of employment over three years in multiple sclerosis: clinically meaningful cognitive decline. Clin Neuropsychol 2010; 24(7): 1131-45.	
4	EFPIA welcome the publication of this draft qualification opinion by the EMA and acknowledge the significant work already undertaken by both the MS Outcome Assessment Consortium (MSOAC) and the Agency to get to this stage. In addition to the detailed comments provided in section 2 below, we have the	

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	following major comments: General comments on the approach and need for alternative scales to EDSS: We are disappointed that the draft QO does not provide clear guidance for sponsors to follow in drug development programs regarding what endpoints can be used. It is unclear how EDSS, functional scales (which need to be more clearly defined by the EMA) and MSOAC Performance measures can be used together or separately to advance drug development in MS, particularly in progressive forms of the disease. EFPIA believe that the Expanded Disability Status Scale (EDSS) will continue to be used as a key endpoint in clinical trials (CTs) despite its well-known limitations as reflected in the EMA/CHMP MS guideline (EMA/CHMP/771815/2011, Rev. 2) which also recognises the need to develop alternative sensitive scales of MS related disability to address the remaining unmet needs in MS trials. We believe that the selection of the four different performance outcome measures (T25FW, 9HPT, SDMT and LCLA) for qualification is appropriate to assess the major aspects of disability progression experienced by persons with MS, but not captured well by EDSS. As supported by the significant body of evidence and the appropriate qualification approach, we strongly support the use of these outcome measures in clinical trials (CTs). We would thus welcome a clearer conclusion, with stronger regulatory endorsement on their use in future MS CTs, in the Final Qualification Opinion. We would also appreciate to see a consistent view on these performance outcome measures across regulatory jurisdictions. Concept of interest The draft Qualification Opinion highlights that the selected performance outcome measures (PerO's) do not cover fatigue, pain, sexual dysfunction and sensory	The assessment of what an effect means in terms of clinical significance is an ongoing discussion. Prospective studies that firmly

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	outcomes, which were highlighted in the patient study as being important elements. However, we agree with the proposal of MSOAC to use patient reported outcome measures (PROMs) to assess these areas and that the concept of interest i.e. 'disability in multiple sclerosis' is well covered by the four PerfO's. We believe that PROMs could be included as secondary outcome measures alongside the PerfO's, in order to provide information on the patient experience for dimensions related to disability that cannot be measured using quantitative performance tests. We do not believe that PROMs should be combined with PerfO's as a primary outcome measure, since there is no validated method to combine performance test scores with PROs, e.g. there is no validated approach to weighting patient reports compared with neurologist derived severity scores. Therefore, although it is clear that patient reports on fatigue, pain, sexual dysfunction, and sensory outcomes are important, we would support a proposal to use validated self-report instruments alongside the primary disability outcome, comprised of the PerfO's. Correlation with EDSS The draft Qualification Opinion repeatedly seeks to comment on the lack of correlation between the PerfO's and EDSS. Although we might expect convergent validity, we would not expect correlation of all the endpoints with EDSS since EDSS does not measure these concepts. There are many well-described challenges with the use of EDSS, particularly its insensitivity to cognitive, visual, or upper extremity dysfunction, and low sensitivity above EDSS 4.0 – factors which could contribute to a lack of strong correlation with other scales. For the SDMT and LCLA it is acknowledged within the Qualification Opinion that they measure different parameters than EDSS and therefore a lack of correlation is	establish this connection are limited. For the moment, in the clinical studies the tests performances on the T25W, 9HPT, LCLA and SDMT will have to be set off against simultaneous measures of functioning (e.g. MSIS, MSWS-12, PRO-developed for the interpretation of the clinical relevance and coverage of the tests). This precludes for the moment, accepting these tests as primary endpoint. See also specific comments below.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	expected between these PerfO's and EDSS. Furthermore, reference is made in the draft Qualification Opinion to a publication by Bosma et al (2012), which looked at the relationships between 1-2 year changes in T25FW and EDSS and the long-term outcome (≥ 5 years) in patient PROs of progressive MS patients. Whilst the study demonstrated that changes in T25FW and EDSS were predictors of longer-term PRO disease impact, it showed that early change in T25FW rather than EDSS was significantly associated with the longer-term impact of MS. In our view, this reference supports data submitted by MSOAC rather than undermining the reliability of the aggregated clinical trial data analysis, as stated in lines 222-223. Test battery approach and global disability score The test battery approach is fully consistent with a global disability measure, since each individual test component (e.g. cognition, vision) provides a quantitative assessment of an important dimension of MS disability. Furthermore, the test battery recommendation reflects that MS affects different persons differently (e.g. one patient may have more cognitive impairment than motor impairment, while a different patient may experience the reverse). The test battery approach provides flexibility in CTs to incorporate some, or all, of the performance outcome measures as may be appropriate for a given CT design and MS patient population depending on the purpose of the intervention, i.e. to support a symptomatic or a disease modifying claim. However, the question can be raised as to whether EMA would have been more supportive of a global disability score as a weighted sum of the 4 components? We believe that the recommended test battery approach is not drifting from the concept of interest; it is simply the MSOAC's recommended approach to use the test battery.	In the paper by Bosma: Walking speed, rather than Expanded Disability Status Scale, relates to long-term patient-reported impact in progressive MS (2012) it is stated: Also, early change on the EDSS was associated with long-term reported walking limitations, although in a less pronounced way than the long-term effects seen following early changes in T25FW assessments. No change required.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Combining the test results into a single number would introduce many complexities and would reduce simple interpretation of the result of intervention. Inclusion of functional scales in the primary endpoint & consistency with the EMA MS guideline Based on the work done, the draft qualification opinion states that the relationship between changes in test performance for both the T25FW and 9HPT and activities of daily living (ADL) are considered established (lines 229 and 262-263), or reasonably established (line 332) based on the data submitted. Given this, it is not clear why the overall conclusion states the use of these individual PerfO's as a primary endpoint to measure disability progression would also need the inclusion of additional functional scales. Furthermore, the current MS guidance in the EU also states that if new scales are accepted then it is advised to still use the EDSS as an additional secondary endpoint in order to facilitate cross comparisons with other studies. This approach seems reasonable but does not seem to be aligned with the conclusions of the draft Qualification Opinion, which seems to require not only inclusion of EDSS, but a treatment effect demonstrated by EDSS. We believe this opinion is overly restrictive and will slow progress in introducing better approaches to measure disability in MS CTs. It would be helpful if this point could be revisited and further clarified in the final Opinion. We hope EMA will reconsider its draft QO by acknowledging the substantial evidence for clinical meaningfulness of the MSOAC Performance measures, and by revising EMA guidance on the requirement to base MS studies on EDSS. We also encourage the EMA to clarify how the EDSS, performance measures, and PROs can be used in clinical trials in a scientifically sound and statistically valid way; and the applicability of the discussed measures for assessing MS disability improvement, rather than just	The relationship of these tests alone or in different combinations to functioning (e.g. MSIS or MSWS-12) to understand the clinical meaningfulness on the concept of disability has not yet been established and more work needs to be done to accept the test battery as primary endpoint. No change required.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	disability progression. Cognition parameter The draft qualification opinion includes conflicting statements on the use of speed of processing as a cognition parameter. For example, in line 99-100, the following sentence states 'Focus on speed of processing as cognition parameter needs to be more extensively justified' whereas line 343-345 states that 'Speed of information processing is important for cognitive function but whether it covers cognitive function in multiple sclerosis is not made clear'. Despite these conflicting comments, EFPIA strongly endorses the clinical meaningfulness of the SDMT as a tool to measure cognition in MS patients. This is further supported by recent publications (Sumowski et al. Neurology 2018; 90:278-288; Benedict et al. Mult Scl J 2017;23(5):721-733) which summarize the field and recommend SDMT as a valid measure of cognitive impairment. These publications make the point that processing speed is a fundamental cognitive process, and deficits in processing speed underlie other cognitive functions such as memory and executive functioning. Furthermore, within group studies have shown that slower performance on the SDMT is correlated with activities of daily living such as shopping and cooking, and employment status (Benedict et al. 2017). We would appreciate to see a stronger endorsement on the use of SDMT in the final Qualification Opinion to support the broader use of this well validated measure.	In our opinion this is not contradictory. Speed of information processing is only one aspect of cognitive function (Benedict 2017): "While measuring a construct by a single test such as the SDMT may be practical, such an approach runs the risk that the test does not fully represent the construct in question…" See also Giedraitine et al. Cognition during and after Multiple Sclerosis Relapse as assessed with the Brief International Cognitive Assessment of Multiple Sclerosis Sci Rep. 2018; 8: 8169. SDMT is acceptable as secondary endpoint but it might not reflect the whole concept of cognition. No change required. See also specific comments below.
5	Roche would thank the EMA for the opportunity to provide some feedback on the draft qualification opinion of Multiple sclerosis clinical outcome assessment (MSCOA). First and foremost, Roche wants to express its support to the contribution made by the EFPIA to the consultation. In addition to our contribution to the EFPIA	Acknowledged.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	comments, Roche would like to highlight several specific aspects and to provide supplementary considerations. The most widely recognised disease assessment tool in Multiple Sclerosis (MS) is the Expanded Disability Status Scale (EDSS). EDSS has been consistently used in clinical trials supporting approval of fifteen approved disease modifying treatments for MS, but is not standardly used in routine clinical practice. At the moment, the EDSS is the only validated outcome measurement to detect the effectiveness of clinical interventions and to monitor disability progression in MS. However, the need for alternative scales assessing disability and the development of new scales are acknowledged and supported by the European Medicines Agency (EMA) (Guideline on clinical investigation of medicinal products for the treatment of Multiple Sclerosis, EMA/CHMP/771815/2011, Rev. 2). The work of the MSOAC is encouraged by Roche, especially in trying to address the development of an endpoint aiming at characterizing disability in the mid and upper ranges of EDSS. SPMS and PPMS present a great unmet medical need, as presently only one disease-modifying treatment is available for progressive forms of MS. Development of new scales could unlock the potential for assessing treatment effect in MS patient populations that are currently underserved by available therapeutic options. In addition, we believe the MSOAC consortium has highlighted an important aspect of the MS disease continuum which could exist from early stages and it may not be evident to distinguish the transition between phases of the disease. Irrespective of the disease stage or subtype, an instrument that measures subtle changes in these patients would be of incredible value to the MS community.	Acknowledged. Whereas the EDSS might not be used standardly in routine clinical practice it use in the clinical trial setting the EDSS is widely used as alternative options are limited. It serves addressing study objectives even if not used in clinical practice. See remarks later. The development of new scales that have the potential for assessing treatment effect in MS patient populations is supported. The problem so far is however that changes in the performance test have not been anchored to a scale measuring clinical relevant change in ADL and/or disability which is the concept of interest. Moreover, disability refers to the inability to execute activities, less involvement in life situation and ability in performing social roles. The T25FW, 9HPT, LCLA and SDMT are sole performance tests. Interpretation how subtle changes in these test performance translate to an effect on daily functioning and/or disability remains

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
		difficult an this is required in support of a labelling claim.
	In the opinion, the EMA acknowledged the limitations of EDSS in patients with most unmet need today. The EMA also recognised the limitations of seeking convergent validity of LCLA and SDMT with EDSS. Regarding 9HPT and T25FW, the EMA highlighted in the main body of the text these outcomes demonstrated a clear link with function in MS patients. However, this is not reflected in the conclusion statement and no suggestion is provided on how to effectively use these two outcome measures in the context of drug development. Therefore, Roche believes a few points create a high level of confusion in the opinion.	In the voice of patient study for each dimension (ambulation, arm functioning vision , cognition) a limited number of functional questions were rated by the patient. For instance in the Mobility ADL there are 5 simple questions rated by patient on a 1-10 point severity scale. As an example 1 of the 5 questions concern difficulties getting up from the floor. Point is these the question used in the Voice of Patient study and overall scale is anchored i.e. not fully validated e.g. against the MSIS. Unfortunately the EDSS or MSIS were not measured in the Voice of patient study. Shortly, despite convergent validity with the questions, at least with respect to the ambulatory and arm functioning, the instrument used in the voice of patient study the measurements used in the Voice of Patient study itself is not fully validated/anchored.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	EMA values a global assessment scale referring to EDSS, while EDSS is a scale that focuses on ambulation As the EMA notes, EDSS is a global assessment, however it is acknowledged that EDSS scoring is driven primarily through ambulation in the mid-range and is becoming insensitive in the upper range (above 6.5). Indeed, subchapter 9 from the neurostatus, ie, the calculation of the final score, is from 3.5 and beyond heavily driven by the capacity of the patient to walk a certain distance. This is confirmed by a Danish study in which the authors concluded that there is a substantial variability and potential lack of concordance between patients reported and actual walking distance (Skjerbaek A. et al. Can we trust self-reported walking distance when determining EDSS scores? – A part of the Danish MS Hospitals Rehabilitation Study. ECTRIMS Online Library. Oct 26, 2017; 200031; P376). In a post hoc analysis on data from placebo-treated RRMS patients from four large, randomized, multicenter, phase 3 clinical trials (where sustained disability progression was defined as a ≥ 1.0-point EDSS score increase over a ≥ 3- or ≥ 6-month period), (Scott, T., et al. Relationship between Sustained Disability Progression and Functional System Scores in Relapsing-Remitting Multiple Sclerosis: Analysis of Placebo Data from Four Randomized Clinical Trials. Neuroepidemiology 2015;44:16-23) reported that worsening of the pyramidal, cerebellar, and sensory domains of the FSS most often coincided with sustained disability progression. It is of note although that pyramidal changes might be more directly linked to ambulation (monoparesis, hemiparisis), however both cerebellar (eg, ataxia) and sensory changes (eg, proprioception) impact walking ability (Kalron, A., et al. Gait characteristics according to pyramidal, sensory and cerebellar EDSS subcategories in people with multiple sclerosis. J Neurol, 2016;263(9), 1796-1801) although they may have a different impact on	Agreed. For this reason, it is unexpected that in the aggregated clinical data the correlation between change in T25FW and change in EDSS was only around 0.25. See remarks above and later. The pros and cons of the EDSS are well known and acknowledged. Therefore this will not be repeated as this is not the issue here. The issue is whether the current alternatives proposed are sufficient. As long as this is at discussion the EDSS is the anchor point as the only validated measure of progression to date as also stated by stakeholder. For this reason it is not considered strictly necessary to validate an alternative measurement of disability in MS to the EDSS especially with respect to those dimensions a priori considered remotely covered by the EDSS. However, convergent validity against for instance QOL-scale, SF-36 should be part of the validation process of such new measurement. The observation that worsening of the pyramidal, cerebellar, and sensory domains of the FSS, stated to most often coincided

