Assessing treatment outcomes in multiple sclerosis trials and the clinical setting

Working title: Outcomes in multiple sclerosis

Carmen Tur[1], Marcello Moccia[1,2], Frederik Barkhof[1,3,4,5], Jeremy Chataway[1,5], Jaume Sastre-Garriga[6], Alan J Thompson[5,7], Olga Ciccarelli[1,5]

1 Queen Square Multiple Sclerosis Centre, University College of London, UCL Institute of Neurology, London WC1B5EH, UK

2 Multiple Sclerosis Clinical Care and Research Centre, Department of Neuroscience, Federico II University, Naples, Italy.

3 Institute of Healthcare Engineering, UCL, London

4 VU University Medical Centre - Radiology & Nuclear Medicine, 1007 MB Amsterdam, the Netherlands

5 National Institute for Health Research University College London Hospitals Biomedical Research Centre

6 Multiple Sclerosis Centre of Catalonia (Cemcat), Neurology-Neuroimmunology Department, Vall d'Hebron University Hospital, Barcelona, Spain

7 UCL Faculty of Brain Sciences, Institute of Neurology, Department of Brain Repair and Rehabilitation, Queen Square, London, UK

Acknowledgements

CT acknowledges the ECTRIMS post-doctoral research fellowship for the financial support. JC, FB, AJT and OC acknowledge funding from the UK National Institute for Health Research (NIHR) University College London Hospitals/University College London Biomedical Research Centre.

Conflicts of interest

CT has received an ECTRIMS post-doctoral research fellowship in 2015. She has also received honoraria and support for travelling from Bayer-Schering, Teva, Merck-Serono and Serono Foundation, Biogen, Sanofi-Aventis, Novartis, and Ismar Healthcare.

MM has received salary from the Federico II University of Naples, Italy. He has also received honoraria and support for travelling from Merck-Serono, Genzyme, Coloplast and Almirall.

FB has received research support from the IMI (H2020) program AMYPAD, the Dutch MS Society and the NIHR UCLH Biomedical Research Centre. He has received consulting fees for work done in steering committees, data-safety monitoring boards and advisory boards from Bayer Schering Pharma, Sanofi-Aventis, Biogen-IDEC, Teva, Merck-Serono, Novartis, Roche, Synthon BV, Jansen Research, and Genzyme.

JC acknowledges the UK National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centres Funding scheme and

University College London (UCL). He has received support from the Efficacy and Mechanism Evaluation and Health Technology Assessment Programmes (NIHR); and the UK and National Multiple Sclerosis Societies. In the last three years he has been a local principal investigator for trials in multiple sclerosis funded by: Receptos, Novartis and Biogen Idec, and has received an investigator grant from Novartis outside this work. He has taken part in advisory Boards/consultancy for Roche, Merck, Biogen, MedDay and Apitope.

~~JC has received support from the Efficacy and Mechanism Evaluation Programme (NIHR) and UK Multiple Sclerosis Society, is the local principal investigator for trials in multiple sclerosis funded by Novartis, Biogen, and GSK, and has received an investigator grant from Novartis outside this work. He has taken part in Advisory Boards for Roche and Merck.~~

JSG has received honoraria as speaker and for participation in Advisory Boards in the last 36 months from Novartis, Biogen, Merck, Almirall, TEVA, Celgene and Genzyme.

AJT receives grant support from the Multiple Sclerosis Society of Great Britain and Northern Ireland, and has received honoraria/support for travel for consultancy from Eisai, Biogen (Optum Insight), and Excemed. He received support for travel from the International Progressive MS Alliance, as chair of their Scientific Steering Committee and the National MS Society (USA) as member of their Research Programs Advisory Committee, and receives an honorarium from SAGE Publishers as Editor-in-Chief of Multiple Sclerosis Journal.

OC receives research grant support from the Multiple Sclerosis Society of Great Britain and Northern Ireland, the NIHR UCLH Biomedical Research Centre, and she is a consultant for Teva, Roche, Novartis, Biogen, Genzyme and GE. She is an Associate Editor for Neurology, for which she receives an honorarium.

Word count/tables/figures/boxes/supplementary tables/references:

Abstract: 222 words; Main text: 8,443 words; # tables: 43; # figures: 2; # supplementary tables: 8; # boxes: 21; # references: 1966.

Abstract

Increasing numbers of drugs are being developed for the treatment of multiple sclerosis. Measuring outcomes is key to assessing the efficacy of drugs in clinical trials and monitoring response to disease-modifying drugs in individual patients treated upon registration. In both clinical trials and the clinical setting, most outcomes reflect relevant aspects of the disease, from clinical or neuroimaging perspectives, such as the presence of clinical relapses and accrual of disability, or the presence of visible inflammation and brain tissue loss, respectively. However, most of the measures employed in clinical trials to assess treatment effects on these relevant outcomes (i.e. outcome measures) are not used in routine practice. In the trial setting, the choice of outcome measures is crucial because they determine whether a drug is considered effective and can move to the next step of development; in the clinic, such outcome measures may be used for individual decision-making, such as choosing a first-line disease-modifying drug or escalating to a second-line treatment. This review discusses the clinical, neuroimaging, and combined outcome measures, including patient-reported ones, that are used in both trials and the clinical setting, to help clinicians and researchers to navigate through the multiple options when choosing an outcome measure. The barriers and limitations that need to be overcome to translate outcome measures from trials to a clinical setting are also discussed.

Introduction

Multiple sclerosis (MS) is a major cause of irreversible disability in young adults. Neurological disability in MS may occur as a consequence of acute relapses with incomplete recovery, or as a result of a clinical progression that occurs independently of the presence of relapses[1]. The pathological processes that lead to the development of acute disability are different from those that contribute to clinical progression. Acute inflammatory demyelination is responsible for the development of relapses, whilst neurodegeneration is the main determinant of progressive disability[2]. There are ~~no~~ few licensed treatments to slow progressive~~ion in~~ MS, whilst numerous disease-modifying treatments (DMT), which reduce the frequency of relapses in relapsing-remitting (RR) MS, are available. Current efforts are shifting towards progressive MS[3], and the number of trials has increased steadily over the last five years.

Measuring appropriate outcomes is central to assessing the efficacy of novel drugs, determining whether a drug can be moved to the next step of a drug development programme, and its regulatory approval. The efficacy of an experimental therapy cannot be demonstrated if the selected measure is unable to capture it, and no trial designs can compensate for inappropriate and poor measures. Outcome measures in RRMS trials focus on clinical (relapse) and radiological (lesion count) disease markers of inflammation, whilst in progressive MS the emphasis is on measures of clinical progression and (brain) atrophy as markers of neurodegeneration. Ideally DMTs would prevent both inflammation and neurodegeneration[4].

In the clinical setting, similar measures are used to monitor the response to DMTs in the individual patient, and, consequently, for decision-making, such as choosing a

specific initial DMT or escalating to second-line treatment. Although ~~M~~most of the outcome measures used in clinical trials are ~~not~~ used in routine practice, the level of standardization and the quality control are lower in the clinic than in trials, because of technical, financial and logistic barriers. However, important efforts have been made to standardise outcome measures in the clinic, especially in relation to monitoring treatment efficacy, in order to allow comparisons across centres[5,6].

The answer to the question what makes an outcome measure appropriate is a complex one. The psychometric properties of the measure must be appropriate for the study, and the chosen measure should be reliable and valid. Reliability indicates that the data collected are accurate and reproducible, while validity refers to the ability of the tool to measure what it is supposed to measure. In addition, the outcome measure must be responsive, i.e., detect changes in the specific functions and areas that are expected to occur as a consequence of the intervention/therapy[7]. The degree of the predicted changes in the outcome measure and the period over which they are expected to happen are also factors that need to be considered[8]. Well-known, traditional endpoints used in MS trials have the advantage that are immediately understood by clinicians, whereas novel outcomes may provide insights into more subtle, but relevant, treatment effects that would have been overlooked when using traditional endpoints. In the clinical setting, the choice of a response measure needs to consider whether the administration of the tool is easy, the data collected are clinically useful, and the interpretation of the test results is straightforward.

This review discusses the clinical and imaging outcomes used in clinical trials, stressing their advantages and limitations, which need to be considered when interpreting the results of clinical trials or designing new studies, with a particular focus on combined outcomes, as recently employed in progressive MS trials. The response measures used in routine clinical practice are also reviewed, and attention is given to their value and practicality. Clinically meaningful outcomes from the perspectives of patients and healthcare professionals are also discussed, with a view on their complementary role to more classical (objective) outcomes to detect treatment effects.

Outcomes in clinical trials

In this section, we first describe the clinical, neuroimaging and the other outcome measures that have been used in clinical trials, especially in phase III trials, and then the combined clinical and MRI measures.

Clinical outcomes

We have divided the clinical outcomes used in clinical trials into: clinical relapses, measures of disability progression, and patient reported outcome measures (PROMs). Relapse-based outcomes are prevailing in trials with RRMS patients, whereas progression-related outcomes are prominent in progressive MS trials. PROMs can be observed in all types of trials. ~~, but may be particularly relevant in trials with progressive MS patients, who are more likely to present with symptoms such as fatigue, pain or depression, than RRMS[9,10].~~ Regulatory agencies have therefore shown

a growing interest in the use of PROMs for trials in MS over recent years[9,10], to measure common and disabling symptoms such as pain, fatigue and depression.

Clinical relapses

The majority of phase III trials have been carried out in patients with RRMS, and, to a lesser extent, with the clinically isolated syndrome (CIS) (**Figure 1**). Since these trials aim to reduce (or suppress) the inflammatory activity responsible of acute relapses, the~~ir~~ main outcome measure is relapse count~~ing~~ (**Tables 1** and **Supplementary Tables 1** and **2**).

These relapse-centred outcome measures can be classified into four groups (**Supplementary Tables 1** and **2**): (i) quantification of the number of relapses in a discrete fashion (which are the most widely used) (ii) those that quantify the number of relapses as a binary phenomenon, such as the proportion of patients without relapses (relapse-free population) –or its opposite - the proportion of patients with at least one relapse (non-relapse-free population)–, (iii) metrics that quantify the time to the first relapse while on treatment (which are common in trials in CIS patients), and (iv) composite outcome measures.

~~An additional group that could be considered is based on the severity of the relapses, such as those associated with hospital admissions and intravenous steroids.~~

A relapse is generally defined as new or recurrent neurological abnormalities that are separated by at least 30 days from the onset of the preceding event. It lasts at least

24h, and occurs without fever or infection[11]. The definition of a relapse has changed over time and has become more stringent in recent trials compared with early trials[12,13]. For example, in the phase III ALLEGRO trial, which compared laquinimod with placebo in RRMS, neurological symptoms had to last at least 48h to be considered relapses[13]. The vast majority of ~~Some~~ trials demand an objective assessment by the examining neurologist[~~16~~], and request a specific increase in the Expanded Disability Status Scale (EDSS) score and associated Functional System sub-scores[14-16].

The most widely used outcome measure is the annualised relapse rate (ARR: number of relapses during the treatment period per patient-year), which belongs to the ~~first abovementioned~~ group (i) above and has been used so far in more than 40 phase III trials, most of which are in RRMS (**Table 1** and **Supplementary Table 1).** In more than half of these trials, and in all trials with RRMS, the ARR has been used as the primary trial endpoint (**Table 1**). The ARR is easy to understand and compute, and it is thought to reflect well the extent of inflammatory activity of the disease. However, it may lack specificity in respect to MS course severity, since the background level of disability and the severity of the attack are not captured. This limitation has prompted the development of the annualised rate of severe relapses, which are those relapses that require intravenous steroid treatment and/or hospitalisation[13], or those that entail a high-level of disability[17], which has been used since 1993 as a secondary endpoint (**Table 1**). However, the lack of standard guidelines to treat MS relapses implies there is an enormous inter-site variability in terms of management of relapses and it might not be appropriate to consider these measures as potentially eligible clinical outcomes in trials.