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	walking characteristics and might not impact the distance of walking. It has to be acknowledged that any effort to correlate EDSS with non-ambulatory disability is thus doomed to fail, at least with an EDSS between 4.0 and 7.0. Moreover, it is therefore important to appreciate that events derived from the Functional Systems Scores are also unlikely to be captured in the final global score. As such, to require a new composite score to represent all the possible domains of MS symptomatology is to apply a more stringent bar than exists already and prevents the development of new and objective measures that are urgently needed. While recognized as the only validated measure of progression to date, the EDSS has important limitations which should be acknowledged. The EDSS is the most widely used measure of disability in MS to the extent that it has been used in almost every MS clinical trial for several decades (Scott et al., 2015). Yet EDSS is "widely disparaged as a flawed tool" (Institute of Medicine Committee on Multiple Sclerosis: Current, S., & Strategies for the, F. (2001). In J. E. Joy & R. B. Johnston, Jr. (Eds.), Multiple Sclerosis: Current Status and Strategies for the Future. Washington (DC): National Academies Press (US)), especially for patients in the mid and upper range. Reasons for this include: Inter- and intra-rater variability: Variability between (inter-) and within (intra-) raters alters the reliability of the EDSS. Inter-rater reliability kappa values between 0.32 to 0.76 for the EDSS and	with sustained disability progression may form a basis for a new assessment scale. That any effort to correlate EDSS with non-ambulatory disability is potentially doomed to fail is acknowledged but this does not imply that such a new scale should not be anchored or validated. See remark above. Where the need for alternative scales is acknowledged again it is repeated that such new scale should be validated as well in accordance to the state of art.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	between 0.23 to 0.58 for the individual functional systems were reported, suggesting inadequate agreement amongst raters (McHugh, M. L. Interrater reliability: the kappa statistic. Biochem Med (Zagreb), 2012;22(3), 276-282). Variance could be due to the subjective nature of the EDSS and reinforces the need to be trained and certified for EDSS administration. The intra-rater is also variable albeit slightly higher than the inter-rater (Hobart, J., et al. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. Brain, 2000;123 (Pt 5), 1027-1040; Meyer-Moock, S., et al. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. BMC neurology, 2014;14(1), 58). Variability is even higher for lower EDSS scores (1.0-3.5) than for higher score value (Meyer-Moock et al., 2014) Linearity: EDSS is not a linear scale; it is bimodal. Patient scores are generally at the low or higher ranges of the scale, with relatively few in the mid-ranges (Meyer-Moock et al., 2014), and patients spend more time at some levels than at others (Zurawski, J., et al. Time between expanded disability status scale (EDSS) scores. Mult Scler Relat Disord, 2019;30, 98-103). Insensitivity to change: Rates of change vary depending on baseline score. Higher rates of change are observed for patients with low baseline scores (Ravnborg, M., et al. Responsiveness of the Multiple Sclerosis Impairment Scale in comparison with the Expanded Disability Status Scale. Multiple Sclerosis Journal, 2005;11(1), 81-84) but from an	Apart for what has been mentioned it is noted that an alternative explanation for the insensitivity to change of the EDSS could be the slow progression of Multiple Sclerosis

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	EDSS score of 6, the EDSS shows very little change despite continued decline in functioning. Once a patient has limited ambulation, other aspects of the neurological exam are substantially less to not taken into account (Institute of Medicine Committee on Multiple Sclerosis: current strategies for the future, 2001) in the global scoring. From 4.0 until 5.0, EDSS final scoring is driven by ambulation distance and the assumption that such a limited distance might be associated with a certain degree of disability based on FSS. From 5.5 and beyond, the only driver for EDSS final scoring is only related to ambulation distance. This underlines the fact that using the EDSS as it is today, it would not be possible to demonstrate the clinical and real effect of a drug/intervention that would impact on disability progression/improvement (specifically in non-ambulatory related domains) in patients with higher EDSS score. One illustration of this issue is the patient with an EDSS of 7.0 - wheelchair bound, who observes a worsening/benefit in hand motor function following any intervention. Sponsors have to use another endpoint to appreciate those changes, as it is in the Oratorio Hand study, where high EDSS patients will be monitored using 9HPT. Finally, the scoring system is problematic, rather than the use of FSS, which capture a substantial amount of clinical information, however not impacting the final scoring if not related to ambulation. Responsiveness: EDSS general scoring might not detect any change (response) when patients are fully aware that they have deteriorated or improved because the domain impacted might not have a direct relation to ambulation and specifically walkable distance, whereas at low EDSS scores the patient is unaware of subtle changes in the nervous system; they can only be detected by a neurologist (Scott et al., 2015). Those subtle changes might as well not be captured by EDSS.	itself. Note that the EDSS is sensitive to pick up a deterioration due to an exacerbation and the improvement after such exacerbation. Shortly insensitive to change of the EDSS is rather relative.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Content validity: EDSS does not capture the full range of symptoms that are important in MS. Specifically, EDSS does not measure with sufficient precision cognitive impairment, although this deficiency is common to all scales based on the standard neurological exam. CEREBRAL FSS is very superficial in its qualification of the cognitive impact of MS. It has been proposed by Saccà and colleagues (Sacca, F., et al. The EDSS integration with the Brief International Cognitive Assessment for Multiple Sclerosis and orientation tests. Multiple sclerosis (Houndmills, Basingstoke, England), 2017, Aug;23(9), 1289-1296) the integration of the BICAMS within the EDSS scaling in order to bring granularity in the cognitive evaluation and avoid the underestimation of a cognitive impairment not correctly captured by the FSS; the study showed a better global evaluation of the clinical situation. Moreover, a patient with MS can be unable to sustain a full day of work due to fatigue or other factors and yet have an EDSS score of zero (Institute of Medicine Committee on Multiple Sclerosis: current strategies for the future, 2001). The MSOAC has met the requirements for a COA validation for 9HPT and for T25FW, and this should be reflected in the conclusions of the opinion along with an associated context of use The measurement of functional status and impairment is central to all aspects of clinical research on MS, and the development and validation of acceptable measures must remain a priority for MS research (Institute of Medicine Committee on Multiple	This opinion or position statement is challenged and cannot be seen unrelated to the context of use. Point is the predictability for disability is remote. It is not up to the stakeholder to determine our interpretation of the data based on the total body of data. Single item claims is not considered sufficient in support of a disability claim. It remains

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Sclerosis: current strategies for the future, 2001). Therefore, there is an urgent need for novel outcomes in MS and we believe the work of the MSOAC has demonstrated that those clinical outcome assessments proposed for the qualification have provided an adequate level of validation to be used as endpoint of disability for specific domains (ex: "hand motor function disability" for the case of 9HPT) as standalone. 9HPT and T25FW We believe that the link between function and disability has been made, especially regarding 9HPT and T25FW. The evidence provided by the MSOAC is considered adequate, as evidenced by a link between those outcome measures and the ADL of patients, and concordance in agreement with EDSS. This is acknowledged by EMA for 9HPT and T25FW in line 225-225 and line 259-263. We do not agree that a correlation coefficient of change in T25FW or change in 9HPT vs change in EDSS of 0.2 to 0.25 is considered weak in the context of a COA. The aim of a COA development is to establish its clinical relevance De Novo, and therefore high levels of correlations are not expected since this would lead to measuring a concept that is already measured by existing outcomes. 9HPT and T25FW have demonstrated a clear link with function of patients and the conclusions reflected in the core report should lead to a statement on how sponsors can use them in the future in clinical trials as standalone primary outcome measures of hand motor disability and motor function disability.	difficult to related a xx seconds worsening in T25W or 9HPT to a clinical meaningful effect an predictability of disability. For the link between outcome measures and ADL or Disability see earlier remarks. The correlation is weak. A correlation on its own is insufficient for the establish the predictive value of these outcome for disability. See also later remarks under specific comments. The value of 9HPT and T25FW as standalone primary outcome measures of hand motor disability and motor function disability is

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	SDMT While SDMT did not demonstrate a clear link with function, the poor correlation between self-reported cognitive impairment and objective tests is a common finding across most neuropsychiatric conditions. There are a number of reasons for this (i) It is well established that perceived cognitive deficits in MS are more closely correlated with mood, fatigue and anxiety than with objective cognitive performance (Strober, L. B., et al. The Perceived Deficits Questionnaire: Perception, Deficit, or Distress? International journal of MS care 2016;18(4): 183-190; Kinsinger, S. W., et al. Relationship between depression, fatigue, subjective cognitive impairment, and objective neuropsychological functioning in patients with multiple sclerosis. Neuropsychology 2010;24(5): 573-580; Oreja-Guevara, C., et al. Cognitive Dysfunctions and Assessments in Multiple Sclerosis. Frontiers in Neurology 2019;10: 581). These confounders make patient-reported measures of cognitive function a poor choice for establishing construct validity. Indeed, inability to reliably determine the existence of a cognitive impairment is the reason that neuropsychological tests are administered both in clinical and trial settings to diagnose and monitor cognitive conditions. With the endorsement of the CMSC, SDMT was recommended as a tool to detect and monitor cognitive decline in MS (Kalb, R., et al. Recommendations for cognitive screening and management in multiple sclerosis care. Multiple sclerosis (Houndmills, Basingstoke, England) 2018;24(13): 1665-1680). (ii) Performance on many cognitive assessments (including SDMT) is impacted by educational background and intellectual ability, thus high achieving patients may experience deteriorations in performance but still fall in the normal range (Feinstein, A., et al. Sub-threshold cognitive impairment in multiple sclerosis: the association with cognitive reserve. Journal of neurology 2013;260(9): 2256-2261). Whilst	questionable. See remarks throughout this document . The SDMT did not demonstrate a clear link with function which excludes it value as endpoint in support of a claim in cognition. The explanation why there was not a clear link does not change this. The same as above applies, explanation why the linkage is difficult to establish does not form an argument that the linkage between SDMT performance and cognitive functioning is there. SDMT test performance as measure o cognition should be anchored against neuro-psychological scales, as stated by the stake holder. The latter was not accomplished.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	comparison to baseline counteracts this effect in longitudinal trials, it may explain the lack of agreement in the voice of the patient study described in the briefing package. As such, findings from the voice of the patient study should not be seen to discredit the meaningfulness of impairments captured by the SDMT. Further assessment of clinical validity could be assessed against objective evidence on real world outcomes related to cognitive disability. For example, impact on employment or assessing the relationship between SDMT and caregiver-reported ADLs may also support the clinical meaningfulness of changes or specific milestones in cognitive disability. Finally, some data suggests that cognitive impairment is associated with MRI changes, namely abnormalities in cortico-thalamic tracts (in)directly related to regional thalamic atrophy (more pronounced in the anterior regions)). Confirming those assumptions, a study has been conducted in RRMS patients, with or without cognitive impairment, and demonstrated the role of thalamic involvement in cognition impairment (Bisecco, A., et al. Connectivity-based parcellation of the thalamus in multiple sclerosis and its implications for cognitive impairment: A multicenter study. Human Brain Mapping 2015;36(7): 2809-2825) as measured by SDMT (Bisecco, A., et al. Attention and processing speed performance in multiple sclerosis is mostly related to thalamic volume. Brain Imaging and Behavior 2018;12(1): 20-28) thus underlying the importance of a thorough cognitive assessment in MS population, in order to assess subclinical abnormalities. Recent data revealed that cognitive impairment in treatment naive RRMS patients with low clinical disability (mean EDSS 1.7) substantially overlapped with results from RRMS patients who had a suboptimal response to a previous DMT (mean EDSS 2.1; enrolled in ENSEMBLE and CASTING respectively). This data suggests that problems with cognitive impairment appears in newly diagnosed patients with low clinically	Not a relevant argument in this context of assessment of the SDMT as a representative outcome measure of cognitive decline in multiple sclerosis.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	disability (Benedict R. H., et al. Baseline Cognitive Functioning Using the Brief International Cognitive Assessment for MS (BICAMS) Tests in Patients With Relapsing-Remitting Multiple Sclerosis Enrolled in Phase IIIb Studies of Ocrelizumab (ENSEMBLE and CASTING). ECTRIMS Online Library. Sep 13, 2019; 278372; P1170). Subclinical abnormalities might be difficult to be recognized by the patient/caregiver/clinician but play a critical role in the evolution of the disease and have to be recognized as soon as possible with the intention to offer the patient the best options available. LCLA Although the LCLA was not correlated with patient reported visual functioning in the voice of the patient study, it was correlated with patient-ratings on a well-established and validated tool, the NEI-VFQ in the literature review. Furthermore, the MSOAC submitted evidence of meaningful change thresholds, established using gold standard methodology, and demonstrated that changes of this magnitude were associated with meaningful deterioration in the NEI-VFQ. On that basis it is not clear what additional evidence should be provided to increase confidence in these outcome measures to lead to a successful validation in the future. Moreover, trials assessing remyelination agents will be using reliable biomarker to assess the regain of function. Numerous biomarkers are under exploration. LCLA has been suggested as a functional measure of the integrity of the visual pathway, a recent study demonstrated that the degree of demyelination contributes significantly to worsening of LCLA and thus support the feasibility of using LCLA as a functional biomarker in remyelination therapy trials (Triplett, J. D., et al. Pathophysiological	The subclinical abnormalities even if recognise need to be linked to clinical cognitive decline. The connection between LCLA and ADL/function as suggested in the literature review, was not reflected in the results of the Voice of Patient study and aggregated data analysis. Considering this all for the LCLA the connection between LCLA and functionality is not considered established. In conclusion the LCLA did not demonstrate a clear link with function which excludes it value as primary endpoint in support of a claim in vision let alone disability.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	basis of low contrast visual acuity loss in multiple sclerosis. Annals of clinical and translational neurology 2018;5(12): 1505-1512). This biomarker is indeed used in a remyelination agent (clemastine) phase II randomised, double-blind, placebo-controlled trial (ReCOVER, NCT02521311). Proposed Context of Use for 9HPT and T25FW We believe the MSOAC has applied a rigorous development for their outcome measures for development of clinical outcome assessments. The concordance in agreement with EDSS change seem to have been sufficiently demonstrated for T25FW and 9HPT. In essence the development of a COA should not focus on existing scales and therefore with the current EMA conclusions we have concerns on what level of evidence would be required for novel MS endpoints; which are highly needed to address areas of unmet needs. We propose the following context of use could be supported; "9HPT can be used as a primary endpoint in patients with an EDSS below 8, in order to characterize disability progression as measured by 9HPT. A >=20% increase in 9HPT is considered clinically meaningful" (Feys, P., et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England) 2017;23(5): 711-720). An associated label claim could be "treatment X suppressed/delayed disability progression as measured by 9HPT".	Novel endpoints should be validated. To be used for a claim for disability their ability to predict disability should be shown . This requires showing convergence validity against known functional and disability scales e.g. MSIS, SF-36. The T25FW and 9HPT can be used as secondary endpoints and considered in section 5.1 in the labelling if assessed as of relevance. Note: Endpoints are usually not mentioned in the indication as per SmPC guideline.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	"T25FW can be used as a primary endpoint in patients with an EDSS below 6,5; in order to characterize disability progression as measured by T25FW. A >= 20% increase in T25FW is considered clinically meaningful" (Cohen, J. A., et al. The Clinical Meaning of Walking Speed as Measured by the Timed 25-Foot Walk in Patients With Multiple Sclerosis Walking Speed in Patients With Multiple Sclerosis Walking Speed in Patients With Multiple Sclerosis. JAMA Neurology 2014;71(11): 1386-1393). An associated label claim could be "treatment X suppressed/delayed disability progression as measured by T25FW".	The Timed 25-foot walk (T25FW), hand dexterity (9 Hole peg Test, 9HPT), visual function (Low contrast Letter acuity, LCLA) mental tests assessing processing speed (Symbol Digit Modalities Test, SDMT) tests can neither be used as single variable or in combination with each other as primary endpoint for measurement of disability without including functional scales as well in the primary endpoint. The inclusion of these tests in clinical studies as secondary endpoints in comparison to functional scales is acceptable.
6	General Comments Biogen welcomes the publication of this draft qualification opinion by the EMA and acknowledges the significant work already undertaken by both the MS Outcome Assessment Consortium (MSOAC) and the regulatory agency to get to this stage. Biogen is providing comments in response to the publicly available information in the draft qualification opinion and the additional background information submitted by the applicant which was included with the published opinion. Biogen acknowledges that EDSS will continue to remain an important tool for measuring disability progression in MS patients. However, we were disappointed that the strength and depth of the submission by MSOAC did not lead to a clear endorsement of the use of the clinically meaningful performance outcome measures (SDMT, 9HPT, LCLA) on a flexible basis in future MS studies. Indeed, the test	Acknowledged. The lack of a clear endorsement of the 9HPT LCLA and SDMT is due to the difficulty of relating improvements in these performance tests into clinically meaningfulness let alone

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	battery approach provides flexibility in clinical trials to incorporate some, or all, of the performance outcome measures as may be appropriate for a given CT design and MS patient population depending on the purpose of the intervention, i.e. to support a symptomatic or a disease modifying claim. Furthermore, consistency in the language used to describe these measures is important to all stakeholders, but particularly patients e.g. T25FW itself is a functional measure of disability – this should be reflected in the final qualification opinion. Finally, whilst outside the remit of this specific qualification opinion, we would like to see a consistent view on these PerfO's across regulatory agencies from major jurisdictions.	that a single item is representative for disability.
	General comments on the approach and need for alternative scales to EDSS. Whilst the Expanded Disability Status Scale (EDSS) continues to be used widely as a key endpoint in clinical trials, limitations of the scale, including insensitivity to detect small changes, a strong orientation towards mobility, and insensitivity to cognitive impairment, means that there is an urgent need for alternative measures of disability in MS patients. Indeed, this point is already reflected in the current EU MS guidance (EMA/CHMP/771815/2011, Rev. 2) which states that the advantages and the disadvantages of the EDSS in assessing disability are well known and that there is a recognised need for the development of alternative sensitive scales.	Referred is to the earlier comments on the EDSS under Stakeholder no 5. It is noted that T25FW, 9HPT, LCLA, SDMT performance tests considered in isolation does not cover the concept disability. Considered in combination the tests do not cover fatigue, pain, sexual dysfunction among others
	Regulatory qualification of new, more sensitive scales would also be an important step forward to expand their broader use and acceptance. This would be a positive move which would benefit all MS patients, including those with progressive MS where EDSS is least sensitive and has largely failed to detect differences between	This is part of the problem. There is a trade of between the sensitivity of a scale and interpretation of small changes in terms of clinical meaningfulness of a xx seconds or xx points improvement in the performance

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	treatment groups; and in RRMS where active arm comparisons will be increasingly required in the future, requiring more sensitive, yet clinically meaningful measures.	scale.
	In Biogen's view, the selection of the four different performance outcome (PerfO) measures (T25FW, 9HPT, SDMT and LCLA) for validation is appropriate to assess the major aspects of disability progression experienced by persons with MS, and not captured well by EDSS. The acknowledgement within the draft qualification opinion that all four of these PerfO's share the key attributes of objectivity, reproducibility, reliability and sensitivity to detect change is welcomed and supports their selection for validation and potential use as primary endpoints in clinical trials.	See earlier remarks. The position statement is not shared.
	Furthermore, the significant work done by MSOAC to validate each of these PerfO's through the literature review, the patient evaluation study and by assessing a significant body of aggregated clinical data is a comprehensive approach which further supports Biogen's own experience that these particular outcome measures are clinically meaningful in patients with MS. Given the body of data submitted in support of the clinical outcome measures, Biogen strongly endorses the potential use of these PerfO's in clinical studies, and would welcome a clearer conclusion, with stronger regulatory endorsement on their potential use, in the final qualification opinion.	Given the discussion above, from a regulatory perspective regulatory endorsement of the their potential use of the performance tests , as primary endpoint either isolated or in all possible combination cannot be given .
	General comments on specific points raised in the Qualification Opinion. Concept of interest The draft qualification opinion highlights that the selected PerfO's do not cover	

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	fatigue, pain, sexual dysfunction and sensory outcomes, which were highlighted in the patient study as being important elements. However, we agree with the proposal of MSOAC to use patient reported outcome measures (PROMs) to assess these areas and that the concept of interest i.e. "disability in multiple sclerosis" is well covered by the four PerfO's.	The position is not shared. See earlier comments
	Biogen believes that PROMs should be included as secondary outcome measures alongside the PerfO's, in order to provide information on the patient experience for dimensions related to disability that cannot be measured using quantitative performance tests. Biogen does not believe that PROMs should be combined with PerfO's as a primary clinical trial outcome measure, simply because there is no validated method to combine performance test scores with patient reported outcomes, e.g. there is no validated approach to weighting patient reports compared with neurologist derived severity scores. Therefore, although it is clear that patient reports on fatigue, pain, sexual dysfunction, and sensory outcomes are important, we would support a proposal to use validated self-report instruments alongside the primary disability outcome, comprised of the PerfO's.	Point is that impact of changes the performance test s T25FW, 9HPT should have been validated again a function scale (+/- patient reported) in order to link improvement i*n* these performance tests to a function/disability. It is repeated that this not necessarily would be the EDSS. E.g. the Multiple Sclerosis Impact Scale - 29 items (MSIS) incorporating more dimensions and is not dominantly based on ambulation.
	Test battery approach The use of each PerfO either individually or in different combinations as a test battery approach is supported. Although the tests will not cover the entire scope of the domains they represent i.e. SDMT is the best available cognitive test but it does not encompass the entire domain of cognitive function, each individual test component (e.g. cognition, vision) provides a quantitative assessment of an important dimension of MS disability. Furthermore, it is appropriate to propose using of all 4 PerfO's, because patients are affected in different ways by MS. For example,	The test battery approach is not endorsed. Initially the MSOAC was proposed a global disability score that would be built as a weighted sum of their components. This seems to drift away from the concept of interest disability. See remarks above.

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	some MS patients are affected much more by cognitive, or visual impairment than motor impairment, others are affected more by motor impairment, and still others by a combination of impairments resulting in disability. The test battery approach provides flexibility in clinical studies to incorporate some, or all, of the PerfO's as may be appropriate for the particular study design and MS patient population depending on the purpose for the intervention. By introducing this flexibility, medicine developers would have the opportunity to discuss the incorporation of these PerfO's into their MS studies with regulators on a case-by-case basis. Inclusion of functional scales in the primary endpoint Based on the work done, the draft qualification opinion states that the relationship between changes in test performance for both the T25FW and 9HPT and activities of daily living (ADL) are considered established (line 229 and 262-263), or reasonably established (line 332). Given this, it is not clear why the overall conclusion states the use of these individual PerfO's as a primary endpoint to measure disability progression would also need the inclusion of additional functional scales. Furthermore, the current MS guidance in the EU also states that if new scales are accepted then EDSS should still be included as an additional secondary endpoint in clinical trials in order to facilitate cross comparisons with other studies. This approach seems reasonable but is not aligned with the conclusions of the Qualification Opinion, which requires not only inclusion of EDSS, but a treatment effect demonstrated by EDSS. This approach is overly restrictive, will slow progress in introducing better approaches to measuring disability in MS clinical trials and should be revisited in the final opinion.	The issue is more the context of use as separate and independent primary endpoint or as part of composite primary endpoint in support of a claim of delay of progression. The battery approach is nor endorsed Initially the MSOAC was proposed a global disability score that would be built as a weighted sum of weighted sum of their components. Note that for Zinbryta the composite disability endpoint defined as baseline EDSS ≥ 3.5 and at least one of the three confirmed 24 week worsening of EDSS, or ≥ 20%

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	Correlation with EDSS The draft Qualification Opinion repeatedly seeks to comment on the lack of correlation between the PerfO's and EDSS. There are many well-described challenges with the use of EDSS, particularly its insensitivity to cognitive, visual, or upper extremity dysfunction, and low sensitivity above EDSS 4.0 – factors which could contribute to a lack of strong correlation with other scales. For the SDMT and LCLA it is acknowledged within the qualification opinion that they measure different parameters than EDSS and therefore a lack of correlation is expected between these PerfO's and EDSS. Furthermore, reference is made in the Qualification Opinion to a publication by Bosma et al (2012), which looked at the relationship between changes in T25FW and EDSS over 1-2 years and the long-term outcome (≥ 5 years) in patient reported outcomes (PRO) of progressive MS patients. Whilst the study demonstrated that changes in T25FW and EDSS were predictors of longer-term PRO disease impact, it showed that early change in T25FW rather than EDSS was significantly associated with the longer-term impact of MS. In our view, this reference supports data submitted by MSOAC rather than undermining the reliability of the aggregated clinical trial data analysis, as stated in lines 222-223. Cognition parameter The draft qualification opinion includes conflicting statements on the use of speed of processing as a cognition parameter. For example in line 99-100, the following sentence states "Focus on speed of processing as cognition parameter needs to be more extensively justified" whereas line 343-345 states that 'Speed of information processing is important for cognitive function but whether it covers cognitive	decline on Timed 25-foot Walk (T25FW), or, ≥ 20% decline on 9-Hole Peg Test (9-HPT), has been accepted based on the followimg considerations. The incorporation of the EDSS was considered mandatory, the overall results of the composite endpoint should not be driven by one of the components and all components should show a trend in ther same direction. That an effect on the performance tests still can be related to EDSS as the EDSS is measured anyway , in order to facilitate cross comparisons with other studies, is beyond. the point. As long as the performance test is needed to interpret the change in performance test the latter cannot be a standalone primary endpoint. Referred is to the earlier comments and other comments throughout this document dealing with the same issue. The argument of conflicting comments is not seen. Both further justification of speed of processing as cognitive variable as well as the

Stakeholder no. *(See cover page)*	General comment (if any)	Outcome (if applicable)
	function in multiple sclerosis is not made clear'. Despite these conflicting comments, Biogen strongly endorses the clinical meaningfulness of the SDMT as the tool to measure cognition in MS patients. This is further supported by recent publications (Sumowski et al, Neurology 2018;90:278-288 and Benedict et al, Mult Scl J; 2017 Vol 23(5) 721-733) which summarize the field and recommend SDMT as a valid measure of cognitive impairment. These publications make the point that processing speed is a fundamental cognitive process, and deficits in processing speed underlie other cognitive functions such as memory and executive functioning. Furthermore, within group studies have shown that slower performance on the SDMT is correlated with activities of daily living such as shopping and cooking, and employment status (Benedict et al. Mult Scler 2017 Apr; 23(5): 721-733). A strong endorsement on the use of SDMT, supported by the data provided by MSOAC, should be included in the final Qualification Opinion to support the broader use of this well validated measure.	representativeness of SDMT for cognition express the uncertainty of the SDMT performance test as a primary endpoint. The representativeness of the SDMT for cognitive disturbances in MS as well as clinical meaningfulness in the change of SDMT score and impact on disability needs further substantiation.