The second group of relapse-centred outcome measures includes the relapse-derived binary outcome measures, which have been used since the very beginning of the trials in MS, but have become more popular over recent years with the testing of highly effective drugs that may lead to a relapse-free status. The percentage of relapse-free patients and the percentage of patients with at least one relapse may depend on the length of the study, as the risk of getting a relapse may increase with time; therefore, the design of the study needs to be considered when comparing these outcome measures among trials. For example, the GATE study, a 9-month placebo-controlled phase III trial, where generic glatiramer acetate (GA) was compared to brand GA and placebo in RRMS patients, the percentage of relapse-free patients in the placebo group was 79~~.3~~%. Instead, in RRMS trials with longer durations, usually 24 months, such as the FREEDOMS[14] or the ALLEGRO[13] studies, that percentage is around 50-60%. This has immediate consequences from a statistical point of view: to be able to detect a given difference in relapse-free patients between placebo and active arms, we will need much greater sample sizes if the percentages in both groups are around 50% than if they are closer to 0% or 100%.

The most relevant measure within the third group is time-to-relapse, often used in CIS studies, where the occurrence of the first relapse since study entry indicates conversion to clinically definite MS (CDMS)[18-20]; therefore, time to CDMS is often the primary trial endpoint (**Table 1** and **Supplementary Table 1**). Since the development of a new lesion on MRI in patients with CIS can also confirm a diagnosis of MS (assuming that the dissemination in space criteria are also fulfilled), according to the

2001 McDonald criteria[20], time to McDonald MS has also been used as trial endpoint in CIS trials, although this measure requires a trial design with repeated MRI scans and is heavily dependent on frequency of MRI assessments. At present, a few phase III trials have used time to McDonald MS as trial endpoint: the BENEFIT study[21], which compared interferon beta-1b 250µg SC every other day versus placebo, the REFLEX study[22], comparing three-weekly and weekly INTERFERON beta-1a versus placebo, and the TOPIC study[23], comparing oral teriflunomide 7mg and 14mg versus placebo (**Table 1**). In the both BENEFIT and REFLEX studies, where time to McDonald MS was the primary outcome, this reached statistical significance well before time to first relapse and allowed for a dose differentiation in REFLEX that was not apparent using clinical outcomes[22].

The most important outcome within the fourth group is "time to treatment failure", which is a primary composite endpoint, recently introduced in the TENERE study, which compared oral teriflunomide 7mg and 14mg versus interferon beta-1a in RRMS[15]. The time to treatment failure is defined as the occurrence of the first confirmed relapse while on treatment, or permanent treatment discontinuation for any cause[15] (**Table 1**); this outcome is thought to account for all the factors that determine the effectiveness of a therapy, such as efficacy, safety and tolerability, and, therefore, may be applicable to the real-life clinical setting.

Measures of disability progression

Measures of disability progression are generally used as primary outcome measures in phase III trials in progressive MS (**Table 1** and **Supplementary Table 3**). Most p~~P~~hase

III trials in progressive MS using these outcome measures have reported negative results[24,25], with the exception of the ORATORIO study, which compared IV ocrelizumab versus placebo in primary progressive (PP) MS[26] and the EXPAND trial, which compared oral siponimod to placebo in secondary progressive (SP) MS[27]. Many trials in RRMS (and CIS) patients have also included disability progression as a trial endpoint (**Table 1** and **Supplementary Tables 1** and **2**), either secondary or primary, suggesting that targeting clinical progression is also a ~~may be a~~ priority ~~even~~ in the relapsing forms of MS.

~~Similarly to the relapse-centred outcomes, d~~Disability progression-related outcomes can be classified into ~~four~~ five groups: (i) those that quantify the amount of progression in a continuous fashion, such as changes in the Expanded Disability Status Scale (EDSS)[28] scores, or the EDSS score at follow-up, (ii) metrics that quantify the amount of progression as a binary phenomenon, such as the proportion of patients with (or without) (confirmed) disability progression, (iii) quantification of the (confirmed) *improvement* in disability progression also binary, (iv~~ii~~) ~~metrics~~ those that quantify the time to confirmed disability progression (CDP), and (~~i~~v) composite outcome measures (see **Table 1**).

The most frequently used outcome measure in the first group is the absolute change in the EDSS score from baseline to follow-up (**Table 1** and **Supplementary Table 3**). Of note, in some trials, such as the PRISMS[17] and CARE-MS I[29] and II[30] trials, changes in the EDSS raw scores are reported, but in other trials, such as the Copolymer-1 trial in RRMS[31], the EDSS-step methodology, instead of raw EDSS changes, is used. It consists

of assigning new values to observed EDSS changes depending on the position of the initial EDSS score in the whole scale. This approach was meant to overcome the non-linear behaviour of the EDSS. The main limitations of the EDSS-based measures are that a worsening in EDSS does not reflect which functional system changes and that a relapse-associated transient deficit may lead to a (transient) change in the EDSS[32]. Additionally, the EDSS may not be sensitive to deterioration of the upper limb motor function, cognitive function or short-distance walking, which may occur in patients with progressive MS and high EDSS scores[33]. Besides, the absolute change in EDSS, especially when relying on a small number of visits, may be affected by noise due to the low inter-rater and intra-rater reproducibility of the scale, namely in the lower end of the scale[34]. The EDSS score does not reflect the whole patient's functional impairment, since it has a low ability to discriminate people with different levels of disability according to the Barthel Index[35], a measure of functional independence in 10 daily activities[36]. Therefore changes in scores other than EDSS, such as MS Functional Composite (MSFC)[37], and its subtests[38,39], Regional Functional System Score (RFSS), ambulation index, arm index, and cognitive tests[40,41], from baseline to follow-up, have been included into some trials to complement the EDSS (**Table 1** and **Supplementary Table 3**). Cognitive tests that have been used in phase III trials include the: Paced Auditory Serial Addition Test (PASAT), which is one the subtests of the MSFC[37]; Rao's Brief Repeatable Battery (Rao's BRB)[42]. With the PASAT, the changes in the z-score over the trial period time was used[43,44]. For Rao's BRB, different trials have used different outcome measures: whereas in the phase III North American trial of SC interferon beta-1b in SPMS the outcome measure was the change in a composite neuropsychological score[41], in the ARIANNA study (atorvastatin add-on vs. placebo

add-on in RRMS patients on SC interferon beta-1b treatment), the outcome measure was the change in the percentage of patients with mild or severe cognitive impairment, defined as failure in one-two or three or more tests, respectively[40].

~~With regard to the o~~Outcomes in the second ~~and third~~ and fourth groups ~~, they~~ vary considerably between studies and are numerous (**Table 1** and **Supplementary Table 3**). Confirmed disability progression (CDP) is defined as a worsening of the EDSS (usually 1.5-step EDSS progression when starting EDSS is 0, 1-step EDSS progression for EDSS≤5.5, or 0.5-step EDSS progression for EDSS>5.5) that persists for either three or six months. It has been demonstrated that 3-month and 6-month CDP overestimate the long-term accumulation of irreversible disability by 30% and 26%, respectively[45]. Longer disability confirmation periods (12 and 24 months), although not completely free from such bias (overestimation of 20% and 11% respectively), would be recommended to detect true, irreversible disability, with a possible little effect on the sensitivity of the progression criteria[45]. However, so far, no trials have used such long periods to confirm disability progression. Most trials have used both 3-month and 6-month disability progression, although some recent studies, such as CARE-MS I[29] and II[30], have used only the 6-month CDP outcome. If a trial uses the time to 3-month CDP (or the percentage of patients with 3-month CDP) as primary endpoint, then the time to 6-month CDP is a secondary endpoint.

The MSFC or its subtests, which are the 25-foot Timed Walk Test (TWT), the 9-Hole Peg Test (9-HPT) (which reflects the motor impairment in the upper limbs), and the Paced Auditory Serial Addition Test (PASAT) (which reflects the speed of (auditory)

information processing and calculation ability)[39], can be used instead of the EDSS to define the CDP. Although the training effects often seen on the PASAT could theoretically be responsible for a lower responsiveness of the MSFC than the EDSS to detect disability progression[46], this is not supported by the results of the trials published so far, where MSFC-derived outcomes seem to be more sensitive than those derived from EDSS. For example, the CARE-MS II[30] or the FREEDOMS II[47] trials, carried out in RRMS, or the IMPACT trial, in SPMS[43], showed significant results in the MSFC but not in the EDSS. Instead, trials that showed significant effects in the EDSS, such as CARE-MS I[29] and the FREEDOMS[14], tended to show also significant results in the MSFC.

Further attempts have been made to improve the sensitivity of MSFC and its subtests to disease progression, and therefore increase its sensitivity to treatment effects. For instance, it was suggested that only increases of at least 20% in MSFC subtests were clinically meaningful and had an acceptable signal-to-noise ratio, suggesting that clinical trials should use outcomes based on these subtests as binary metrics[48]. However, so far, only one phase III trial, the ARIANNA study, which compared oral atorvastatin add-on to SC interferon beta-1b in RRMS, has used this 20% cut-off to define the MSFC-related outcome measure[40].

Among the outcome measures of the third group, the most widely used one is the sustained improvement in the EDSS score, which was used as a secondary outcome in the CARE-MS II trial[30] and The Copolymer 1 Multiple Sclerosis study[31] (**Table 1**). In phase III trials, it has only been used when drugs were to be tested in patients with RRMS, possibly reflecting the role of acute inflammation in the development of

disability in these patients. Quite recently, a phase II study carried out in progressive MS, the *biotin study*, also used the improvement of disability as an outcome measure –in particular, as a primary outcome measure[49]. In this study, which showed positive results, the improvement of disability was not only reflected by improvements in the EDSS score, but also in the TWT score[49]. Improvement was considered if there was a decrease in the EDSS of ≥0.5 or ≥1 points, if baseline score was between 6 and 7 or between 4.5 and 5.5, respectively, or if there was a decrease in the TWT of at least 20%. Sustained improvement of disability as outcome measure may therefore reflect clinical changes secondary to not only remission of inflammation but also tissue regeneration, which may be expected in the new era of drugs being tested in progressive MS, such as the abovementioned biotin[49], simvastatin (tested in the phase II MS-STAT trial[50]) or oxcarbazepine (being currently tested in the phase II PROXIMUS trial[51]).

Composite endpoints, which are in the fourth group of disability progression measures, facilitate higher event rates and theoretically increase the sensitivity of the progression parameters, thereby reducing the length of the trial and the sample size. Besides, they theoretically reduce the risk of multiplicity and so the risk of type I error[9]. However, composite endpoints should be pre-specified before starting the trial and their individual components should only be tested when there is a statistically significant treatment effect for the composite, unless the components have been pre-specified as outcome measures too[9]. A recent reanalysis of a PPMS trial showed that composite endpoints including different disability measures allows detection of larger treatment effects, then reducing the sample size needed for clinical trials[52]. The

highest efficiency and event rate estimates were obtained by using a sustained disability progression endpoint confirmed by any two of the following: [EDSS and TWT] or [EDSS and 9-HPT] or [TWT and 9-HPT]. This endpoint usefully combines the logical "and" and "or" criteria, maximizing the likelihood to detect a clinical event. However, composite endpoints are only valid when the composite includes outcomes that are causally related to the treatment[53].

A recent phase III trial in PPMS used as primary outcome measure the time to 3-month CDP based on a composite endpoint, defined as the presence of at least one of the following three changes: increase in EDSS (1 if EDSS<5.5 or 0.5 if EDSS ≥5.5), increase in ≥20% in 9-HPT, and increase ≥20% in TWT[54]. Post-hoc re-analyses of trial data have suggested that this composite endpoint may separate MS patients with ongoing progression from those who are stable[54], thereby representing an improved endpoint for disability progression trials. Another composite outcome used as secondary endpoint in a progressive MS trial[55] is the time to a 3-month CDP or to a confirmed 20% worsening in the 9HPT treatment failure (**Supplementary Table 3**).

Patient-reported outcome measures

Patient-reported outcome measures (PROMs) are self-completed questionnaires that measure the impact of the disease on daily activities, social functioning and quality of life. In 2009, the Food and Drug Administration (FDA) published a guidance on PROMs[9], which were defined as 'any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else'[9]. In 2016, the European Medicines Agency

defined PROMs as any data directly reported by a patient that is based on his/her perception of a disease and its treatment (www.ema.europa.eu), thereby further developing the concept of "personal perspective". The term PROM is an umbrella term, which includes evaluations of health-related quality of life, health status, well-being, satisfaction with treatment, adherence to treatment, and symptoms. Therefore, PROMs complement and support the outcome measures based on clinical assessments, and, as mentioned in the 2009 guidance of the FDA, they can be used in clinical trials to measure the risks of a given treatment as well as its benefits[9,56].

PROMs can be divided into two groups: condition-specific and generic PROMs. In the first group, there are tools designed for MS, which cannot be extrapolated to the general population and are sensitive to detect an MS-induced change. Examples of MS-specific health-related quality of life PROMs are the 29-item MS impact scale (MSIS-29)[57], the patient-reported indices in MS (PRIMUS)[58], and the MS quality of life-54[59]. Thirteen fatigue-centred PROMs have been proposed in 20 years, and the most popular are the fatigue severity scale (FSS)[60] and the fatigue impact scale (FIS)[61] (**Table 2**). The MS-specific PROMs that measure the impact of motor impairment on daily activity, such as the Arm Index[37] and the Multiple Sclerosis Walking Scale (MSWS-12)[62,63], have been frequently used as trial endpoints over the last 5 years[54,64,65] (**Table 2**). It has been suggested that a reduction of 4-6 points on the MSWS-12 is clinically meaningful[66], although the MSWS-12 has also been used as a continuous measure, without any thresholding, in a symptomatic trial (i.e. the Fampridine trial)[67]. Many generic PROMs, such as those that focus on symptoms, such as pain, tremor, and

spasticity, have been used in symptomatic trials in MS[68,69], but a deep discussion of these is outside the scope of this review.

PROMs that in future may be further studied and validated for use in clinical trials and clinical practice are the patient-determined disease step (PDDS), which is a simple and economical scale compared with the EDSS, but correlates with it and its functional system scores[70], and the subscales of both the MFIS and MSIS-29. A recent trial has included the physical subscale of the MSIS-29 as a co-primary endpoint of the study together with time to EDSS-based 6-month CDP[65]. This indicates that composite endpoints may be obtained by combining objective scales (e.g., EDSS) and PROMs, although the same limitations associated with the combined scores discussed above apply to these combined endpoints.

Neuroimaging outcomes

We have divided the neuroimaging outcomes used in clinical trials into: focal brain lesions, brain and spinal cord atrophy measures, and novel MR outcomes for neurodegeneration and remyelination.

Focal brain lesions

MRI measures of focal brain lesions often serves as primary endpoints in phase II trials and typically secondary outcomes in phase III trials. They are particularly relevant to trials carried out in patients with RRMS and CIS, which test the efficacy of medications targeting the inflammatory activity[71] (**Table 3** and **Supplementary Tables 4** and **5**), although they are also used in trials in progressive MS (**Table 3** and **Supplementary**

Table 6). The most commonly used MRI measures are based on T1 gadolinium enhancing and new T2 brain lesions, which reflect the occurrence of new inflammatory activity. In particular, Gadolinium enhancement signifies breakdown of the blood-brain barrier as a consequence of acute inflammation in the CNS. However, there is a fundamental difference between T1 gadolinium enhancing and new T2 brain lesions, since T1 gadolinium enhancing lesions are transient (average duration of 3 weeks[72]) and a single scan will miss cumulative new inflammation over a period of time. Instead, given the (generally) non-transient nature of the T2 lesions, 'new T2 lesions' with respect to the last scan would capture cumulative new inflammation between the last and the current scans. Nonetheless, ~~In particular,~~ the 'number of gadolinium-enhancing lesions' during or at the end of follow-up is the most widely used trial outcome in all phase III trials (**Table 3**). ~~Gadolinium enhancement signifies breakdown of the blood-brain barrier as a consequence of acute inflammation in the CNS.~~ T1-hypointense lesions are visible in both the acute phase of a lesion development (corresponding to the lesional oedema) and the chronic phase[73,74]; in the latter case they are called permanent black holes (PBH), which have been mostly used as a post-hoc measure of tissue destruction and recovery[13].

Lesion-derived measures can be divided into three categories: (i) outcomes that measure the occurrence of new lesional activity during the trial, such as the number of new and/or enlarging T2 lesions or new T1 gadolinium enhancing lesions, (ii) outcomes that quantify the total lesion volume, either T2-hyperintense, T1-hypointense or gadolinium-enhancing lesion volume, and (iii) those that estimate the inflammatory activity as a binary phenomenon, such as the proportion of patients

without gadolinium enhancing lesions. Finally, there would be a set of metrics that could be included within the first group, since they reflect new, acute lesional activity, and that are derived from the combination of different MRI measures. An example of these composite MRI measure is the number of combined unique active (CUA) lesions, which describes the total number of active lesions in the widest sense and includes all new, enlarging T2 lesions or new enhancing lesions, provided that the same focal lesion is counted only once. This endpoint was originally proposed by Paty and Li and was already used in the first clinical trials in RRMS. In CIS trials, it was used for the first time in the early 2000 by the ETOMS study[75], and in SPMS trials, it was first used in the SPECTRIMS study[76,77]. So far, at least 13 phase III trials have used it (**Table 3**).

The greatest advantage of lesion-related markers is that they provide objective measures of the underlying pathology and correlate with clinical outcomes in RRMS, in particular with relapses, at least in the short/medium term[78]. It has been demonstrated that more than 80% of the between-trial variability in terms of treatment effects on relapses is explained by the between-trial variability in terms of treatment effects on new T2 lesions on MRI[79]. In addition, treatment effects on relapses of phase III trials can be predicted by the treatment effects on lesion-related outcome measures in the corresponding phase II trials that used the same drug[80]. Another advantage of lesion-related measures is that, given their high sensitivity, they allow the comparison of two active drugs, which can be difficult when the outcome is clinical relapses. For instance, in the GATE study, which compared generic glatiramer acetate with the originally branded drug, lesion-related outcomes were used to show equivalence of the two drugs[81].

The counting of new T2 lesions can be limited by factors such as high pre-existing lesion load, suboptimal repositioning of serial scans and poor inter-observer reproducibility. Image subtraction has been proposed to overcome these issues, thus providing good visualization and quantification of both active and shrunken or resolved T2 lesions[82]. The combination of automated identification of new/enlarging lesions with automated lesion subtraction may be useful to improve cost-effectiveness and reduce the risk of adverse events associated with gadolinium administration[83].

Brain atrophy measures

The rationale behind the use of brain atrophy in clinical trials is that it reflects neurodegeneration, which is the pathological process most consistently linked to accrual of disability[84-86]. Total brain volume/fraction is the non-lesional outcome measure most commonly used in phase III trials (**Table 3).** It is generally used as a secondary outcome measure in phase II and III trials, such as the FREEDOMS study[14], where fingolimod was compared to placebo, or the CARE-MS I[29] or II[30] studies, where alemtuzumab was compared to interferon beta-1a. Nonetheless, it has recently been used for the first time as primary endpoint in phase II[50] ~~and phase III~~ trials in secondary progressive MS (http://www.ms-smart.org, accessed on 29/06~~1~~/2017; and https://clinicaltrials.gov/ct2/show/NCT01910259?term=MS+smart&rank=1, accessed on 29/06~~1~~/2017), and also in the ongoing phase II ARPEGGIO trial in PPMS[87]. In RRMS, the treatment effect on brain atrophy correlates with the effect on disability progression over 2 years, independently of the effect on active MRI lesions[66].

There are two types of brain volume-derived metrics (**Table 3** and **Supplementary Tables 5** and **6**): (i) metrics that calculate global brain atrophy, as either brain parenchymal volume[88] or fraction[40] (which is the ratio of brain parenchymal volume to the total volume within the brain surface contour), and their change over time, and (ii) metrics that estimate regional volumes, such as white matter and grey matter, and change thereof during the trial[89].

The most widely used measures in the first group are the brain parenchymal fraction (BPF), a segmentation-based technique that reduces the variability caused by individual variation in brain size and has high test–retest reproducibility when compared with raw brain volume[90], and the percentage brain volume change (PBVC), a registration-based difference map of brain contours over time[91,92]. BPF has been used in studies such as the phase II trial with natalizumab in RRMS[93] or the phase II trial with interferon beta-1b in PPMS[94]. PBVC has been used in the phase III fingolimod trials, i.e. the TRANSFORMS[95] and FREEDOMS I[14] and II[47] studies, and the phase III laquinimod trials, i.e. the BRAVO[96] and the ALLEGRO[13] studies.

In addition to the well-known technical sources of measurement error, such as changes in magnet, gradients, coils, distortion corrections and image-contrast changes that affect tissue segmentation, global atrophy metrics are susceptible to: (i) the phenomenon of pseudo-atrophy, likely due to resolution of inflammation and oedema and especially seen in patients on active treatment with greater gadolinium-

enhancing lesion volume at baseline[97,98], (ii) physiological (circadian) variations in hydration status[99], and (iii) smoking and other cardiovascular risk factors[100].

The measures in the second group most commonly used are the grey and the white matter volumes. The change in the volume of CSF (normalised by the total intracranial volume) has also been used in phase III trials[101,102], as an attempt to quantify indirectly loss of neural tissue. A single phase II trial used the partial (central) cerebral volume, a surrogate estimate of global atrophy[89]. The same trial showed that a reduction in grey matter volume over time is greater than that in the white matter, and is less affected by pseudoatrophy[98], as other observational studies have also reported[103]. Grey matter and thalamic volumes have also been used as additional outcome measures in the phase III ALLEGRO study[13]. Therefore, if these partial volumes are confirmed to show a greater change over time than global measures[89,104,105], they will result in higher sensitivity and a smaller sample size.

Spinal cord atrophy

Spinal cord atrophy is usually measured at the cervical level, and has been associated with long-term development of motor disability, not only in progressive MS but also in relapse-onset MS[106,107]. The rate of brain atrophy in MS is about 0.5% a year[108], whilst that of spinal cord atrophy has been shown to be higher, up to 2.2% a year in SPMS[109], suggesting that spinal cord atrophy may be a sensitive and meaningful marker of neurodegeneration. Trials in PPMS or SPMS have used the change in cord area[54] as a secondary endpoint (**Supplementary Ttable 7**). However, there are methodological factors that affect the noise of this measurement in multi-centre

trials, mostly related to the limited spatial resolution of current MRI scanners relative to the small cord size and cord movement. This translates into larger sample sizes than those estimated from a single centre/scanner study[110]. Additionally, spinal cord atrophy-related measures are calculated using semi-automated segmentation-based methods, which are subject to inter-rater variability.