2. Specific comments on text

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
16-17	4	Comment: We assume the four tests proposed are applied in the original versions (i.e. paper, watch, Pegboard etc.). As electronic versions and apps are under development including the 4 tests, it would be helpful to clarify which versions are to be used.	Confirmed
36-41	4	Comment: The SDMT is the most accepted test to assess cognition in MS, having shown the ability to evaluate treatment effect and sensitivity to change in several MS clinical trials (see references below): Benedict, R. H., et al. Improved cognitive outcomes in patients with relapsing-remitting multiple sclerosis treated with daclizumab beta: Results from the DECIDE study. Multiple sclerosis (Houndmills, Basingstoke, England) 2018;24(6):795-804. Benedict, R.H., et al. Impact of ocrelizumab on cognition in patients at increased risk of developing progressive disease. Presented at: 32nd Annual Meeting of the Consortium of Multiple Sclerosis Centers. May 30-June 2, 2018; Nashville, Tennessee. Abstract DX67. Benedict et al. Impact of Siponimod on Cognition in Patients with Secondary Progressive Multiple Sclerosis: Results from Phase III EXPAND Study. Abstract no. 004. Oral presentation at the 70th Annual Meeting of the American Academy of Neurology, Los Angeles, CA, April 21-27, 2018. However, discrepancies between objective cognitive assessments and subjective PROs (such as the ADLs presented in the VOP study) is not surprising since loss of cognitive insight may have had an impact on these results. Depression, and to a greater extent	Referred is to the general remarks for stakeholders 5 and 6 dealing with the same comments.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		fatigue have a significant impact on the subjective evaluation of cognition (DeLuca, G. C., et al. Cognitive Impairment in Multiple Sclerosis: Clinical, Radiologic and Pathologic Insights. Brain Pathology 2015;25(1):79-98.). Since these concepts were not assessed by the MSOAC group, this could have confounded responses provided in the VOP study, leading to a weaker correlation. Such loss of insight supports the need for more objective cognition assessments, such as the SDMT. Proposed change (if any): The intent is for this COA instrument to serve as a primary, co-primary, or secondary endpoint to assess efficacy in clinical trials at various stages of drug development, including proof of concept, dose-ranging, confirmatory and registration trials. The four performance measures are considered as a battery of tests, some or all of which could be used as a dysconjugate composite endpoint by sponsors in a clinical trial. For example, the T25W measure would not be used in PPMS and SPMS trials in which participants are non-ambulatory. In the future, the outcome measures proposed by the MSOAC could be used as standalone assessments of disability for the 9HPT and T25FW. The following context of use is supported: 9HPT can be used as a primary endpoint in patients with an EDSS below 8, in order to characterize disability progression as measured by 9HPT. A >=20% increase in 9HPT is considered clinically meaningful [Feys, P., et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England). 2017;23(5):711-720]. An associated label claim could be "treatment X suppressed/ delayed disability progression as measured by 9HPT". T25FW can be used as a primary endpoint in patients with an EDSS below 6,5; in order	Again referred is to the general remarks for stakeholders 5 and 6 dealing with the same issues. Idem It is repeated that the lack of a clear endorsement of the 9HPT LCLA and SDMT as primary endpoint either separately or in all thinkable combinations lies in relating improvements in these performance test

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		to characterize disability progression as measured by T25FW. A >= 20% increase in T25FW is considered clinically meaningful (Cohen, J. A., et al. The Clinical Meaning of Walking Speed as Measured by the Timed 25-Foot Walk in Patients with Multiple Sclerosis Walking Speed in Patients With Multiple Sclerosis Walking Speed in Patients With Multiple Sclerosis. JAMA Neurology 2014;71(11):1386-1393.). An associated label claim could be "treatment X suppressed/delayed disability progression as measured by T25FW".	into clinically meaningfulness let alone for being representative for disability.
41-43	4	For example, the T25W measure would not be used in PPMS and SPMS trials in which participants are non-ambulatory. If used in registration trials, the ultimate language included in product labeling will reflect which measures were used in the trials and would describe the effect of treatment on each measure.' Comment: We suggest that this language be reflected in the conclusion to make it clear to Sponsors how this could be used to support a label claim.	This text is from the executive summary from the Consortium presented for setting the scene. The text reflect the Consortium's opinion of the context of use. It does not reflect a position statement of the EMA.
49-56	4	Consequently, the qualification of an instrument that includes SDMT would fill an unmet need; since determinantal effects on cognition accounts for much of the socioeconomic impact of MS and this dimension of MS is extremely important to PwMS. Importantly, worsening cognitive function, as measured by SDMT, occurs independently from worsening physical function, as captured by the EDSS or performance measures such as the T25FW, 9HPT and LCLA. Therefore, an instrument that measures a critical aspect of cognition with SDMT, in combination with important physical measures of ambulation, dexterity and vision, fills a measurement gap and provides a much more complete assessment of MS-related disability.'	Idem.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome *(To be completed by the Agency)*
		Comment: This should be acknowledged when evaluating SDMT. Link to function for cognitive measures is very difficult to make and alternative methods to address the impact on patient's life, such as rate of unemployment should be accepted concepts to demonstrate the impact of cognitive deficits.	For comments with respect to the SMDT see to the general remarks for stakeholders 5 and 6 dealing with the same issue.
Figure 3: Framework for developing a COA Performance Measure for MS clinical trials	4	Comment: The table provides a useful framework for developing performance outcome assessments. Although the ADLs in step 3 are ADL examples, it is not clear if these examples were selected using quotes from qualitative research carried out with patients. Furthermore, it is not clear if step 5 (subcomponents of bodily functions) are variables voiced as important to patients. For example, why speed was selected as an endpoint to measure walking ability, when in fact distance may be voiced as a more important variable. Such endpoints should be justified using evidence from qualitative research and a developed conceptual framework. The lack of conceptual framework could also provide a reason why correlations are not as strong for the SDMT and LCLA as endpoints used in these tests may not capture variables considered important to patients, whilst responses in the voice of the patient study used ADLs evidenced as important to patients. Proposed change (if any): Clarification would be appreciated.	The figure reflects the working hypothesis presented by the Consortium in 2014. The substantiation of the linkage between level 5-6 to the levels 3 and 4-6-is part of the validation procedure. Face-validity is based on assumptions which is not sufficient on its own.

91	4	Comment: Typographical change Proposed change (if any): The attractiveness of the performance tests chosen i.e. T25FW, HPT, LCLA and SDMT lies in there their objectivity, reproducibility…	This comment seems to be unfinished. See below stakeholder 6 comment.
99-100 343-345	4	Comment: Extensive literature documenting the importance of processing speed in several domains of cognition is available. Processing speed provides the underpinning for the successful use of other cognitive functions such as learning, organizing, judgement, word finding, etc. Deficits in cognitive processing speed can have a devastating effect on employability since many jobs, particularly white collar, depend on the speed of information processing for acceptable job performance. Slowed information processing can spell the difference between working and being unemployed. Slowed cognitive processing speed can interfere significantly with socialization. It can lead to difficulty participating in conversations, especially if more than two people are involved. This can lead to social isolation and difficulty in a wide variety of daily activities. We know in neuroscience that patients with cognitive impairment may not perceive it themselves, and as such it may not then be seen in PROs. We also know that even in RIS and CIS, there is reduction in thalamus volume, which correlates with SDMT worsening and correlated with unemployment. (See references below): For RIS: Azevedo, C. J., et al. Early CNS neurodegeneration in radiologically isolated syndrome." Neurology(R) neuroimmunology & neuroinflammation. 2015;2(3):e102-e102. For CIS: Henry, R., et al. Connecting white matter injury and thalamic atrophy in clinically isolated syndromes." Journal of the neurological sciences. 2009;282:61-66. For Thalamus and SDMT: Bisecco, A., et al. Attention and processing speed performance	It is acknowledged that speed of processing might be necessary condition for cognitive function but what is not clear is whether this is sufficient on its own and representative for cognitive function in MS general. This requires that changes on processing speed as measured by the SDMT is related to a functional scale. In the VOPS study this was not observed. There is a point that patients with cognitive impairment may not perceive it themselves, and as such this is not picked up in PROs. However, then this has to be established otherwise i.e. by caregiver questionnaire, neuropsychiatric cognitive tests.

		in multiple sclerosis is mostly related to thalamic volume. Brain Imaging and Behavior. 2018;12(1):20-28. Other objective data described above that shows that change in disease status can be objectively measured and do having impact on patient's lives, beyond PROs, needs to be considered in order to have measures that capture concepts like cognition that are meaningful to patients. Numerous tests are available to assess cognitive functions. SDMT is a valid measure of information processing speed (IPS), is correlated with activities of daily living (ADLs) in MS patients such as employment and driving (see references below) and has been shown to be particularly sensitive to slowed processing of information that is commonly seen in MS. (see references below): For Fatigue and Employment: Goverover, Y., et al. Factors That Moderate Activity Limitation and Participation Restriction in People with Multiple Sclerosis. American Journal of Occupational Therapy; 2015;69(2):6902260020p6902260021-6902260020p6902260029. For Driving: Schultheis, M. T., et al. Examining the Relationship Between Cognition and Driving Performance in Multiple Sclerosis. Archives of Physical Medicine and Rehabilitation. 2010;91(3):465-473. Benedict, R. H., et al. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England). 2017;23(5):721-733. Moreover, congruent conclusions have been observed when considering employment (see previous reference) and driving (see previous reference). According to experts, SDMT is the best psychometric measure available for assessing IPS in MS patients (Benedict et al. 2017). SDMT and IPS are not obviously represented within the EDSS and Functional System Scores (FSS) – Cerebral FSS lack the level of precision in the	

		definition of cognitive alterations, as well as the sensitivity to change. Proposed change (if any): Further, the domain of cognition was broader and did not only include pace of thought (SDMT) but also memory (California verbal learning test[CVRT]; Benton visual retention test [BVRT]; 7/24 Spatial Recall Test [SRT]) and attention. The focus on speed of processing as cognition parameter needs to be 100 more extensively justified. The MSOAC used information processing speed to assess cognition because processing speed is a basic, elemental cognitive function, required by, and therefore influencing downstream processes such as learning, memory, word-retrieval and executive function.	The validity of the SDMT as primary endpoint in support of general claim on cognition in MS should be further worked out.
101-108	4	'However, disability refers to the inability to execute activities, less involvement in life situation and ability in performing social roles. The T25FW, 9HPT, LCLA and SDMT are sole performance tests. How changes in test performance translate to an effect on daily functioning and/or disability remains unclear. In other words, whether the connections between the yellow boxes drawn in the figure above are substantiated by data is not clear from the above. Change in speed (T25FW, 9HPT) or scores (LCLA, SDMT) of the performance tests cannot be accepted to reflect disability at face value. Hence, whether these tests reflect the concept of interest can only be determined when the connections mentioned are further substantiated.' Comment: The reviewers have ignored extensive evidence presented documenting the clinical meaningfulness of T25W, 9HPT, LCLA, and SDMT as related to 'inability to execute activities, less involvement in life situation and ability in performing social roles.' This information was provided in 3 forms: 1) Analysis of a large amount of pooled clinical trial data provided by MS drug developers; 2) An extensive formal literature review; and 3) The voice of patient study, incorporating input from persons with MS. The evidence cited clearly documents the relationship between these performance measures and clinically meaningful disability.	Nothing is ignored. There is a difference in opinion on the context of use and whether the provided data support that context of use. 1) The almost absence of concordance in agreement of worsening of EDSS and worsening on the T25FW or

		We do not agree that the link between function and disability has not been made, especially regarding 9HPT and T25FW. The evidence provided by the MSOAC is considered adequate, as evidenced by a link between those outcome measures and the ADL of patients, and concordance in agreement with EDSS. This is acknowledged by EMA for 9HPT and T25FW in line 225-225 and line 259-263. A combination of predominantly motor and sensory symptoms causes upper limb disability, which hampers the ability to perform ADLs and social activities, resulting in a decreased quality of life (van Munster et al, Tasks of activities of daily living (ADL) are more valuable than the classical neurological examination to assess upper extremity function and mobility in multiple sclerosis, MSJ 2018). Distal upper limb dysfunction is frequently referred to as impaired manual dexterity or hand dysfunction. The 9HPT is recommended as a standard test for measuring manual dexterity in MS patients, and it can be used as reference value to investigate validity of other, newly developed upper limb outcome measures, due to its excellent psychometric properties regarding reliability, discriminant, concurrent, and ecological validity (Feys, P., et al. (2017). "The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis." Multiple sclerosis (Houndmills, Basingstoke, England) 23(5): 711-720.). In assessment of the qualitative study the EMA seems to conclude that the patient relevance has been established but later in the document concludes that the endpoint could not be used as a primary endpoint. This is inconsistent. As already outlined earlier in the document, a link between SDMT and ADLs is very difficult to make in the context of such qualitative studies. It is not clear what EMA is expecting to substantiate further the link between those performance tests and ADLs. Proposed change (if any): However, disability refers to the inability to execute daily activities. The evidence provided by the MSOAC has adequately demonstrated that 9HPT and T25FW reflect	9HPT in the aggregated data analysis is unexpected and sets doubts on the reliability of the aggregated clinical trial data analyses. 2) The intended context of use has changed from a global disability score that would be built as a weighted sum of these components into four separate measures that can be used in combination or on a single primary endpoint in support of a descriptive indication for instance *delay in accumulation of disability as measured by arm dexterity assessed by the 9 Hole Peg Test.* 3) The voice of patient study for each dimension a limited number of functional questions were rated by the patient. For instance in the Mobility ADL there are 5 simple questions rated by patient on a 1-10 point severity scale. As an example 1 of the 5 questions concern

		disability at face value, as evidenced by a link between those outcome measures and the ADL of patients, and concordance in agreement with EDSS.	difficulties getting up from the floor. Point is that despite convergent validity with function based on the voice of patient study the separate measurements used in the Voice of Patient study and overall sum scale is not anchored. Shortly despite convergent validity with function based on the voice of patient study the measurements used in the Voice of Patient study itself is not fully validated/anchored.
111-114	4	Comment: The EMA has acknowledged that it is not scientifically correct to judge the value of SDMT or LCLA based on concordance with EDSS. From a regulatory science point of view, it is also not correct to expect biomarker type correlations; by definition a tool that covers a new concept cannot be expected to correlate with a tool that adequately or not at all captures this. Direction correlation is all that can be expected in this context. Proposed change (if any): We suggest that the paragraphs on concordance with EDSS be removed from lines 280-284 and lines 307-315.	The arguments to delete the lack of concordance with the EDSS is not understood. It is a presentation of the observation followed by a discussion why this is not unexpected.
115-124	4	Comment: We agree that capturing fatigue, pain and other concepts best known to the patient are essential as part of a comprehensive measurement strategy in MS. However, many of these symptoms are highly variable and are strongly influenced by confounding factors, including mood. Such symptoms would be inconsistent with the current approach to	Not accepted. What is important stays important irrespective whether it is highly variable or not. Whether these can be primary or secondary endpoint is not the issue here.

		confirmed disability progression as a relatively permanent and persistent clinical symptom. Proposed change (if any): We propose to capture such symptoms as secondary endpoints in clinical trials, consistent with the current approach.	
125-128	4	Initially (2014), the MSOAC proposed a global disability score that would be built as a weighted sum of these its 4 components. The decision has been made not to pursue a global disability score is a change of concept i.e. the four measures are now considered as a battery of tests, all or some of could be used in a clinical trial as the primary endpoint. This seems to drift away from the concept of interest.' Comment: The test battery approach is fully consistent with a global disability measure, since each individual test component (e.g. cognition, vision) provides a quantitative assessment of an important dimension of MS disability. Furthermore, the test battery recommendation reflects that MS affects different persons differently (e.g. one patient may have more cognitive impairment than motor impairment, while a different patient may experience the reverse). Another advantage of the test battery approach is that the same tests can be used for trials of symptomatic therapy as for disease modifying therapy. Tests included in the primary outcome measure could be selected based on the purpose for the intervention. In the future we encourage an adaptive qualification approach, driven by the data. While it is acknowledged that a global disability scale would be of interest in MS, there are differences in the importance of the different domains that may lead to different definitions of a global disability score in MS patients. To address this challenge, establishing the validity of single domains is a preferred approach in the future. In the meantime, we believe the MSOAC has demonstrated the importance of the hand motor function and lower limb single domains thus far.	Referred is to the earlier comments. As an example It is not seen why this is different from an indication *delay of <u>disability</u> as measured by arm dexterity as measured by the 9 Hole Peg Test.* Hence the concept of disability should be covered.

		The recommended test battery approach is not drifting from the concept of interest; it is simply the MSOAC's recommended approach to use the test battery. Combining the test results into a single number would introduce many complexities and would reduce simple interpretation of the result of intervention. Would EMA have been more supportive of a global disability score as a weighted sum of the 4 components? Proposed change: Initially (2014), the MSOAC proposed a global disability score that would be built as a weighted sum of these its 4 components. The decision has been made not to pursue a global disability score is a change of concept i.e. the four measures are now considered as a battery of tests, all or some of could be used in a clinical trial as the primary endpoint.	See earlier comments
147-150	4	Comment: It is not clear what the following sentence means 'The value of the literature review is limited as the data dominantly concern cross-sectional data'. Moreover, since these performance measures have been studied for 20 years, and the literature review, and the review papers represent extensive support for the validity of these measures, we propose the following changes. Proposed change: *Correlations between the performance measures (or between performance measures and EDSS) are weak to moderate, though statistically significant. This strongly supports use of these performance measures together in the same trial, as they are only weakly correlated. This means that the performance measures are testing largely independent aspects of MS disability, i.e. independent from one another and from the EDSS.*	Not accepted. Statistical significance is not considered sufficient on its own. The effect size if not predictability of disability is more relevant.

159-160	4	Comment: Typographical change Proposed change (if any): 'this is the major study where the hypnotized hypothesised linkage can be substantiated'.	Already spotted and changed.
203	4	Comment: Based on the definition of disability in MS, any assessment that measures the neurological or neuropsychological impairment that limits patient's important activities of daily living should be considered as measuring disability, no matter whether it is used in studies for disease modifying therapies or symptomatic treatments. In other words, the efficacy of symptomatic treatment can be evaluated with a disability measurement. Proposed change (if any): ~~However, the context of use of the T25FW was symptomatic treatment not for assessing disability.~~	In the context of substantiating a claim of delay in disability this is not acceptable. Note that for a symptomatic improvement also the xx seconds improvement in T25FW performance has to be related to a clinical meaningful change. Note that for Fampyra full approval has been given only after could be related to a clinical meaningful change in MSWS-12 and CGI-I. Reference is made to the EPAR of Fampyra.
214	4	Comment: The instructions of the Timed 25-Foot Walk (T25FW) does not specify to control for the use of walking aids, in particular important to see changes from Baseline. Proposed change (if any): Please clarify and consider editing.	The point made here is not clear.
216-223	4	Comment: We do not agree that a correlation coefficient of 0.25 is considered weak in the context of a COA. The aim of a COA development is to establish its clinical relevance *De Novo*, and therefore high levels of correlations are not expected since this would lead to	Acknowledged. Indeed full correlation would be useless as would be no correlation at all. However it always has been argued that the EDSS dominantly focusses on ambulation. Therefore the

		measuring a concept that is already measured by existing outcomes. Reference is made to Bosma 2012, which looked at the relationships between 1-2 year changes on T25FW and EDSS and the long-term outcome (≥ 5 years) in PRO of progressive MS patients. Whilst the study demonstrated that changes in T25FW and EDSS were predictors of longer-term PRO disease impact, it showed that early change in T25FW rather than EDSS was significantly associated with the longer-term impact of MS. In our view, the conclusions of this paper are in line with the data submitted by MSOAC and does not appear to undermine the reliability of the aggregated clinical trial data analysis, as stated in line 222-223. Moreover, we disagree with the reviewers. The use of pooled clinical trial data, as opposed to smaller individual trials, ensures that no single small trial would provide misleading information. Also, MSOAC showed in the analyses presented that no single trial within the pooled clinical trial data set had a large effect on the observed correlations. As a result we recommend revising Lines 216-223. Proposed change (if any): Further whereas the correlation between the absolute values of the T25FW and absolute EDDS values is relatively high (0.39-0.62 Table 39 127/205 of the briefing document) the correlation between the change in T25FW and change in EDSS was ~~only~~ around 0.25 ((table 39 p 117/205 of the briefing document). ~~This is unexpected considering that in the paper Bosma et al. (2012) it was shown that early changes in EDSS and T25FW are independently good predictors of long term EDSS (3 years). This is what would be expected as the two scales focusing on ambulation. It set some doubt on the reliability of the aggregated clinical trial data analyses~~. ~~It set some doubt on the reliability of the aggregated clinical trial data analyses~~. *This demonstrates concordance in agreement between EDSS and T25FW.*	low correlation is unexpected. This also considering that in the literature the correlation is much stronger. Given the predictability of the T25WT for long tem EDSS based on the literature (Bosma at all 2012) this is unexpected. This is fully compatible with the remark that this finding sets some doubts on the reliability of aggregated clinical data. It is speculated that the aggregated clinical data is not as homogenous as expected. No change of the text is warranted.
225-229	4	Comment: This conclusion should be reflected in the overall conclusion and we consider that T25FW	[...] Reasonable is used and considered justified as it was not clear from the aggregated clinical data. The question

		can be used as a valid single measure of motor function disability in MS patients. Proposed change (if any): [...] Thus, the connection between T25FW test performance and functionality ~~may be~~ *is* considered ~~reasonably~~ established. *We therefore consider the evidence provided by the MSOAC supports the use of T25FW as a primary endpoint to measure disability in MS.*	whether this can be used a s single primary endpoint for a disability claim is a different one. See earlier comments**.**
240-241	4	Comment: The statement included in the draft Qualification Opinion is not supported by EFPIA. Within the literature review there was a systematic review of 9HPT by Feys [*Feys, P., et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England) 2017;23(5):711-720].* This review documented clinical meaningfulness of 15-20% change in 9HPT, using EDSS, Guys Neurological Rating Scale, MS Impact Severity Scale, and Global Disability Ratings. Indeed, a 20% change in test score is commonly used to define clinically meaningful worsening as it corresponds to pre-defined clinically meaningful changes of established clinician and patient-reported measures. A >=20% worsening threshold confirmed 3 or 6 months after the initial worsening can reliably identify subjects who are experiencing sustained progression in upper-extremity function (*Cadavid, D., et al. The EDSS-Plus, an improved endpoint for disability progression in secondary progressive multiple sclerosis. Multiple Sclerosis. 2017;23(1):94-105*). Proposed change (if any): ~~However, that a 15~~ *A* 20% difference in 9HPT is clinically relevant ~~has not been convincing demonstrated as the information in the literature review is anecdotal~~. Quantitative data that relates a change in 9HPT test performance to a change in for instance MSIS-score are not presented.	As stated the scarceness or lack of quantitative data that relates a change in 9HPT test performance to a change in for instance MSIS-score precludes this conclusion. Note that the paper of Cadavid the T25FW and 9HPT test performance was evaluated in the context of EDSS-Plus as potential primary endpoint i.e. a

			composite endpoint that incorporated the EDSS. See the comment on Zinbryta earlier.
253-257	4	Comment: We do not agree that a correlation coefficient of 0.2 is considered weak in the context of a COA. The aim of a COA development is to establish its clinical relevance *De Novo*, and therefore high levels of correlations are not expected since this would lead to measuring a concept that is already measured by existing outcomes. Proposed change (if any): Further *concordance in agreement* ~~whereas there is a rather modest correlation~~ between the *change in* ~~absolute values of the~~ 9HPT and *change in* ~~absolute~~ EDDS values *has been demonstrated (0.20 (table 40 page 128/205 of the briefing document))*.~~0.37-0.59 table 40 page 128/205 of the briefing document), the correlation between the change in 9HPT and in change in EDSS was only around 0.20 ((table 40 page 128/205 of the briefing document). Also the correlation between 9HPT test performance and Physical Component Summary (PCS) score of the SF-36 was low.~~	See earlier comments. Indeed, full correlation would be useless as would be no correlation at all. However, as this was also the case for the 9HPT test performance correlation for both EDSS as well as Physical Component summary score of the SF-36 this gets more weight. No changes are implemented.
258	4	Comment: Typographical error. T25FW should be replaced by 9HPT to avoid any confusion. Proposed change (if any): ~~T25FW summary~~ *9HPT- Summary*	Already spotted and changed.
259-263	4	Comment: We consider the evidence provided by the MSOAC supports the use of 9HPT as a primary endpoint to measure Hand Motor Function disability. We suggest this conclusion be reflected in the overall conclusions. Comparison made to T25FW is not relevant as both measure different concepts and T25FW is naturally closer	Earlier comments made for the T25FW as independent separate primary endpoint also here applies.

		to EDSS as EDSS is heavily weighted on gait. Proposed change (if any): Nevertheless, considering the literature and Voice of Patient study, for the 9HPT the connection between 9HPT test performance and functionality *is established* ~~may be considered reasonably established although to a lesser extent as compared to the T25FW. Again main weight in this assessment is given by the "Voice of the Patient" study. Thus the connection between 9HPT test performance and ADL (see figure above) may be considered established.~~ *We therefore consider the evidence provided by the MSOAC supports the use of 9HPT as a primary endpoint to measure disability in MS.*	
280-284	4	Comment: Please refer to comment on lines 111-114. We believe this paragraph is not relevant since LCLA is not expected to correlate with EDSS as EDSS poorly captures cognition in MS. If reference to the correlation must be made then this should be a clear statement with sufficient context. The reference to the briefing document should be removed unless this is also being published with the final opinion. Proposed change (if any): ~~In the analysis of aggregated clinical trial data there was limited concordance in agreement between Disability Worsening at Endpoint as defined by EDSS and worsening as defined by LCLA (Kappa coefficient around 0.10 table 31 page 120/205 of the briefing document). Correlation between LCLA and the physical component of the SF-36 is more than weak (table 41 page 41/205 of the briefing document).~~	Referred is to the earlier comments. The arguments to delete the lack of concordance with the EDSS is not understood. It is a presentation of the observation followed by a discussion why this is not unexpected.
289	4	Comment: Typographical change Proposed change (if any):	Already spotted and changed.