Novel imaging outcomes for neurodegeneration and remyelination

New outcomes have been proposed and used over the last 5 years to detect the effect of drugs at a microscopic level. The advantage of such measures is that they are expected to be more tissue-specific for the underlying pathophysiological processes than conventional MRI measures, and, therefore, may detect changes reflecting the underlying mechanisms of damage caused by the action of the experimental medication. These novel measures may provide complementary information to that given by conventional imaging endpoints and insights into the mechanistic efficacy of the medication.

The most widely used measure is the change in magnetic transfer ratio (MTR) in the whole brain[13,16,111] (**Table 3** and **Supplementary Table 4**). MTR changes are thought to reflect the process of demyelination[112] and remyelination[113]. Apart from whole brain MTR, regional MTR, such as grey matter, white matter and lesional MTR, have also been used (e.g., in the phase III, ALLEGRO trial in RRMS[13]).

Other measures –used mostly in the past– to show an effect of DMTs are metabolite concentrations, estimated by MR spectroscopy imaging, such as N-Acetyl

Aspartate[13,114]. Novel secondary outcome measures currently used in phase II trials in secondary progressive MS are diffusion ~~metrics~~ parameters derived from NODDI (Neurite orientation dispersion and density imaging), which estimate the microstructural complexity of dendrites and axons in vivo[115] and sodium imaging[116] (https://clinicaltrials.gov/ct2/show/NCT02104661?term=oxcarbazepine+multiple+sclerosis&rank=1, accessed 29/06/~~January~~ 2017).

Optical Coherence Tomography (OCT) measures axonal and neuronal loss within the anterior visual pathway, which not only correlate with the visual function[117,118], but also reflect whole-brain process of neurodegeneration, especially in progressive MS[119]. For that reason, it has been proposed as outcome measure in both optic neuritis[120] and non-optic neuritis MS trials, such as the PROXIMUS (add-on oral oxcarbazepine vs. placebo in progressive MS)[51], the FLUOX-PMS (oral fluoxetine vs. placebo in monotherapy in progressive MS)[121] and the ACTiMuS (bone marrow-derived cellular therapy in progressive MS)[122] trials. Please see **Box 1** and **Supplementary Table 8** for more details on OCT-related outcome measures.

Combined clinical and MRI outcomes

Although the use of these types of measures emerged in MS trials in 2012 with the CombiRx trial, the concept dates back to 2006, when Rio et al. showed that the absence of relapses, disability, and inflammatory activity visible in the MRI (at certain thresholds) after a given time on treatment would possibly indicate so minimal disease activity that the risk of progression over a longer follow-up was negligible[5]. In 2014, the outcome measure called “no evidence of disease activity” (NEDA)[4] was defined as

no relapses, no progression of disability, and no MRI activity (new/enlarging T2 lesions and T1 gadolinium enhancing lesions). It had been initially defined as "Disease Activity Freedom" (DAF) in the natalizumab AFFIRM trial[123] and later re-termed as NEDA. It has been recently used in phase III (**Table 3** and **Supplementary Table 1**)[29,30,101] and phase II trials[124,125]. NEDA has also been used to compare the efficacy of medications among trials; for example, AHSCT (autologous haemopoietic stem cell transplantation) trials have shown a greater proportion of patients reaching the NEDA status than other treatments[126]. Since brain volume loss reflects neurodegeneration (the main determinant of progressive disability), it has been proposed to include it in the definition of no evidence of disease activity (so-called "NEDA-4"), together with relapses, MRI disease activity and clinical progression[127].

Another combined endpoint is the event-free survival[128], used in AHSCT trials, which includes death as an outcome in addition to worsening of disability, relapse and new MRI lesions, suggesting that combined measures can be designed to reflect the expected efficacy and main adverse events of the drug.

The main objections to the use of these combined measures in clinical trials are that the net effect of the experimental drug on the composite ~~metric~~ may be difficult to interpret, if the effect on the different components is not the same, and there is uncertainty in respect to the clinical relevance for individual cases[53,129].

Outcomes in the clinical setting

In this section, we describe the clinical and neuroimaging measures that are currently used in clinical practice.

Clinical measures

In clinical practice, the most widely used clinical measures are related to the occurrence of relapses and clinical progression, generally measured with the EDSS.

Relapses

The number of relapses occurred within a given time frame, usually 6-12 months, is the clinical outcome most commonly used in clinical practice. It traditionally requireds taking a medical history (which ~~may~~ could be associated with a recall bias) and inspecting the clinical notes. The use of high-quality prospectively designed databases can allow a more precise retrieval of relapse-related data in the clinic, successfully enabling clinicians to assess treatment effects in clinical practice[130,131]. The presence of relapses while on treatment, in combination with other factors such as EDSS increase[5] or MRI activity[132], has been considered as a surrogate for future disability. Along these lines, a recently published study from the MAGNIMS group, which included 1,280 patients with RRMS on disease-modifying treatment, showed that the presence of at least 2 relapses (or 1 relapse and ≥3 new T2 lesions) during the first year of treatment with interferon beta was associated with 48% risk of treatment failure, defined as a confirmed EDSS worsening (≥1 point increase in EDSS if starting EDSS <5.5, or ≥0.5 increase if EDSS ≥5.5) or a switch to other therapies for lack of efficacy, and 29% risk of EDSS worsening over 3 years[133].

Measures of disability

The most common measure collected in clinical practice is the EDSS, which is used in the outpatient clinics to assess the severity of clinical relapses and monitor treatment effects. This scale is based on the standard neurological examination, which is part of any clinical assessment, and clinicians are very familiar with the meaning of scores above 4.0, which are based on walking ability. Therefore, the EDSS may be easy to interpret clinically. However, as mentioned above, it has low intra- and inter-rater reproducibility, especially for patients with mild to moderate disability. Besides, the EDSS is not sensitive to important aspects of clinical progression, such as cognitive dysfunction.

The MSFC is not used in the clinic as frequently as the EDSS or as often as in clinical trials. One of the MSFC subtests, the PASAT test[134,135], assesses the speed of (auditory) information processing and calculation ability, and may compensate for the fact that cognitive impairment is not captured by the EDSS. The TWT may be routinely performed in the clinical setting when assessing patients' ability to walk before and after fampridine, to know whether the patient has benefited from the drug[136]. However, the MSFC and its subtests have been designed to be used in clinical trials, for group analyses, rather than to be used in the clinic, at the individual level[39]. To use the MSFC or its components, it is required an a priori definition of a clinically meaningful change. Besides, the reference population affects the values of the MSFC z-scores, which means they cannot be easily interpreted in the clinic. Other limitations include the practice effects[137,138], which may influence the PASAT, and the fact that the PASAT can be too distressful[139].

Considering the prevalence of cognitive dysfunction in MS and its impact on patients' day-to-day lives, a committee of experts on cognitive dysfunction in MS agreed on the need of regular cognitive assessments in patients with MS and proposed a brief battery to be administered in the clinic, the Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS)[140]. This includes the Symbol Digit Modalities Test (SDMT)[141], which is also included in the Rao's battery[142], and is the most widely used cognitive test. It measures attention and speed of information processing and lower scores have been associated with the severity of white matter damage[143]. It has been shown to be more valid and reliable than the PASAT, in part because it is a less ~~di~~stressful test[144]. It requires a few minutes in total to be performed and the person who administers the test does not require a specific training[142]. For all these reasons, it is considered the best test to be administered if the time allocated to cognitive assessment is very limited[140]. In addition to the SDMT, the BICAMS includes The California Verbal Learning Test (Second Edition) and The Brief Visuospatial Memory Test (Revised Version), and tests of verbal and visuospatial memory[140]. Apart from the SDMT and the PASAT, the remainder of the tests included in the Rao's battery can also be used in the clinic, although training of the health professional is required[142]. Finally, the Cogstate battery, a computerized tool made of simple rapid tests measuring processing speed, attention, working memory, executive function and verbal learning has been used in several neurological conditions, including MS[145]. In general, cognitive tests in the clinic are difficult to administer due to time constraints. Thus, more novel batteries such as the Cogstate, which can be self-administered online, are potentially more promising in clinical practice. Additionally, it is neutral to language and culture, being therefore preferable to other tests that may be influenced

by education. Additional factors to consider are the effects of depression, anxiety and fatigue on performance~~Besides, age, education, depression and anxiety, and fatigue may affect performance on all cognitive tests~~.

The PROMs discussed above can also be used in the clinic. In particular, the fatigue scales, such as the FSS~~atigue Severity Scale~~[60], the Modified Fatigue Impact Scale (MFIS)[146] or the Visual Analogue Scale for fatigue[147], may be used. Other useful PROMs are those that relate to depression, anxiety, pain or quality of life. Interestingly, in the near future, the usefulness of PROMs in the clinic may substantially increase with the help of the new technologies, since PROM-related information can be collected and displayed to clinicians electronically.

Neuroimaging measures

In this section we review the T2 lesions, which is the most commonly used response measure in the clinical setting, followed by brain atrophy and combined outcome measures, which have recently started to emerge and are therefore also discussed.

Lesion-related measures

MRI has become a ~~very useful~~vital tool in clinical practice. According to international recommendations, patients should be scanned regularly, usually at least once a year[148,149], especially if they are on treatment, or even more frequently, if they are on certain treatments such as natalizumab, fingolimod or dimethyl fumarate, and considered to be at risk of John Cunningham virus (JCV)-positive progressive multifocal

leukoencephalopathy[150]. However, other time frames may still be possible and it is not fully clear which is the best to adopt for routine, non-urgent MRI scans[148,151]. International consensus recommends to perform a brain and/or a spinal cord MRI scan when unexpected or atypical symptoms appear[148,151]. Ideally, when brain MRI is used for monitoring of disease activity and treatment efficacy, it should be performed on the same MRI system, using the same imaging protocol (i.e., the same pulse sequences and spatial resolution) as the reference (baseline) scan[148].

The most common response measure is the number of new (or enlarging) T2 lesions, as compared with the previous scan, which is also referred as the number of *active T2 lesions*[148]. The number of active lesions is useful to monitor treatment response, since the presence of new T2 lesions while on treatment has been associated to a worse clinical outcome[6,148] and may indicate the need for a treatment change[6]. The occurrence of at least 3 new T2 lesions in the first year of interferon beta therapy was associated with 27% risk of treatment failure (defined as confirmed EDSS increase or switch to other therapies for lack of efficacy) and 22% risk of EDSS worsening over 3 years[133]. A disadvantage of the number of active T2 lesions as a response measure in the clinic is that it requires previous MRI scans of the patient to be available for comparison, and an experienced radiologist. Recently, the feasibility and reliability of automated lesion segmentation algorithms using clinically acquired scans has started to be assessed, showing promising results[152]. Therefore, in the near future, these algorithms may allow the automatic computation of total T2 lesion load in the clinic, potentially improving the monitoring of patients with MS.