		Symbol digit modalities test ~~(SDTM)~~ (SDMT)	
290	4	Comment: Typographical change needed Proposed change (if any): ~~SDTM~~ SDMT literature review	Already spotted and changed.
294- 300	4	Comment: MSOAC used information processing speed to assess cognition because processing speed is a basic, elemental cognitive function, required by, and therefore influencing, downstream processes such as memory, executive functioning and language. Extensive literature documenting the importance of processing speed in several domains of cognition, and the extensive consensus in the field about the primacy of mental processing speed as the best single test to capture neuropsychological disability in MS patients has been provided. It is acknowledged that the link of SDMT with ADLs has not been demonstrated longitudinally. However, such a link is very difficult to make for a measure of cognitive function. The choice of SDMT is supported as one of the best outcome measures of cognition in MS in order to evaluate treatment effect since it has shown sensitivity to change and sensitivity to treatment in several MS clinical trials (Some references below: • Benedict, R. H., et al. Improved cognitive outcomes in patients with relapsing-remitting multiple sclerosis treated with daclizumab beta: Results from the DECIDE study. Multiple sclerosis (Houndmills, Basingstoke, England) 2018;24(6):795-804. • Benedict, R.H., et al. Impact of ocrelizumab on cognition in patients at increased risk of developing progressive disease. Presented at: 32nd Annual Meeting of the Consortium of Multiple Sclerosis Centers. May 30-June 2, 2018; Nashville,	Referred is to the earlier comments. Without longitudinal data validation is incomplete Sensitivity to change and sensitivity to treatment is not sufficient on its own. The relevance of a mean 1 point difference in SDMT score, responder rates remains to be established. See earlier comments.

		Tennessee. Abstract DX67. • Benedict et al. Impact of Siponimod on Cognition in Patients With Secondary Progressive Multiple Sclerosis: Results From Phase III EXPAND Study. Abstract no. 004. Oral presentation at the 70th Annual Meeting of the American Academy of Neurology, Los Angeles, CA, April 21-27, 2018).	
		In addition, the Benedict paper (2017) referred to in the draft Opinion is supportive of an association between SDMT decline and employment. While further work would be needed to fully validate SDMT as a primary endpoint for MS cognition, we would like to support the use of SDMT as one component of a global assessment scale in the future, pending further validation work demonstrating impact on ADL.	With the exception of rare diseases, the use of a primary endpoint implies as principle that that endpoint isolated or as a component of a composite endpoint is fully validated. This is not the case here.
		While SDMT did not demonstrate a clear link with function, the poor correlation between self-reported cognitive impairment and objective tests is a common finding across most neuropsychiatric conditions. There are a number of reasons for this (i) It is well established that perceived cognitive deficits in MS are more closely correlated with mood, fatigue and anxiety than with objective cognitive performance (some references below: • Strober, L. B., et al. The Perceived Deficits Questionnaire: Perception, Deficit, or Distress? International journal of MS care 2016;18(4):183-190. • Kinsinger, S. W., et al. Relationship between depression, fatigue, subjective cognitive impairment, and objective neuropsychological functioning in patients with multiple sclerosis. Neuropsychology 2010;24(5): 573-580. • Oreja-Guevara, C., et al. Cognitive Dysfunctions and Assessments in Multiple Sclerosis. Frontiers in Neurology 2019;10:581.	See earlier remarks. Further it raised the question whether then the focus should not be on mood, fatigue and anxiety rather than cognition.
		These confounders make patient-reported measures of cognitive function a poor choice for establishing construct validity. Indeed, inability to reliably determine the existence of a cognitive impairment is the reason that neuropsychological tests are administered both in clinical and trial settings to diagnose and monitor cognitive conditions. With the endorsement of the CMSC, SDMT was recommended as a tool to detect and monitor cognitive decline in MS (*Kalb, R., et al. Recommendations for cognitive screening and*	See earlier remarks. There is a point that patients with cognitive impairment may not perceive it themselves, and as such this is not picked up in PROs. However, then this has to be established

		management in multiple sclerosis care. Multiple sclerosis (Houndmills, Basingstoke, England) 2018;24(13):1665-1680). (ii) Performance on many cognitive assessments (including SDMT) is impacted by educational background and intellectual ability, thus high achieving patients may experience deteriorations in performance but still fall in the normal range (*Feinstein, A., et al. Sub-threshold cognitive impairment in multiple sclerosis: the association with cognitive reserve. Journal of neurology 2013;260(9):2256-2261*). Whilst comparison to baseline counteracts this effect in longitudinal trials, it may explain the lack of agreement in the voice of the patient study described in the briefing package. As such, findings from the voice of the patient study should not be seen to discredit the meaningfulness of impairments captured by the SDMT. Further assessment of sensitivity to change and meaningfulness of change could be explored longitudinally in relation to clinician and caregiver assessments of cognitive impairment to determine clinical validity.	otherwise i.e. by a caregiver questionnaire, neuropsychiatric cognitive tests.
299-300	4	Comment: It is stated that, "Moreover, SDMT performance can be influenced e.g. by visual acuity and ocular motor functions and there are learning effects (Benedict 2017)." As reported in the literature, the SDMT may be influenced by some incidental learning of symbol–digit associations. Alternative versions were therefore developed and were shown to yield nearly identical results to the original form while maintaining good test-retest reliability, in healthy subjects and in MS patients (Drake *et al.* 2010) and are recommended to be used in the clinical trial setting. We believe that the planned use of alternate version can help control and minimize learning effects from repeated administration of the SDMT. Proposed change: Moreover, SDMT performance can be influenced e.g. by visual acuity and ocular motor functions and there are learning effects (Benedict 2017). *Learning effects from repeated use of the SDMT can be minimized by the utilisation of alternate versions of this instrument.*	Acknowledged. However, it is not seen why this is an argument that the text should be amended as suggested. It does not take away stated uncertainties with respect to the SDMT stated earlier.

301-305	4	Comment: For SDMT Voice of the Patient, the draft opinion states: "Based on the Voice of Patient Study the correlation between Cognitive Functioning and SDMT score was weak to modest (see table 12, page 75/205, figure 13, page 76/205 of the briefing document). A linear relationship between SDTM and patient related level of interference in daily activities could not be established." As stated in the background-briefing package that is provided on page 22 of 46 in the draft opinion, the Consortium acknowledged the apparent lack of ecological validity of the SDMT as one of its shortcomings. However, the Consortium further stated that the task entailed in the SDMT does not resemble anything familiar to most people, although Patients with MS will often report symptoms that are suggestive of processing speed problems, e.g., inability to do things as quickly as before, "brain-fog", etc. Furthermore, it was noted that despite its lack of intuitive significance, the clinical relevance and meaningfulness of the SDMT has been amply documented in the literature along with estimates of what constitutes a clinically meaningful change or difference. Scores on the SDMT are correlated with instrumental activities of daily living such as cooking, managing finances, and using the Internet. Among cognitive measures, the SDMT is the best predictor of employment status. A 3 or 4 point difference on the SDMT reliably discriminates those who stopped work from those still working. In the course of a relapse, scores on the SDMT are likely to decline by 2 or 3 points and in one study stable vs. relapsing patients with MS differed by 5 points on the SDMT. Lastly, the SDMT has been shown to be sensitive to the effects of MS disease-modifying therapies based on a 3 or 4-point difference. In summary, the Consortium concluded that the review of the SDMT has shown that this simple, quick and inexpensive test, among the brief cognitive tests available, stands out as offering the best array of the qualities desired in a measure of cognitive function for use in MS trials. Moreover, the literature provides strong support for the clinical	This opinion statement is shared in of context the use of SDMT as secondary endpoint. This position is not shared the context of use i.e. as primary endpoint either independently or as part of a composite endpoint.

		meaningfulness of a 3 to 5 point change or difference. With respect to SDMT, we concur with the Consortium's conclusion (found on page 26 of 46 of the draft opinion). "This relative lack of alignment should not lead to the interpretation that the measure is not clinically meaningful, as the literature data demonstrate otherwise, but only that PwMS do not relate scores on two unfamiliar tests to interference with their ADLs related to those disease dimensions. We therefore encourage its use (either singly or as part of a composite instrument) in future MS clinical trials.	
307- 315	4	Comment: Please refer to comment on lines 111-114. We believe this paragraph is not relevant since SDMT is not expected to correlate with EDSS as EDSS poorly captures cognition in MS. We believe the poor correlation between the SDMT and SF-36 mental component is not unexpected since the subdomains and associated items that comprise the mental domain score do not focus on cognitive ability. The subdomains that form the mental component of the SF-36 consist of vitality, social functioning, emotional and mental health. Given these consist of items referring to tiredness, social extent and mood, it is not surprising such differing concepts do not strongly correlate with a performance outcome assessing an element of cognition. The literature also lacks consistency when correlating the SDMT with the SF-36 mental component score with some studies showing a correlation, whilst others do not (*Baumstarck-Barrau, K., et al. (2011). "Cognitive function and quality of life in multiple sclerosis patients: a cross-sectional study." BMC neurology 11: 17-17.*). Given the MSOAC group did not collect information on patient education, it is difficult to compare these results with those previously published, since this could be considerably different between studies and hence influence the results. Proposed change (if any): ~~Based on the analysis of aggregated clinical trial data there is no concordance in agreement between Disability Worsening at Endpoint as defined by EDSS and worsening~~	This paragraph describes the analysis provided by the Applicant and should therefore be kept. We are of the opinion that a stronger correlation with the mental component of the SF-36 would have been expected. The point remains that in order to accept the SDMT as a primary endpoint of cognition it should be anchored against known cognition assessments Not accepted and no change required.

		~~as defined by SDMT (Kappa 309 coefficient around 0. (See table 31 page 120/205 of the briefing document). Correlation between the absolute values of the SDMT and absolute EDDS values was modest at best 0.34 (table 38 p 125/205 311 of the briefing document). However, the correlation between change in SDMT and change in EDSs was less i.e. 0.12. This is not unexpected as the correlation between EDSS and SDTM a priori is remote as the EDDS has no cognitive dimension. More important is the modest correlation between SDMT and the mental component of the SF-36 (table 38 p 125/205 of the briefing document) as here a stronger correlation is expected[CS{12]~~ *Although a modest correlation between the SDMT and the mental component of the SF-36 is presented (table 38 p 125/205 of the briefing document), the published literature presents inconsistent results regarding an association between these measures, which could be due to differences in the population studies and education. Furthermore, as the mental component score includes items used to assess concepts unrelated to cognitive ability, it is unsurprising a strong correlation is not presented here.*	
307- 315	4	Comment: Please refer to comment on lines 111-114. We believe this paragraph is not relevant since SDMT is not expected to correlate with EDSS as EDSS poorly captures cognition in MS. We believe the poor correlation between the SDMT and SF-36 mental component is not unexpected since the subdomains and associated items that comprise the mental domain score do not focus on cognitive ability. The subdomains that form the mental component of the SF-36 consist of vitality, social functioning, emotional and mental health. Given these consist of items referring to tiredness, social extent and mood, it is not surprising such differing concepts do not strongly correlate with a performance outcome assessing an element of cognition. The literature also lacks consistency when correlating the SDMT with the SF-36 mental component score with some studies showing a correlation, whilst others do not (*Baumstarck-Barrau, K., et al. (2011). "Cognitive function and quality of life in multiple sclerosis patients: a cross-sectional study." BMC neurology 11: 17-17.*). Given the MSOAC group did not collect information on patient education, it is difficult to	This paragraph describes the analysis provided by the Applicant and should therefore be kept. We are of the opinion that a stronger correlation with the mental component of the SF-36 would have been expected. Not accepted and no change required.

		compare these results with those previously published, since this could be considerably different between studies and hence influence the results. Proposed change (if any): ~~Based on the analysis of aggregated clinical trial data there is no concordance in agreement between Disability Worsening at Endpoint as defined by EDSS and worsening as defined by SDMT (Kappa 309 coefficient around 0. (See table 31 page 120/205 of the briefing document). Correlation between the absolute values of the SDMT and absolute EDDS values was modest at best 0.34 (table 38 p 125/205 311 of the briefing document). However, the correlation between change in SDMT and change in EDSs was less i.e. 0.12. This is not unexpected as the correlation between EDSS and SDTM a priori is remote as the EDDS has no cognitive dimension. More important is the modest correlation between SDMT and the mental component of the SF-36 (table 38 p 125/205 of the briefing document) as here a stronger correlation is expected[CS{12]~~ *Although a modest correlation between the SDMT and the mental component of the SF-36 is presented (table 38 p 125/205 of the briefing document), the published literature presents inconsistent results regarding an association between these measures, which could be due to differences in the population studies and education. Furthermore, as the mental component score includes items used to assess concepts unrelated to cognitive ability, it is unsurprising a strong correlation is not presented here.*	
316	4	Comments: Typographical change Proposed change (if any): ~~SDTM~~ SDMT summary	Already spotted and changed

316-322	4	Comments: Please refer to comment on line 294-300. Proposed change (if any): Thus the connection between SDMT and ADL/function as suggested by the literature review was not reflected in the results of the Voice of Patient study and aggregated data analysis. Considering this all for the SDMT the connection between SDMT and functionality is not considered established. *However, SDMT is the strongest predictor of major socioeconomic outcomes, such as employment, independent living with a direct impact on ADL (Benedict, R. H., et al. (2017). "Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis." Multiple sclerosis (Houndmills, Basingstoke, England) 23(5): 721-733.). Further assessment of clinical validity could be assessed against objective evidence on real world outcomes related to cognitive disability. For example, impact on employment or assessing the relationship between SDMT and caregiver-reported ADLs may also support the clinical meaningfulness of changes or specific milestones in cognitive disability.*	Addition not accepted since these data were not provided. Speed of information processing is only one aspect of cognition. *Benedict, R. H., et al. (2017*) also states "While measuring a construct by a single test such as the SDMT may be practical, such an approach runs the risk that the test does not fully represent the construct in question…" See also Giedraitine et al. Cognition during and after Multiple Sclerosis Relapse as assessed with the Brief International Cognitive Assessment of Multiple Sclerosis Sci Rep. 2018; 8: 8169. SDMT is acceptable as secondary endpoint but it might not reflect the whole concept of cognition. No change required. See also general comments above.
322-326	4	Comment: In the future, we encourage a qualification process supporting an adaptive approach, driven by the data. While it is acknowledged that a global disability scale would be of interest in MS, there are differences in the importance of the different domains that may lead to different definitions of a global disability score in MS patients. Furthermore, the	Comment acknowledged but no change required for the current qualification which was not for symptomatic treatment but for the concept of disability.