Another MRI measure used in the clinic is the number of Gd-enhancing lesions, which provides information on acute inflammation and does not require the availability of previous MRI scans. The predictive value of Gd-enhancing lesions seems to be equivalent to that of the presence of new/enlarged (active) T2 lesions[148]. ~~Additionally, the enhancement, as happens with the presence of new lesions, has a role in demonstrating the dissemination in time, as defined in the revised McDonald criteria[21]. For the dissemination in space criteria, the recent MAGNIMS consensus guidelines[146] for the MRI criteria for the diagnosis of MS have suggested to include (i) cortical lesions (together with the juxta-cortical lesions); and (ii) optic nerve lesions. Yet at present, these lesions are looked for in selected, ad-hoc cases.~~

Over longer periods of observation, though, the number of new T2 lesions may be preferable to Gd-enhancing lesions to detect subclinical disease activity, as the latter only depicts disease activity in recent weeks. Other reasons for this include the higher costs associated to gadolinium usage and the fact that gadolinium infusions entail ~~some~~ rare medical risks, the most serious of which is the nephrogenic systemic fibrosis, although the risk may depend on the type of the gadolinium-containing contrast media[153]. Gadolinium can also deposit in the brain[154], yet the clinical consequences of this deposition remain unknown. Gadolinium administration is not recommended in routine MRI safety monitoring of patients receiving natalizumab[155].

Brain atrophy and other MRI measures

The use of atrophy in the clinic is currently controversial[156-158]. Although the contribution of brain atrophy to clinical and cognitive deficits is well-established at a

group level[148], there are several factors that may limit the application of atrophy in the clinical setting. These are: the lack of normative values for brain volume changes in healthy individuals and in patients with MS, the intra-individual variability, due to physiological variations (for example, dehydration, alcohol consumption), the presence of co-morbidities and disease-related factors, such as the initiation of a DMT, which may induce "pseudoatrophy"[97,103,148]. There a number of current techniques in development to try to overcome these issues: Jacobian integration[159] or lateral ventricle volume estimation[160], using T1-weighted or T2-weighted images, respectively, are being developed to improve the reliability of atrophy metrics in the clinic. It is important to bear in mind that ~~D~~differences in the MRI hardware and software packages used for analysis or processing can generate variability in brain atrophy measures[148]. Additionally, ~~;~~MRI scanner upgrades or replacements can make the images acquired at different time points ~~non-~~less comparable[161]. Ideally, of course, the same MS patient should be scanned on the same scanner and with the same protocol, whenever possible.

Combined clinical and MRI measures

A MAGNIMS study mentioned above showed that combining MRI activity with clinical relapses during the first year of treatment with interferon may identify patients who have a high risk of treatment failure and EDSS worsening in the short term[133]. In actual fact, escalation from first line DMT to a second line DMT is routinely advised in the clinical setting as a consequence of clinical and radiological evidence of disease activity.

There is no strong evidence to support the use of NEDA in clinical practice. In 2015 Rotstein et al. found, in a longitudinal study carried out in 219 patients, that those who maintained NEDA for 2 years had a very high probability (78.3%) of not showing any disability progression (defined as an increase in EDSS of >0.5 points), at 7 years of follow-up. However, a recent study that included 517 consecutive MS patients has found that achieving NEDA after the first two years of follow-up was not associated to a better prognosis at 10-year follow-up[162]. Although this was an observational study carried out in a heterogeneous cohort, where not all patients were on treatment ~~(which may have been adjusted based on MRI and clinical findings)~~, NEDA might not be a useful measure to predict a long-term outcome. In fact, it is likely that despite its high positive predictive value, NEDA has a low negative predictive value, so losing NEDA during the follow-up does not necessarily mean that prognosis is significantly worse, whereas maintaining NEDA is definitely a good prognostic marker. The implementation of NEDA-4, which includes brain atrophy, in the clinical setting is associated with the limitations described above and has not been validated for use in individual patients.

Translation from Trials to Clinical usage

We have demonstrated in the two sections above that most outcome measures used in clinical trials are not used in routine practice, and when they are, their use is limited and simplified. This is because in the clinical trials they are used for investigating drug effects at a population level, whilst in the clinical setting they are employed at the

individual level to assess the response to the medication (response measure), monitor patients (monitoring measure), or guide treatment decisions. In this section, we will compare the outcomes in clinical trials versus those used in the clinic. Although a translation of outcome measures used to demonstrate the effects of the drug to the clinical setting should be sought, there are elements in the clinical practice that go beyond treatment efficacy and influence patient management, such as patient's perception of risks and patient's priorities. ~~An attractive field of outcome measure which may overcome some barriers to the translation of outcome measures from trials to the clinical setting, such as the lack of time in the outpatient clinics, concerns the development of novel outcome measures driven by the introduction of electronic devices.~~

Outcomes in clinical trials versus monitoring in the clinic

Clinical or MRI outcome measures in clinical trials must be sensitive enough to be able to detect subtle, though highly relevant, treatment changes. This is especially important when the trial aims to compare a new drug not with placebo, but with another active drug[39]. In clinical trials, if the outcome measures are specific but not too sensitive, there may be a high risk of a falsely negative result, ultimately implying that a potentially efficacious drug may never be launched. Response measures in the clinic, instead, should probably be more specific than sensitive, since the consequences of prematurely (or incorrectly) starting or stopping a drug may have harmful consequences for the patient.

In clinical trials, clinical and MRI outcomes do not need to be meaningful at the individual level, as far as they are meaningful at the group level. For example, the outcome 'changes in MSFC z-scores' is only meaningful at the group level, and its usefulness stems from the comparison between treatment groups. In particular, it has been suggested that an increase in at least 20% in MSFC score or its subscores is a clinically relevant increase[48]. Instead, in the clinic, any type of monitoring instrument (or response measure) must be meaningful at the individual level. Importantly, in both clinical trials and the clinical setting, outcomes must reflect relevant functional or structural/pathological aspects of the condition and must be reproducible.

Regarding combined outcomes, whereas they have been extensively and successfully used in clinical trials, their use in the clinic will again depend on their meaningfulness at the individual level. Some of these combined outcomes, such as NEDA, have mainly been used in the trials, although they could be valid at the individual level and used in the clinic. In fact, when the factors associated with treatment response started to be defined[5], the underlying concept was the same as NEDA, although with a less restrictive threshold.

In relation to PROMs, their implementation in the clinic may be hampered by their inter and intra-patient variability. In clinical trials, this high variability may be compensated by large numbers. Further limitations for the use of PROMs in the clinic include that they can be time-consuming, that there is a very large number of measuring tools available without a clear evidence of superiority of one over the

others, and that the large amount of information that is produced needs to be interpreted and turned into useful data.

Another difference between outcomes in clinical trials and in the clinic is that in clinical trials there seems to be a trend towards a greater number of outcomes used over time (**Figure 2a**), whereas this is not happening in the clinic, where the EDSS score has been dominant for long time already. Interestingly, this increase in the number of trial endpoints is accompanied by a clear increase in the number of participants per trial (**Figure 2b**), which all together may be considered as an attempt to increase the power of the trial to detect a treatment effect, without prolonging the trial duration (**Figure 2c**).

Finally, we need to acknowledge that patients and clinicians may have a different perspective on what outcomes are relevant and desirable. For example, a comparison of the opinions and judgements of clinicians with those of patients utilising the short-form-36 showed that patients tend to prioritise general health and vitality, mental health, and emotional role limitation, whilst clinicians consider that physical disability, bodily pain and social functions are more important to the patient[163]. Undoubtedly, these are also factors that need to be taken into account when translating outcomes from trial to the clinic setting. Ultimately holistic approaches~~Thus~~, ~~rather holistic approaches~~ accommodating both patients' and clinicians' priorities, are probably preferred in the clinical setting, whereas this may not be a priority in clinical trials.

Conclusions

There are now over a dozen agents that can reduce the inflammatory component of MS, but there is an unmet and urgent need to treat progressive MS and promote tissue repair and neuroprotection. The availability of clinical and imaging measures in trials is of the utmost importance to ensure the detection of drug efficacy – nowhere more needed than in phase II trials of progression. The choice of the best set of outcomes for a given trial may be difficult because of the large amount of possible response measures described and used in the literature. Yet all trials should surely include clinical measures of disease progression, ideally based on the EDSS, for which there is a high experience, and other motor and/or cognitive measures, for which there is less experience, but which potentially have a higher sensitivity to capture subtle but relevant changes in disability. ~~Besides, t~~The time periods used to decide *confirmed* disability progression should be as long as possible, even 12 months if possible. Neuroimaging outcomes should include more traditional measures such as those related to lesion load, and also measures of brain atrophy. The inclusion of more novel measures is encouraged and their choice will possibly depend on the mechanism of action of the drug or the mechanistic research question that needs to be answered.

In the clinic, the choice of response measures determines the decisions about treatments and patient management. Although it would be ideal to use in the clinic the same tools to measure treatment response as those used in the clinical trial that led to licencing~~drug being licenced~~, at present, most of the endpoints used in trials cannot be used as response measures in the clinical setting. This is due to technical, financial and logistic barriers, such as the time required to obtain these measures,

training/standardisation, and the fact that their clinical meaning, when used at the individual level, is ~~very~~ limited. Most importantly, validated cut-off values that predict a favourable outcome in the long-run are lacking.

The use of PROMs and combined measures is important in both settings, since they capture the impact (and effects) of the intervention on clinical disability, MRI parameters, daily activities and quality of life. Further studies are needed to assess the reliability, accuracy and robustness of the combination of PROMs and objective (clinical and neuroimaging) measures, with the potential to comprehensively capture the intrinsic multidimensional nature of MS.

Review criteria

For this review paper, we performed searches in PubMed and www.clinical.trials.gov using the following search terms: 'multiple sclerosis', 'phase trial', 'EDSS', 'progression', 'relapse rate', 'MRI', 'neuroimaging', 'OCT', 'PROMS', 'cognition' (clinical trials sections); and 'multiple sclerosis', 'EDSS', 'progression', 'relapse rate', 'MRI', 'neuroimaging', 'OCT', 'PROMS', 'cognition', 'electronic devices'. We did not include any date limitations (the last date that we searched was June 2017). Papers were included in this review only if they were written in English. For the clinical trial section, only phase II or phase III controlled trials were included (uncontrolled and/or phase 0/I trials were not included).