		test battery recommendation is based on the variable effects of MS on different persons (e.g. one patient may have more cognitive impairment than motor impairment, while a different patient may experience the reverse). To address this challenge, starting to establish the validity of single domains would be a preferred approach and we believe the MSOAC has demonstrated the importance of those single domains thus far, especially for 9HPT and T25FW. Another advantage of the test battery approach is that the same tests can be used for trials of symptomatic therapy as for disease modifying therapy. Tests included in the primary outcome measure could be selected based on the purpose for the intervention.	
327-330	4	Comment: There is no validated method to combine performance test scores with patient reported outcomes, and no validated approach to weighting patient reports compared with neurologist derived severity scores. Therefore, we agree that that capturing fatigue, pain and other concepts best known to the patient are essential as part of a comprehensive measurement strategy in MS. However, many of these symptoms are highly variable and are strongly influenced by confounding factors, including mood. Such symptoms would be inconsistent with the current approach to confirmed disability progression as a relatively permanent and persistent clinical symptom, which assumes that once disability has occurred it is permanent if untreated. We propose that such symptoms should be captured as secondary endpoints in clinical trials, consistent with the current approach. Proposed change (if any): The T25FW, 9HPT, LCLA, SDMT tests do not incorporate fatigue, pain, sexual dysfunction and sensory outcomes. These impairments are also considered important by the consortium but thought to be better covered by PRO measures. ~~However, this raised the question if the concept of interest i.e. "disability in multiple sclerosis" or impact on ADL is fully covered~~ *However, many of these symptoms are highly variable and are strongly influenced by confounding factors, including mood. Such symptoms would be inconsistent with the*	Deletion not accepted see lines 121-124: "This begs the question whether a general questionnaire e.g. Multiple Sclerosis Impact Scale- 29 items (MSISI) incorporating all these dimensions is not an alternative way forward." No change deemed necessary.

		current approach to confirmed disability progression as a relatively permanent and persistent clinical symptom, which assumes that once disability has occurred it is permanent if untreated. Such symptoms should be captured as secondary endpoints in clinical trials, consistent with the current approach.	
334-336	4	Comment: We do not agree that a correlation coefficient of 0.2-0.25 is considered weak in the context of a COA. The aim of a COA development is to establish its clinical relevance De Novo, and therefore high levels of correlations are not expected since this would lead to measuring a concept that is already measured by existing outcomes. Therefore, it is not scientifically valid to expect a high correlation with EDSS for clinical outcome assessments in MS, provided they have demonstrated their clinical relevance De Novo which is the case for 9HPT and T25FW through the VOP study Proposed change (if any): ~~The almost absence of concordance in agreement of worsening of EDSS and worsening on the T25FW or 9HPT in the aggregated data analysis is unexpected and sets doubts on the reliability of the aggregated clinical trial data analyses.~~	Not accepted since the submitted Aggregated Data Analysis of clinical studies did not confirm the data of the literature review and the Voice of the patients study. Information should be kept.
343-348	4	Comment: In our view, there is a considerable level of experience with the use of the SDMT as an endpoint in clinical studies of MS patients, described below: Analysis of the impact of ocrelizumab on a 4-point sustained worsening in SDMT in patients with RRMS showed that, consistent with other outcomes from the trial, ocrelizumab has a statistically significant benefit on sustained 4-point worsening on SDMT. This is considered to be an important, clinically meaningful result. Furthermore, SDMT has been incorporated widely into RRMS studies conducted by Biogen since 2005. SDMT was used successfully as an exploratory endpoint in EXPAND, a siponimod SPMS Phase 3 by Novartis. Again, this is considered to be a clinically important test with a 4-point change reflecting a clinically meaningful change.	Deletion not accepted since so far data were not sufficient to justify an indication claim.

		Proposed change (if any): ~~So far there is limited experience with the SDMT as endpoint in clinical studies in MS. Speed of information processing is important for cognitive function but whether it covers cognitive function in 345 multiple sclerosis is not made clear. The quality of cognitive processing e.g. executive functioning is not assessed. Whereas inclusion of cognitive impairment scales as endpoint in MS trials is generally endorsed the usefulness/validity/relevance of the SDMT as representative measure for cognitive 348 function is still at discussion.~~	
349-353	4	Comment: The distinction between RRMS and SPMS is increasingly blurred. Although the draft opinion refers to the fact that only data from RRMS patients with the SDMT was available, support for the sensitivity of the SDMT in SPMS and PPMS is provided through literature reviews (reference page 41 of the posted document). Proposed change (if any): Apart from that, literature data (Borghi et al., Front Hum Neurosci 2016) suggest differences in cognitive scoring as assessed by PASAT for patients affected with different courses of the disease (SPMS vs. RRMS). ~~The transferability to other MS forms (SPMS and PPMS) needs to be justified since~~ *Although* only data in RRMS patients are available for the SDMT *the transferability of the SDMT to other MS forms (SPMS and PPMS) is supported by several literature references.* Only one randomized double- blind controlled study was analysed (ADVANCE) that contained data on both the SDMT and the PASAT.	The classical distinction between RRMS, SPMS and PPMS is primarily based on phenotype. It is agreed that there are no clearer demarcation criteria that mark the transition from RRMS to SPMS. There is a general recommendation to use the term activity and progression as meaningful descriptors as modifiers as basis for describing MS instead (Lublin 2014). In this sense the first sentence is acknowledged. However, this does not mean that the cognitive disturbances in the earlier stage of MS are the same as compared to the cognitive disturbances at a later stage. Therefore the principle that the transferability form the early to late stage should be clear still applies. Change proposed *Apart from that, literature data (Borghi et al., Front Hum Neurosci 2016) suggest differences in cognitive scoring*

			as assessed by PASAT for patients affected with different courses of the disease (SPMS vs. RRMS). The transferability to later stage MS needs to be justified since only data early stage patients are available for the SDMT. Only one …
389	4	Comment: Typographical change Proposed change: The attractiveness of the performance tests chosen lies in ~~there~~ their objectivity, reproducibility, reliability...	Already spotted and changed.
390	4	Comment: Typographical change Proposed change: They lack the subjectivity of a PRO (e.g. MSIS)	Typographical change accepted.
410-423	4	Comments: 9HPT and T25FW We believe that the link between function and disability has been made, especially regarding 9HPT and T25FW. The evidence provided by the MSOAC is considered adequate, as evidenced by a link between those outcome measures and the ADL of patients, and concordance in agreement with EDSS. This is acknowledged by EMA for 9HPT and T25FW in line 225-225 and line 259-263. We do not agree that a correlation coefficient of change in T25FW or change in 9HPT vs	Not accepted since the aim is to measure disability. See lines 405-408: "/…/the relationship of these test performances either as single test or in different combinations to functioning (e.g. MSIS, MSWS-12, PRO-developed) and thus the interpretation of the clinical relevance of the test performances remains to be

		change in EDSS of 0,2 to 0,25 is considered weak in the context of a COA. The aim of a COA development is to establish its clinical relevance De Novo, and therefore high levels of correlations are not expected since this would lead to measuring a concept that is already measured by existing outcomes. 9HPT and T25FW have demonstrated a clear link with function of patients and the conclusions reflected in the core report should lead to a statement on how sponsors can use them in the future in clinical trials as standalone primary outcome measures of hand motor disability and motor function disability. SDMT While SDMT did not demonstrate a clear link with function, the poor correlation between self-reported cognitive impairment and objective tests is a common finding across most neuropsychiatric conditions. There are a number of reasons for this (i) It is well established that perceived cognitive deficits in MS are more closely correlated with mood, fatigue and anxiety than with objective cognitive performance (some reference below: Strober, L. B., et al. The Perceived Deficits Questionnaire: Perception, Deficit, or Distress? International journal of MS care 2016;18(4):183-190. Kinsinger, S. W., et al. Relationship between depression, fatigue, subjective cognitive impairment, and objective neuropsychological functioning in patients with multiple sclerosis. Neuropsychology 2016; 24(5):573-580. Oreja-Guevara, C., et al. Cognitive Dysfunctions and Assessments in Multiple Sclerosis. Frontiers in Neurology 2019;10:581. These confounders make patient-reported measures of cognitive function a poor choice for establishing construct validity. Indeed, inability to reliably determine the existence of a cognitive impairment is the reason that neuropsychological tests are administered both in clinical and trial settings to diagnose and monitor cognitive conditions. With the endorsement of the CMSC, SDMT was recommended as a tool to detect and monitor cognitive decline in MS [Kalb, R., et al. Recommendations for cognitive screening and	established. This precludes for accepting the tests as primary endpoint in support of an effect on disability." A meaningful assessment of the results on EDSS or correlation with function is still required. Since the T25W measures walking the low correlation with the EDSS is not understood.

		management in multiple sclerosis care. Multiple sclerosis (Houndmills, Basingstoke, England) 2018;24(13):1665-1680]. (ii) Performance on many cognitive assessments (including SDMT) is impacted by educational background and intellectual ability, thus high achieving patients may experience deteriorations in performance but still fall in the normal range (*Feinstein, A., et al. Sub-threshold cognitive impairment in multiple sclerosis: the association with cognitive reserve. Journal of neurology 2013;260(9) 2256-2261*). Whilst comparison to baseline counteracts this effect in longitudinal trials, it may explain the lack of agreement in the voice of the patient study described in the briefing package. As such, findings from the voice of the patient study should not be seen to discredit the meaningfulness of impairments captured by the SDMT. Further assessment of clinical validity could be assessed against objective evidence on real world outcomes related to cognitive disability. For example, impact on employment or assessing the relationship between SDMT and caregiver-reported ADLs may also support the clinical meaningfulness of changes or specific milestones in cognitive disability. Finally, some data suggests that cognitive impairment is associated with MRI changes, namely abnormalities in cortico-thalamic tracts (in)directly related to regional thalamic atrophy (more pronounced in the anterior regions). Confirming those assumptions, a study has been conducted in RRMS patients, with or without cognitive impairment, and demonstrated the role of thalamic involvement in cognition impairment (*Bisecco, A., et al. Connectivity-based parcellation of the thalamus in multiple sclerosis and its implications for cognitive impairment: A multicenter study. Human Brain Mapping 2015;36(7):2809-2825)* as measured by SDMT (*Bisecco, A., et al. Attention and processing speed performance in multiple sclerosis is mostly related to thalamic volume. Brain Imaging and Behavior 2018;12(1):20-28*) thus underlying the importance of a thorough cognitive assessment in MS population, in order to assess subclinical abnormalities. Subclinical abnormalities might be difficult to be recognized by the patient/caregiver/clinician but play a critical role in the evolution of the disease and have to be recognized as soon as possible with the intention to offer the patient the best options available.	

		LCLA Although the LCLA was not correlated with patient reported visual functioning in the voice of the patient study. it was correlated with patient-ratings on a well-established and validated tool, the NEI-VFQ in the literature review. Furthermore, the MSOAC submitted evidence of meaningful change thresholds, established using gold standard methodology, and demonstrated that changes of this magnitude were associated with meaningful deterioration in the NEI-VFQ. On that basis it is not clear what additional evidence should be provided to increase confidence in these outcome measures to lead to a successful validation in the future. Moreover, trials assessing remyelination agents will be using reliable biomarker to assess the regain of function. Numerous biomarkers are under exploration. LCLA has been suggested as a functional measure of the integrity of the visual pathway, a recent study demonstrated that the degree of demyelination contributes significantly to worsening of LCLA and thus support the feasibility of using LCLA as a functional biomarker in remyelination therapy trials (*Triplett, J. D., et al. Pathophysiological basis of low contrast visual acuity loss in multiple sclerosis. Annals of clinical and translational neurology 2018;5(12):1505-1512)*. This biomarker is indeed used in a remyelination agent (clemastine) phase II randomised, double-blind, placebo-controlled trial (ReCOVER, NCT02521311). *Role of composite endpoint* Revised wording is proposed below to reflect the above comments. However, notwithstanding this we do also recognise the EMA's current position that the PerfO's could be included in a composite primary endpoint if a meaningful assessment of the results on EDSS or a correlation with function is possible by not stopping double blind treatment and follow-up after progression on other elements of the composite. In this case, clinical investigations would need to plan for an adequate number of EDSS-events (but not necessarily basing the formal power calculation on EDSS). We find this an encouraging approach which would warrant further discussion. A retrospective analysis looking at clinical trial results from the EXPAND study assessing the clinical efficacy of	Further dialogue is appreciated.

		siponimod in SPMS patients combining EDSS and SDMT as a composite endpoint resulted in more progression events leading to higher sensitivity to detect treatment effect (Kappos, 2019 EAN Platform Presentation (EPR2075). The two endpoints (EDSS and SDMT) captured progression in two different domains in a similar frequency; none of these occurred preferentially in a group that could be defined by standard baseline characteristics. Combining SDMT and EDSS resulted in more progression events leading to higher sensitivity to detect treatment effect. This suggests that the composite endpoint confers more statistical power to assess differences allowing for lower sample sizes in future clinical trials, which should encourage MS drug development. We would encourage further dialogue on the matter of the individual PerfO's and potential use as a composite endpoint. Proposed change(s): ~~While the validation work is acknowledged, the Timed 25-foot walk (T25FW), hand dexterity (9 Hole peg Test, 9HPT), visual function (Low-contrast Letter acuity, LCLA) and mental tests assessing processing speed (Symbol Digit Modalities Test, SDMT) can neither be used as single variable or in combination with each other as primary endpoint for measurement of disability without including functional scales as well in the primary endpoint. They could be included in a composite primary endpoint provided that a meaningful assessment of the results on EDSS or correlation with function is possible by not stopping double blind treatment and follow-up after progression on other elements of the composite and planning for an adequate number of EDSS events (but not necessarily basing the formal power calculation on EDSS). All components should contribute to the overall effect and the overall effect should not be predominantly driven by the performance tests. It is considered that subjects, after meeting the composite event, should be followed up for all the components of the composite endpoint.The inclusion of these tests in clinical studies as secondary endpoints in comparison to functional scales is accepted.~~	Change not accepted since it is the aim to measure and confirm disability progression which is complex. For the time being functional scales should also be included in the primary endpoint.