Additional elements of the article

Tables: 3

- Table 1: Relapse-related and progression-related outcome measures used in phase III trials
- Table 2. Patient-reported outcome measures used as phase III trial endpoints
- Table 3: MRI outcome measures used in phase III trials
- Table 4: Strengths and weaknesses of outcome measures

Boxes: 1

- Box 1: Novel and future outcome measures
- Box 21: Main clinical and neuroimaging outcomes and outcome measures used in the clinical setting

Figures: 2

- Figure 1: Number of phase III trials over time in relapsing and progressive MS
- Figure 2: Trends over time in phase III trials: 2a: Evolution of number of trial endpoints over time; 2b: Evolution of number of participants per trial over time; 2c: Evolution of trial duration over time

Supplementary tables: 7

- Supplementary table 1: Clinical outcomes in phase III trials with relapsing MS
- Supplementary table 2: Clinical outcomes in phase III trials with CIS
- Supplementary table 3: Clinical outcomes in phase III trials with progressive MS
- Supplementary table 4: Brain MRI outcomes in phase III trials with relapsing MS

- Supplementary table 5: Brain MRI outcomes in phase III trials with CIS
- Supplementary table 6: Brain MRI outcomes in phase III trials with progressive MS
- Supplementary table 7: Trials with spinal cord MRI outcomes

Links to web sites

1. MS International Federation: http://www.msif.org
2. NICE guidelines for MS: https://www.nice.org.uk/guidance/cg186?unlid=719853888201626182413
3. NIH: http://www.ninds.nih.gov/disorders/multiple_sclerosis/multiple_sclerosis.htm
4. Clinical trials.gov: https://clinicaltrials.gov
5. Progressive MS Alliance: http://www.progressivemsalliance.org/about-us/2015-progress-report/

Tables

Table 1. Main relapse-related and progression-related outcome measures used in phase III trials

Outcome measure	Number of trials		Trials/References (in alphabetical order) (*: it indicates the outcome measure was the primary outcome)
	Primary outcome*	Primary or secondary outcome	
Relapse-related outcome measures – CIS trials			
Time to CDMS	6*	7	BENEFIT*[21], CHAMPS*[164], ETOMS*[75], ORACLE MS*[165], PreCISe*[88], REFLEX[22], TOPIC*[23]
%CDMS	0	5	BENEFIT[21], CHAMPS[164], ETOMS[75], REFLEX[22], TOPIC[23]
Time to McDonald MS	2*	3	BENEFIT*[21], REFLEX*[22], TOPIC[23]
% McDonald MS	0	3	BENEFIT[21], REFLEX[22], TOPIC[23]
Relapse-related outcome measures – MS trials			
Time to confirmed relapse	1*	18	BEYOND[11], CLARITY[166], CombiRx[101], CONFIRM[167], DEFINE[111], EudraCT 2006-004937-13[168], EUSPMS[169], EVIDENCE[170], FREEDOMS[14], GALA[171], NASPMS[41], PRISMS[17], REGARD*[172], SIMCOMBIN[173], SPECTRIMS[76,77], TEMSO[102], The copolymer 1 multiple sclerosis study[31], The Nordic SPMS study[64]
Time to confirmed relapse or permanent treatment discontinuation	1*	1	TENERE[15]
ARR	23*	41	ADVANCE*[16], AFFIRM*[123], ALLEGRO*[13], ARIANNA[40], BEYOND[11], BRAVO*[96], CARE-MS I*[29], CARE-MS II*[30], CLARITY*[166], CombiRx*[101], CONFIRM*[167], DECIDE*[174], DEFINE[111], ESIMS[55], ETOMS[75], EudraCT 2006-004937-13*[168], European/Canadian glatiramer acetate study[175], EUSPMS[169], EVIDENCE[170], FORTE*[176], FREEDOMS*[14], FREEDOMS II*[47], GALA*[171], GATE[81], LINOMIDE[177], MAESTRO[44], MSCRG[178], NASPMS[41], PRISMS*[17], REGARD[172], SENTINEL*[179], SIMCOMBIN*[173], SPECTRIMS[76,77], TEMSO*[102], TENERE[15], The copolymer 1 multiple sclerosis study*[31], The IFNb multiple sclerosis study*[180], The Nordic SPMS study[64], TOPIC[23], TOWER*[181], TRANSFORMS*[95]
ARSR	0	6	ALLEGRO[13], BEYOND[11], GALA[171], MAESTRO[44], PRISMS[17], SPECTRIMS[76,77], The IFNb multiple sclerosis study[180]
% at least one relapse	1*	9	ADVANCE[16], BEYOND[11], CombiRx[101], CONFIRM[167], DEFINE*[111], ESIMS[55], EudraCT 2006-004937-13[168], PreCISe[88], TENERE[15]

% relapse free	2*	28	AFFIRM[123], ALLEGRO[13], ARIANNA[40], BEYOND[11], BRAVO[96], CARE-MS I[29], CARE-MS II[30], CLARITY[166], CombiRx[101], DECIDE[174], EudraCT 2006-004937-13[168], EVIDENCE*[170], FORTE[176], FREEDOMS[14], FREEDOMS II[47], GALA[171], GATE[81], NASPMS[41], PRISMS[17], REGARD[172], SENTINEL[179], SIMCOMBIN[173], The copolymer 1 multiple sclerosis study[31], The IFNb Multiple Sclerosis Study*[180], The Nordic SMPS Study[64], TEMSO[102], TOWER[181], TRANSFORMS[95]
Other relapse-related measures: mean annualised rate of relapses requiring steroids, relapse risk*, time between first and second relapse	1*	2	BEYOND*[11], SPECTRIMS[76,77]
Progression-related outcome measures			
Change in EDSS	0	21	ARIANNA[40], CARE-MS I[29], CARE-MS II[30], ESIMS[55], ETOMS[75], EudraCT 2006-004937-13[168], EUSPMS[169], FREEDOMS[14], FREEDOMS II[47], GATE[81], MAESTRO[44], NASPMS[41], OLYMPUS[25], PRISMS[17], PROMISE[182], The Copolymer 1 Multiple Sclerosis study[31], The IFNb Multiple Sclerosis Study[180], The Nordic SMPS Study[64], TOPIC[23], TOWER[181], TRANSFORMS[95]
Change in MSFC or its subscores (PASAT, TWT, 9HPT)	1*	11	CARE-MS I[29], CARE-MS II[30], CombiRx[101], CUPID[65], FREEDOMS[14], FREEDOMS II[47], IMPACT*[43], MAESTRO[44], OLYMPUS[25], PROMISE[182], TRANSFORMS[95],
Change in other clinical scales (physical disability)	0	3	ETOMS[75], PRISMS[17], The Nordic SMPS Study[64]
Change in other clinical scales (cognitive disability)	0	2	IMPACT[43], MAESTRO[44]
% of 3m-CDP in EDSS	2*	23	ADVANCE[16], AFFIRM*[123], ALLEGRO[13], BEYOND[11], BRAVO[96], CONFIRM[167], DECIDE[174], DEFINE[111], ESIMS[55], ETOMS[75], EudraCT 2006-004937-13[168], EUSPMS[169], INFORMS[54], LINOMIDE[177], MSCRG[178], OLYMPUS[25], PROMISE[182], SENTINEL*[179], SIMCOMBIN[173], SPECTRIMS[76,77], TEMSO[102], The Copolymer 1 Multiple Sclerosis study[31], TOPIC[23]
% free from 3m-CDP in EDSS	0	7	CLARITY[166], FREEDOMS[14], FREEDOMS II[47], PRISMS[17], The Copolymer 1 Multiple Sclerosis study[31], TOWER[181], TRANSFORMS[95]
% of 6m-CDP in EDSS	0	10	ARIANNA[40], BRAVO[96], CARE-MS I[29], CARE-MS II[30], CombiRx[101], INFORMS[54], MAESTRO[44], OLYMPUS[25], REGARD[172], The Nordic SMPS Study
% free from 6m-CDP in EDSS	0	2	FREEDOMS[14], FREEDOMS II[47]

% sustained improvement in EDSS	0	2	CARE-MS II[30], The Copolymer 1 Multiple Sclerosis study[31]
% 3m-CDP in MSFC subscores	0	2	ESIMS[55], INFORMS[54]
% 6m-CDP in MSFC subscores	0	1	INFORMS[54]
% with 20% worsening in MSFC	0	1	ARIANNA[40]
Time to EDSS 7.0	0	1	EUSPMS[169]
Time to 3m-CDP in EDSS	8*	22	ALLEGRO[13], BEYOND[11], BRAVO[96], CLARITY[166], CONFIRM[167], DEFINE[111], ESIMS*[55], EUSPMS*[169], FREEDOMS[14], IMPACT[43], INFORMS[54], LINOMIDE*[177], MSCRG*[178], OLYMPUS*[25], ORATORIO*[26], PRISMS[17], PROMISE*[182], SIMCOMBIN[173], SPECTRIMS*[76,77], TEMSO[102], TOPIC[23], TOWER[181]
Time to 6m-CDP in EDSS	6*	12	ALLEGRO[13], BRAVO[96], CARE-MS I*[29], CARE-MS II*[30], CUPID*[65], FREEDOMS[14], INFORMS[54], MAESTRO*[44], NASPMS*[41], ORATORIO[26], SIMCOMBIN[173], The Nordic SMPS Study*
Time to 3m-CDP in MSFC subscores	0	2	ESIMS[55], INFORMS[54]
Time to 6m-CDP in MSFC subscores	0	1	INFORMS[54]
Clinical scores at follow-up	0	4	ALLEGRO[13], EUSPMS[169], PRISMS[17], The Copolymer 1 Multiple Sclerosis study[31]
Combined disability outcomes (including NECA)	1*	5	CARE-MS I[29], CARE-MS II[30], CombiRx[101], ESIMS[55], INFORMS*[54]

Footnote table 1. The primary endpoint of the ARIANNA study[40] was the changes in brain volume fraction (i.e. this study did not have a clinical primary endpoint). *Abbreviations:* ARR: annualised relapse rate; ARSR: annualised rate of severe relapses; CDMS: clinically defined multiple sclerosis; CDP: confirmed disability progression; EDSS: expanded disability status scale; 9HPT: nine-hole peg test; MSFC: multiple sclerosis functional composite; NECA: No evidence of clinical activity; PASAT: paced auditory serial addition test; TWT: 25-foot timed walk test.

Table 2. Main patient-reported outcome measures used as phase III trial endpoints

Outcome measure	Number of trials	Trials/References (in alphabetical order)
Arm index	2	PRISMS[17], The Nordic SMPS Study
PRIMUS	1	INFORMS[54]
EQ-5D/MSQoL-54	4	BENEFIT[21], FREEDOMS II[47], INFORMS[54], MAESTRO[44]
FIS	5	INFORMS[54], TEMSO[102], TENERE[15], TOPIC[23], TOWER[181]
MSWS-12	2	CUPID[65], INFORMS[54]
MSIS-29	2	CUPID[65], DECIDE[174]
SF-36	1	TOWER[181]
TSQM	1	TOWER[181]

Footnote table 2. *Abbreviations:* FIS (or UFIS): Unidimensional Fatigue Impact Scale; MSIS-29: Multiple Sclerosis Impact Scale – 29 items; MSWS-12: Multiple Sclerosis Walking Scale; SF-36: Short Form 36 Health Survey; PRIMUS: Patient-Reported Indices for Multiple Sclerosis; TSQM: Treatment Satisfaction Questionnaire for Medication, with domains for Effectiveness, Side-Effects, Convenience and Global Satisfaction.

Table 3. Main MRI outcome measures used in phase III trials

Outcome measure	Number of trials	Trials/References (in alphabetical order)
T2-lesion-related outcome measures		
# new T2 lesions	8	AFFIRM[123], BENEFIT[21], BEYOND[11], European/ Canadian Glatiramer Acetate Study, FORTE[176], IMPACT[43], PreCISe[88], The IFNb Multiple Sclerosis Study[180]
# enlarging T2 lesions	1	AFFIRM[123]
# new or enlarging T2 lesions	28	ADVANCE[16], AFFIRM[123], ALLEGRO[13], BRAVO[96], CARE-MS I[29], CARE-MS II[30], CHAMPS[164], CLARITY[166], CONFIRM[167], CUPID[65], DECIDE[174], DEFINE[111], ESIMS[55], ETOMS[75], EudraCT 2006-004937-13[168], EVIDENCE[170], FREEDOMS[14], FREEDOMS II[47], GALA[171], INFORMS[54], MAESTRO[44], ORACLE MS[165], PRISMS[17], REGARD[172], SENTINEL[179], SIMCOMBIN[173], TRANSFORMS[95], TEMSO[102]
Change in #T2 lesions	4	CombiRx[101], PreCISe[88], TEMSO[102], TOPIC[23]
Change in T2 lesion volume	33	ADVANCE[16], AFFIRM[123], BENEFIT[21], BEYOND[11], CARE-MS I[29], CARE-MS II[30], CHAMPS[164], CLARITY[166], CombiRx[101], CONFIRM[167], DECIDE[174], DEFINE[111], ESIMS[55], ETOMS[75], European/Canadian Glatiramer Acetate Study[175], EUSPMS[169], FREEDOMS[14], FREEDOMS II[47], IMPACT[43], MAESTRO[44], MSCRG[178], NASPMS[41], OLYMPUS[25], ORATORIO[26], PRISMS[17], PROMISE[182], REGARD[172], SIMCOMBIN[173], SPECTRIMS[76,77], TEMSO[102], The IFNb Multiple Sclerosis Study[180], TOPIC[23], TRANSFORMS[95]
Gadolinium-enhancing lesion-related outcome measures		
# Gd-enhancing T1 lesions at follow-up	36	ADVANCE[16], AFFIRM[123], ALLEGRO[13], BENEFIT[21], BEYOND[11], BRAVO[96], CARE-MS I[29], CARE-MS II[30], CHAMPS[164], CLARITY[166], CONFIRM[167], DECIDE[174], DEFINE[111], ESIMS[55], ETOMS[75], EudraCT 2006-004937-13[168], European/Canadian Glatiramer Acetate Study[175], FORTE[176], FREEDOMS[14], FREEDOMS II[47], GALA[171], GATE[81], IMPACT[43], INFORMS[54], MAESTRO[44], MSCRG[178], NASPMS[41], ORACLE MS[165], PROMISE[182], REGARD[172], SENTINEL[179], SPECTRIMS[76,77], TEMSO[102], The IFNb Multiple Sclerosis Study[180], TOPIC[23], TRANSFORMS[95]
% patients with Gd-enhancing lesions at follow-up	9	ARIANNA[40], CLARITY[166], FREEDOMS[14], FREEDOMS II[47], INFORMS[54], LINOMIDE[177,183], REGARD[172], TEMSO[102], TRANSFORMS[95]
Volume of Gd-enhancing lesions at follow-up	11	AFFIRM[123], BENEFIT[21], BEYOND[11], CONFIRM[167], DEFINE[111], European/Canadian Glatiramer Acetate Study[175], IMPACT[43], MSCRG[178], REGARD[172], TOPIC[23], TRANSFORMS[95]
Non-enhancing T1 lesion-related outcome measures		
# new non-enhancing T1 lesions	14	ADVANCE[16], AFFIRM[123], ALLEGRO[13], BENEFIT[21], CLARITY[166], CONFIRM[167], CUPID[65], DECIDE[174], DEFINE[111], GALA[171], INFORMS[54], TEMSO[102], TOPIC[23], REGARD[172]
Change in T1 lesion volume	14	ADVANCE[16], AFFIRM[123], BENEFIT[21], BEYOND[11], DECIDE[174], DEFINE[111], ESIMS[55], European/Canadian Glatiramer Acetate Study[175], FREEDOMS[14], FREEDOMS II[47], REGARD[172], SIMCOMBIN[173], TEMSO[102], TRANSFORMS[95]
Change in # T1 lesions	1	PreCISe[88]
Outcomes related to permanent black holes	1	ALLEGRO[13]
T1 and T2 lesion-related outcome measures		

Change in ratio T1/T2 volume	2	AFFIRM[123], ESIMS[55]
# combined unique active lesions	13	ADVANCE[16], BENEFIT[21], CLARITY[166], CombiRx[101], ETOMS[75], EudraCT 2006-004937-13[168], FREEDOMS[14], FREEDOMS II[47], ORACLE MS[165], REFLEX[22], REGARD[172], SPECTRIMS[76,77], TEMSO[102]
Combined lesional volume + CSF volume	1	CombiRx[101]
Non-lesion-related MRI outcome measures		
Change in whole-brain volume/fraction	25	ADVANCE[16], AFFIRM[123], ALLEGRO[13], ARIANNA[40], BEYOND[11], BRAVO[96], CARE-MS I[29], CARE-MS II[30], CONFIRM[167], CUPID[65], DEFINE[111], FORTE[176], FREEDOMS[14], FREEDOMS II[47], GALA[171], ESIMS[55], INFORMS[54], MAESTRO[44], OLYMPUS[25], ORATORIO[26], PreCISe[88],REGARD[172], SIMCOMBIN[173], TOPIC[23], TRANSFORMS[95]
Change in GM volume/fraction	3	ALLEGRO[13], CombiRx[101], TEMSO[102]
Change in WM volume/fraction	4	ALLEGRO[13], CombiRx[101], CUPID[65], TEMSO[102]
Change in thalamic volume	1	ALLEGRO[13]
Change in whole brain MTR	4	ADVANCE[16], ALLEGRO[13], CONFIRM[167], DEFINE[111]
Change in WM MTR	1	ALLEGRO[13]
Change in GM MTR	1	ALLEGRO[13]
Change in T2 lesion MTR	1	ALLEGRO[13]
Changes in the ratio NAA/creatinine	1	ALLEGRO[13]
Combined MRI and clinical outcomes		
NEDA (no evidence of disease activity)	3	CARE-MS I[29], CARE-MS II[30], CombiRx[101]

Footnote table 3. *Abbreviations:* CSF: cerebrospinal fluid; Gd: gadolinium; MTR: magnetisation transfer ratio; NAA/Cr: N-acetyl aspartate-creatine ratio;

Table 4 (New): Summary of the strengths and weaknesses of the main outcome measures

Outcome measure *Used in clinical trials (T), in the clinic (C) or in both (B)*	**Strengths** *In relation to their use in clinical trials (T), in the clinic (C) or in both (B)*	**Limitations** *In relation to their use in clinical trials (T), in the clinic (C) or in both (B)*
CLINICAL OUTCOME MEASURES		
Relapse-centred outcome measures		
# of relapses (C) or ARR (T)	Easy to compute and understand (B)	Only relevant for relapsing forms of MS (B) No specific for MS severity (B)
# of severe relapses (C) or ARSR (T)	May reflect severity of MS relapses (B)	High inter-site variability due to absence of guidelines for relapse management (T)
% of relapse-free patients (T)	In line with the concept of no disease activity, useful for trials with powerful drugs (T)	Highly dependent on trial duration, with statistical implications (see main text for more details) (T)
Time to confirmed relapse (T)	Useful in CIS trials (T)	Only relevant for relapsing forms of MS (B) No specific for MS severity (B)
Time to treatment failure (T)	Accounts for efficacy, safety and tolerability of the drug (i.e. reflects real-life scenario) (T)	Unspecific (T)
Measures of disability progression		
Change in EDSS and EDSS scores at follow-up (B)	Easy to understand by the MS community (B)	EDSS score changes do not reflect what functional system changes (B) Sensitive to relapse-related transient deficits (B) EDSS is not sensitive to upper limb or cognitive disability (B) Low inter- and intra-rater reproducibility (especially if low EDSS scores) (B)
Change in MSFC or its subscores and MSFC scores at follow-up (B)	No specific training required (B) Sensitive to upper limb (NHPT) and cognitive (PASAT) functions (B) In the clinic, TWT is useful to monitor drug effects, such as fampridine (C)	Designed to be used in trials, at group level (i.e. reduced usefulness in the clinic) (C) Definition of clinically meaningful change is required (mainly CT) Choice of a reference population affects z-scores (T) Practice effects (B) PASAT may be stressful (B)
Change in other clinical (mainly cognitive) scales (B)	For SDMT, no specific training required (B) Sensitive to cognitive impairment (B)	Training may be required for cognitive tests (exc. SDMT) (B) Reference population often needed to interpret results (C) Age, anxiety, fatigue and education may influence results (B)
% of 3m/6m-CDP in EDSS (T)	Easy to understand by the MS community (T)	Overestimation of long-term disability accumulation (T)

		Highly dependent on trial duration, with statistical implications (T)
% free from 3m/6m-CDP in EDSS (T)	Easy to understand by the MS community (T) In line with the concept of no disease activity, useful for trials with powerful drugs (T)	Underestimation of % patients free from long-term disability accumulation (T) Highly dependent on trial duration (T)
% sustained improvement in EDSS (T)	Useful to detect improvements of disability, largely overlooked in MS trials (T)	May be unspecific in relation to the pathophysiological process underlying clinical improvement (T)
% 3m/6m-CDP in MSFC subscores (T)	Strengths of the MSFC-related outcome measures and outcome measures that consider progression as a binary phenomenon (see above) (T)	Limitations of the MSFC-related outcome measures and outcome measures that consider % patients with disability progression (see above) (T)
Time to 3m/6m-CDP in EDSS/MSFC and time to a given EDSS/MSFC score (T)	Strengths of EDSS/MSFC-related measures (T) Informative about the effect of the drug on immediate risk of CDP (as opposed to '% patients with CDP', which considers the risk over a relatively long period) (T)	Limitations of EDSS/MSFC-related measures (T)
Combined disability outcomes (including NECA) (B)	Higher sensitivity than individual components to detection of disability progression, implying a reduction in required sample sizes/trial durations (T) Reduction of the risk of type I error (T) NECA: comprehensive measure of real-life treatment effect (B)	Individual components cannot be analysed independently, unless they were pre-defined as outcome measures (T) Composite outcomes must include measures causally related to treatment (T)
PROMs		
All PROMs (B)	Information comes directly from the patient (B)	Information is subjective and may fluctuate within subjects (B)
NEUROIMAGING & NEUROPHYSIOLOGICAL OUTCOME MEASURES		
Outcome measures related to focal lesions		
T2-lesion-related outcome measures (B)	Information on new and cumulative inflammatory activity (B)	Temporal frameworks for new inflammatory activity are imprecise (B)
Gd-enhancing lesion-related outcome measures (B)	Information on recent inflammatory activity (within 3-6 weeks prior scan date) (B)	No information on cumulative inflammatory activity (B)
Non-enhancing T1 lesion-related outcome measures (B)	May inform about tissue destruction secondary to inflammation and repair (B)	The delineation of hypointense T1 lesions may depend on scanner parameters (B)
Combined unique active lesions (B)	More sensitive than new T2 or gadolinium-enhancing lesions separately (B)	Their computation is slightly more complex than new T2 or gadolinium-enhancing lesions (B)
Non-lesion-related MRI outcome measures		
Brain atrophy-related metrics (B)	Reflect neurodegeneration, the most important substrate of disability accrual (B)	Susceptible to pseudo-atrophy phenomenon (B)

		High intra-subject (physiological) variation (B)
Spinal cord atrophy-related metrics (T)	Reflect neurodegeneration in the spinal cord, highly related to motor disability (T)	Limited spatial resolution, which hampers multi-centre studies (T) Current segmentation methods are semi-automated, implying high inter-rater variability (T)
Novel imaging outcomes (MTR, MR spectroscopy, diffusion-weighted, PET-derived metrics) (T)	Information on microstructural features of brain damage, complementary to that given by lesion-related or atrophy-related measures (T)	Standardisation of acquisition protocols and analysis methods still in progress (T)
OCT (B)	Information on axonal and neuronal loss within the anterior visual pathway (related to neurodegeneration) (B) Useful to monitor drugs' side effects (fingolimod) (B)	Less reliable if previous history of optic neuritis (B)
Combined MRI and clinical outcomes (including NEDA) (B)	NEDA: comprehensive measure of real-life treatment effect (B)	Difficult interpretation of the net effect of drugs on the outcome measure (T) Reduced usefulness in the clinic (high positive predictive value but low negative predictive value) (C)
VEPs (T)	May reflect remyelinating processes secondary to experimental drugs (T)	Not sensitive enough to monitor disease progression (B)

Footnote table 4. *Abbreviations:* ARR: annualised relapse rate; B: both clinical trial and clinical setting; ARSR: annualised rate of severe relapses; C: clinical setting; CDMS: clinically defined multiple sclerosis; CDP: confirmed disability progression; EDSS: expanded disability status scale; Gd: gadolinium; 9HPT: nine-hole peg test; MSFC: multiple sclerosis functional composite; MTR: magnetisation transfer ratio; NECA: No evidence of clinical activity; NEDA: No evidence of disease activity; PASAT: paced auditory serial addition test; OCT: optical coherence tomography; PET: positron emission tomography; PROMs: patient-reported outcome measures; SDMT: symbol digit modalities test; T: clinical trial; TWT: 25-foot timed walk test; VEPs: visual evoked potentials.