		The MSOAC has applied a rigorous development for their outcome measures for development of clinical outcome assessments. *The evidence provided by the MSOAC is considered adequate for T25FW and 9HPT, as evidenced by a link between those clinical outcome assessments and the ADLs of patients, as well as concordance in agreement with EDSS.* *We therefore propose the following context of use:* *"9HPT can be used as a primary endpoint in patients with an EDSS below 8, in order to characterize disability progression as measured by 9HPT. A >=20% increase in 9HPT is considered clinically meaningful" (Feys, P., et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England) 2017;23(5):711-720). An associated label claim could be "treatment X suppressed/delayed disability progression as measured by 9HPT"* *"T25FW can be used as a primary endpoint in patients with an EDSS below 6,5; in order to characterize disability progression as measured by T25FW. A >= 20% increase in T25FW is considered clinically meaningful" (Cohen, J. A., et al. The Clinical Meaning of Walking Speed as Measured by the Timed 25-Foot Walk in Patients With Multiple Sclerosis Walking Speed in Patients With Multiple Sclerosis Walking Speed in Patients With Multiple Sclerosis. JAMA Neurology 2014;71(11):1386-1393). The clinical meaning of walking speed as measured by the timed 25-foot walk in patients with multiple sclerosis. JAMA neurology, 71(11), pp.1386-1393). An associated label claim could be "treatment X suppressed/delayed disability progression as measured by T25FW"* *While SDMT did not demonstrate a clear link with function, the poor correlation between self-reported cognitive impairment and objective tests is a common finding across most neuropsychiatric conditions. These confounders make patient-reported measures of cognitive function a poor choice for establishing construct validity. Further assessment of clinical validity could be assessed against objective evidence on real world outcomes related to cognitive disability. For example, impact on employment or assessing the relationship between SDMT and caregiver-reported ADLs may also support the clinical*	

		meaningfulness of changes or specific milestones in cognitive disability. *Although the LCLA was not correlated with patient reported visual functioning in the voice of the patient study. it was correlated with patient-ratings on a well-established and validated tool, the NEI-VFQ in the literature review. Furthermore, the MSOAC submitted evidence of meaningful change thresholds, established using gold standard methodology, and demonstrated that changes of this magnitude were associated with meaningful deterioration in the NEI-VFQ. This provides a good basis for use of LCLA in the future, in particular as a potential functional biomarker in remyelination therapy trials (Triplett, J. D., et al. Pathophysiological basis of low contrast visual acuity loss in multiple sclerosis." Annals of clinical and translational neurology 2018;5(12):1505-1512).*	
115-124	5	Comment: We agree that capturing fatigue, pain and other concepts best known to the patient are essential as part of a comprehensive measurement strategy in MS. However, many of these symptoms are highly variable and are strongly influenced by confounding factors, including mood. Such symptoms would be inconsistent with the current approach to confirmed disability progression as a relatively permanent and persistent clinical symptom. We propose that such symptoms should be captured as secondary endpoints in clinical trials, consistent with the current approach. Proposed change (if any): As noted the T25FW, 9HPT, LCLA, SDMT tests do not incorporate fatigue, pain, sexual dysfunction, sensory outcomes. The result of the second project in the Voice of the Patient study confirms that fatigue (90.3%), incoordination (88.7%) and spasticity (75.6%) are severe problems in multiple sclerosis impacting overlapping levels of ADL (see table 14 page 81/205 of the briefing document). These impairments are also considered important by the consortium but thought to be better covered by PRO measures. ~~However, this~~ *This* raised the question if the concept of interest i.e. "disability in multiple sclerosis" is fully covered by the 4 dimensions selected *and*~~. This begs the question~~ whether a general questionnaire ~~e.g. Multiple Sclerosis Impact Scale - 29 items (MSIS) incorporating all these dimensions is not an alternative way forward although it is~~	Overall, the message is not different. No change required.

		~~acknowledged that PRO may be less objective and more subject to variability.~~ *incorporating all these dimensions could not be used alongside to aid a more accurate and comprehensive interpretation of the tests.*	
125-128	5	Comment: In the future we encourage an adaptive qualification approach, driven by the data. While it is acknowledged that a global disability scale would be of interest in MS, there are differences in the importance of the different domains that may lead to different definitions of a global disability score in MS patients. To address this challenge, establishing the validity of single domains is a preferred approach in the future. In the meantime, we believe the MSOAC has demonstrated the importance of the hand motor function and lower limb single domains thus far. While EDSS is covering a large spectrum of symptoms in MS, it is a global assessment rating scale and can be considered as well as a composite endpoint (*EMA/CHMP/44762/2017*). The final score calculation is based equally on each FSS scoring and this represents a serious limitation, for the following reasons. First, a subtle change in clinical expression of the disease might not be captured or sufficient to change the FSS. Second, when two FSS move in an opposite fashion, EDSS might not capture this change. Third, because of the calculation system, a change in one FSS might not be reflected in the final score. Fourth, every single FSS harbours the same ponderation, which, based on clinical neurology and VoP, constitutes a strong bias to address and monitor disability. The impact on QoL and ADLs might not be the same after a change in one point in a specific FSS (for example PYRAMIDAL FSS) or another (BOWEL FSS). It would be of value for the future to explain the reasons why the composite score has been abandoned, the challenges in building a composite score in MS and what would be required in the future if outcome measures characterizing single domain disability are established. We need to understand how EMA expects for a/a couple of composite score(s) to be built taking into account the clinical meaningfulness of such composite across broad EDSS ranges.	No clear rationale was provided by the Applicant for the change of concept and the present document provides an assessment of what has been provided. No change required.

		Changes proposed: Please detail in the text the rationale to not pursue a global disability score.	
129-141	5	Comment: The EMA seems to imply that because EDSS is a global assessment, it cannot be replaced with individual measures. The Agency should be clear whether an endpoint replacing EDSS has to cover multiple domains (eg, be a composite). In addition, the EMA notes that the proposed measures do not reflect concepts like fatigue and bladder dysfunction. Although it is well acknowledged that EDSS is a comprehensive assessment, it is also acknowledged that EDSS scoring is highly subjective (high inter and intra rater variability) and is driven by ambulation. So, from a methodological point of view, events in studies are not being driven by fatigue or bladder dysfunction but rather changes in ambulation. The EMA should acknowledge the value of objective changes in function to measure disease status, and as such acknowledge that there is value for patients in having such objective measures to make studies more efficient (shorter, smaller, especially for patients on comparator arms which could include placebo) as long as these domains are measuring concepts of value for patients. We acknowledge the fact that EDSS, by using 8 Functional System Scores (FSS), covers a substantial amount of domains affected in MS. However, as stated by the VOP, the important concepts such as fatigue, spasticity or incoordination are not correctly (in a deeper and meaningful manner) assessed within the EDSS. Especially for fatigue, numerous scales have been developed to fill this gap.	Nothing to add. The proposed approach also does not address this gap as stated in the Summary overall discussion ll320 ff. No change required.
151-152	5	Comment: We believe the VOP study was a high quality study that was well conducted to establish how more objective assessments of upper limb, lower limb, cognition and vision are associated with ADLs. However, we believe that this is more a measure of construct validity and initial qualitative work is needed to establish content validity. This would allow the development of a conceptual framework and help inform variables selected to assess each concept are important to patients. Such an approach is consistent with PRO	Acknowledged, however no change required since no longitudinal data were provided (see line 159).

		FDA guidance and the more recent PFDD guidance 3. Furthermore, we believe correlations may have been stronger if an average of test performance was taken over a period of time (eg, 1 week) and then correlating with ADLs. We acknowledge this could add additional burden to patients if multiple site visits were required, at the same time, this may have provided a stronger association and more accurate representation between test performance and ADLs.	
354-356	5	Comment: While we agree there are some potential confounders with SDMT, this is not a problem in the vast majority of patients. For this reason, SDMT is the strongest predictor of major socioeconomic outcomes, such as employment, independent living, and ADLs, including ability to drive (*Schultheis, M. T., et al. Examining the Relationship Between Cognition and Driving Performance in Multiple Sclerosis. Archives of Physical Medicine and Rehabilitation. 2010;91(3): 465-473.*). The confounders related to SDMT pale in comparison to EDSS confounds, such as rater-dependence, the subjective nature of the ratings leading to significant imprecision even when the rate is held constant, and difficulty distinguishing cerebellar from pyramidal dysfunction, among others. Proposed change (if any): ~~Moreover, there are learning effects and the SDMT performance can be influenced e.g. by visual acuity and ocular motor functions (Benedict 2017). Apart from that the type and degree of cognitive impairment in MS is highly dependent on the location of the lesions.~~	Not agreed since the weaknesses of the SDMT should also be addressed. See also above and reference to Giedraitine et al. Cognition during and after Multiple Sclerosis Relapse as assessed with the Brief International Cognitive Assessment of Multiple Sclerosis Sci Rep. 2018; 8: 8169.
91	6	Comment: Typographical change Proposed change (if any): The attractiveness of the performance tests chosen i.e. T25FW, HPT, LCLA and SDMT lies in ~~there~~ their	Already spotted and changed.
99-100	6	Comment: The draft qualification opinion includes conflicting statements on the assessment of speed of processing as a cognition parameter. For example in line 99-100,	We do not see this as conflicting statements. The focus on speed as a

343-345		the following sentence states 'Focus on speed of processing as cognition parameter needs to be more extensively justified' whereas line 343-345 states that 'Speed of information processing is important for cognitive function but whether it covers cognitive function in multiple sclerosis is not made clear'. Proposed change (if any): Wording to be revised upon clarifying the points above.	sole parameter of cognitive function is questioned. See also Giedraitine et al. Cognition during and after Multiple Sclerosis Relapse as assessed with the Brief International Cognitive Assessment of Multiple Sclerosis. Sci Rep. 2018; 8: 8169. See also specific comment above. No change required.
147	6	Comment: It is not clear what the following sentence means 'The value of the literature review is limited as the data dominantly concern cross-sectional data'. Proposed change (if any): Please clarify and revise if appropriate.	No change required since no longitudinal data were provided (see line 159) and specific comment above.
159-160	6	Comment: Typographical change Proposed change (if any): 'this is the major study where the ~~hypnotized~~ **hypothesised** linkage can be substantiated'.	Already spotted and changed
203	6	Comment: Based on the definition of disability in MS, any assessment that measures the neurological or neuropsychological impairment that limits patient's important activities of daily living should be considered as measuring disability, no matter whether it is used in studies for disease modifying therapies or symptomatic treatments. In other words, the efficacy of symptomatic treatment can be evaluated with a disability measurement. Propose to delete the sentence in line 203 (see below). Proposed change (if any): ~~However, the context of use of the T25FW was symptomatic treatment not for assessing disability.~~	Not accepted since there are different mechanisms of action of disease modifying and symptomatic treatments leading which eventually might lead to different outcomes at least with respect to duration of effects.

219-223	6	Comment: Reference is made to Bosma 2012, which looked at the relationships between 1-2 years changes on T25FW and EDSS and the long-term outcome (≥5 years) in patient reported outcomes (PRO) of progressive MS patients. Whilst the study demonstrated that changes in T25FW and EDSS were predictors of longer-term PRO disease impact, it showed that early change in T25FW rather than EDSS was significantly associated with the longer-term impact of MS. In our view, the conclusions of this paper are in line with the data submitted by MSOAC and does not appear to undermine the reliability of the aggregated clinical trial data analysis, as stated in line 222-223. Proposed change (if any): Delete lines 219-223 ~~This is unexpected considering that in the paper Bosma et al. (2012) it was shown that early changes in EDSS and T25FW are independently good predictors of long term EDSS (3 years). This is what would be expected as the two scales focusing on ambulation. It set some doubt on the reliability of the aggregated clinical trial data analyses~~.	In the paper by Bosma: Walking speed, rather than Expanded Disability Status Scale, relates to long-term patient-reported impact in progressive MS (2012) it is stated: Also, early change on the EDSS was associated with long-term reported walking limitations, although in a less pronounced way than the long-term effects seen following early changes in T25FW assessments. No change required.
240-241	6	Comment: The statement in the draft Qualification Opinion that a '15-20% difference in 9HPT is clinically relevant has not been convincingly demonstrated as being clinically relevant because the information in the literature review is anecdotal' is not supported by Biogen. Based on our own internal data analysis of MS studies (IMPACT), Biogen strongly supports a 15-20% difference in 9HPT as being clinically relevant. Proposed change (if any): Further clarify or delete this sentence	This refers to what has been provide by the Applicant in the literature review. No change accepted.

258	6	Comment: Typographical change Proposed change (if any): ~~T25FW~~ 9HPT - Summary	Already spotted and changed
282-283	6	Comment: The following sentence '*Correlation between LCLA and the physical component of the SF-36 is more than weak (table 41 page 41/205 of the briefing document).''* is not clear and should be revised. Proposed change (if any): If there is only a weak correlation then that should be clearly stated and contextualised as to why this may be the case. The reference to the briefing document should be removed unless this is also being published with the final opinion.	Accepted: Text should be revised: In the analysis of aggregated clinical trial data there was limited concordance in agreement between Disability Worsening at Endpoint as defined by EDSS and worsening as defined by LCLA (Kappa coefficient around 0.10). Deletion of ~~table 31 page 120/205 of the briefing document). Correlation between LCLA and the physical component of the SF-36 is more than weak (table 41 page 41/205 of the briefing document).~~