Boxes

Box 1 (New). Novel and future outcome measures

Possible future clinical outcomes include those obtained through the utilisation of 'smart' technology such as wearable sensors have started to be developed for their use mainly in the clinic. Wearable sensors are electronic devices that can be attached to the body and record information about the user's quantity and quality of movement. This portable technology can provide objective and quantitative data[184] which may be useful to detect response to therapeutic interventions in the real life. Besides, several strategies have been developed to maximise the sensitivity to disease progression of current disability scores. These include re-baselining the EDSS score according to both screening and first visits, and using new metrics such as the area under the curve described by the disability score trajectories over time[131].

Possible future imaging outcomes include markers of remyelination, such as within-lesion MTR[185] or the level of [^{11}C]PIB binding[186], obtained with positron emission tomography (PET). Markers of chronic inflammation, such as the presence of slowly enlarging lesions[187], and microglial activation, such as and level of TSPO binding[188-190], also obtained through PET, can be used as future outcome measures too. These potential outcomes can bring us closer to achieving precision medicine[189].

Advanced OCT techniques provide quantitative measurements of both retinal nerve fibre layer (RNFL, axonal) and ganglion cell layer (GCL, neuronal) loss in vivo, representing an ideal model for assessing the neuroprotective effects of novel agents[118]. Possible advantages of OCT in trials are that the evaluation of the retinal structure might predict the clinical response to treatment[191] and the risk of developing specific ocular side effects[192].

Finally, future neurophysiological outcomes would include visual evoked potentials and multimodal evoked potentials, which have shown some ability to predict clinical evolution in patients with MS[193-195]. Change in full-field VEPs latency at week-24 has been used as the primary outcome measure in a phase 2 trial assessing the efficacy of a remyelinating therapy after the first episode of optic neuritis[196]

Box 2. Main clinical and neuroimaging outcomes and derived outcome measures used in the clinical setting

Clinical outcomes

Relapses
- Number of relapses over a period of time

EDSS
- EDSS score at a given time point
- Change in EDSS score over a period of time

TWT
- TWT score (measured in seconds) at a given time point

9HPT
- 9HPT score (measured in seconds) at a given time point

PASAT
- Number of successes (maximum: 60) during the test

SDMT
- Number of successes (no maximum) during the test (usually 1 minute)

FIS/FSS/MFIS
- Score at a given time point

Neuroimaging outcomes

Brain T2 lesions
- Number of lesions at a given time point
- Number of new or enlarging lesions

Brain Gd-enhancing lesions
- Number of lesions at a given time point

Brain non-enhancing T1 lesions
- Number of lesions at a given time point

Brain cortical lesions (in DIR sequences)
- Number of lesions at a given time point

Spinal cord T2 lesions
- Number of lesions at a given time point

Abbreviations:
DIR: double inversion recovery; EDSS: Expanded Disability Status Scale; FIS (or UFIS): Unidimensional Fatigue Impact Scale; FSS: fatigue severity scale; 9HPT: Nine-Hole Peg Test; MFIS: modified fatigue impact scale; PASAT: Paced Auditory Serial Addition Test; SDMT: symbol digit modalities test; TWT: 25-Foot Timed Walk Test;

Figure legends

Figure1: Number of phase III trials over time in relapsing and progressive MS

Figure 1 (legend).
This figure illustrates the increase in the number of phase III clinical trials carried out over the last five years, especially in relapsing MS patients. *Abbreviations:* CIS: clinically isolated syndrome; MS: multiple sclerosis; PPMS: primary progressive multiple sclerosis; SPMS: secondary progressive multiple[1] sclerosis.

Figure 2: Trends over time in phase III trials

2a: Evolution of number of trial endpoints over time;
2b: Evolution of number of participants per trial over time;
2c: Evolution of trial duration over time.

Figure 2 (legend).
This figure illustrates the evolution over time of (a) the number of trial endpoints per trial; (b) number of participants per trial; (c) trial duration. As can be observed, there has been a clear increase in the number of trial endpoints per trial and the number of participants per trial over the last 5-10 years, whereas the trial duration has remained very similar. Most of the trials have a duration of 2 or 3 years. *Abbreviations:* MS: multiple sclerosis;

References

1 Compston, A. & Coles, A. Multiple sclerosis. *Lancet* **372**, 1502-1517, doi:10.1016/S0140-6736(08)61620-7 (2008).

2 Trapp, B. D. & Nave, K. A. Multiple sclerosis: an immune or neurodegenerative disorder? *Annu Rev Neurosci* **31**, 247-269, doi:10.1146/annurev.neuro.30.051606.094313 (2008).

3 Alliance, P. M. <http://www.progressivemsalliance.org/> (

4 Bevan, C. J. & Cree, B. A. Disease activity free status: a new end point for a new era in multiple sclerosis clinical research? *JAMA Neurol* **71**, 269-270, doi:10.1001/jamaneurol.2013.5486 (2014).

5 Rio, J. *et al.* Defining the response to interferon-beta in relapsing-remitting multiple sclerosis patients. *Ann Neurol* **59**, 344-352, doi:10.1002/ana.20740 (2006).

6 Rio, J., Comabella, M. & Montalban, X. Predicting responders to therapies for multiple sclerosis. *Nat Rev Neurol* **5**, 553-560, doi:10.1038/nrneurol.2009.139 (2009).

7 Hobart, J. C., Cano, S. J., Zajicek, J. P. & Thompson, A. J. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* **6**, 1094-1105, doi:10.1016/S1474-4422(07)70290-9 (2007).

8 Coster, W. J. Making the best match: selecting outcome measures for clinical trials and outcome studies. *Am J Occup Ther* **67**, 162-170, doi:10.5014/ajot.2013.006015 (2013).

9 US Department of Health and Human Services Food and Drug Administration, F. D. A. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. (2009).

10 Fox, R. J. *et al.* Setting a research agenda for progressive multiple sclerosis: the International Collaborative on Progressive MS. *Mult Scler* **18**, 1534-1540, doi:10.1177/1352458512458169 (2012).

11 O'Connor, P. *et al.* 250 microg or 500 microg interferon beta-1b versus 20 mg glatiramer acetate in relapsing-remitting multiple sclerosis: a prospective, randomised, multicentre study. *Lancet Neurol* **8**, 889-897, doi:10.1016/S1474-4422(09)70226-1 (2009).

12 Uitdehaag, B. M. *et al.* The changing face of multiple sclerosis clinical trial populations. *Curr Med Res Opin* **27**, 1529-1537, doi:10.1185/03007995.2011.591370 (2011).

13 Comi, G. *et al.* Placebo-controlled trial of oral laquinimod for multiple sclerosis. *N Engl J Med* **366**, 1000-1009, doi:10.1056/NEJMoa1104318 (2012).

14 Kappos, L. *et al.* A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *N Engl J Med* **362**, 387-401, doi:10.1056/NEJMoa0909494 (2010).

15 Vermersch, P. *et al.* Teriflunomide versus subcutaneous interferon beta-1a in patients with relapsing multiple sclerosis: a randomised, controlled phase 3 trial. *Mult Scler* **20**, 705-716, doi:10.1177/1352458513507821 (2014).

16 Calabresi, P. A. *et al.* Pegylated interferon beta-1a for relapsing-remitting multiple sclerosis (ADVANCE): a randomised, phase 3, double-blind study. *Lancet Neurol* **13**, 657-665, doi:10.1016/S1474-4422(14)70068-7 (2014).

17 Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group. *Lancet* **352**, 1498-1504 (1998).

18 Polman, C. H. *et al.* Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann Neurol* **58**, 840-846, doi:10.1002/ana.20703 (2005).

19 Polman, C. H. *et al.* Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* **69**, 292-302, doi:10.1002/ana.22366 (2011).

20 McDonald, W. I. *et al.* Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* **50**, 121-127 (2001).

21 Kappos, L. *et al.* Treatment with interferon beta-1b delays conversion to clinically definite and McDonald MS in patients with clinically isolated syndromes. *Neurology* **67**, 1242-1249, doi:10.1212/01.wnl.0000237641.33768.8d (2006).

22 Comi, G. *et al.* Comparison of two dosing frequencies of subcutaneous interferon beta-1a in patients with a first clinical demyelinating event suggestive of multiple sclerosis (REFLEX): a phase 3 randomised controlled trial. *Lancet Neurol* **11**, 33-41, doi:10.1016/S1474-4422(11)70262-9 (2012).

23 Miller, A. E. *et al.* Oral teriflunomide for patients with a first clinical episode suggestive of multiple sclerosis (TOPIC): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Neurol* **13**, 977-986, doi:10.1016/S1474-4422(14)70191-7 (2014).

24 Business Wire, B. W. (http://www.businesswire.com/news/home/20151021005273/en/Biogen-Reports-Top-Line-Results-Phase-3-Study, 2015).

25 Hawker, K. *et al.* Rituximab in patients with primary progressive multiple sclerosis: results of a randomized double-blind placebo-controlled multicenter trial. *Ann Neurol* **66**, 460-471, doi:10.1002/ana.21867 (2009).

26 Montalban, X. *et al.* Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. *N Engl J Med*, doi:10.1056/NEJMoa1606468 (2016).

27 MS Society, M. <https://www.mssociety.org.uk/ms-research/treatments-in-the-pipeline/siponimod> (

28 Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* **33**, 1444-1452 (1983).

29 Cohen, J. A. *et al.* Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting multiple sclerosis: a randomised controlled phase 3 trial. *Lancet* **380**, 1819-1828, doi:10.1016/S0140-6736(12)61769-3 (2012).

30 Coles, A. J. *et al.* Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: a randomised controlled phase 3 trial. *Lancet* **380**, 1829-1839, doi:10.1016/S0140-6736(12)61768-1 (2012).

31 Johnson, K. P. *et al.* Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. The Copolymer 1 Multiple Sclerosis Study Group. *Neurology* **45**, 1268-1276 (1995).

32 Lublin, F. D., Baier, M. & Cutter, G. Effect of relapses on development of residual deficit in multiple sclerosis. *Neurology* **61**, 1528-1532 (2003).

33 Ebers, G. C., Heigenhauser, L., Daumer, M., Lederer, C. & Noseworthy, J. H. Disability as an outcome in MS clinical trials. *Neurology* **71**, 624-631, doi:10.1212/01.wnl.0000313034.46883.16 (2008).

34 Noseworthy, J. H., Vandervoort, M. K., Wong, C. J. & Ebers, G. C. Interrater variability with the Expanded Disability Status Scale (EDSS) and Functional Systems (FS) in a multiple sclerosis clinical trial. The Canadian Cooperation MS Study Group. *Neurology* **40**, 971-975 (1990).

35 Mahoney, F. I. & Barthel, D. W. Functional Evaluation: The Barthel Index. *Md State Med J* **14**, 61-65 (1965).

36 Hobart, J. C. *et al.* Getting the measure of spasticity in multiple sclerosis: the Multiple Sclerosis Spasticity Scale (MSSS-88). *Brain* **129**, 224-234, doi:10.1093/brain/awh675 (2006).

37 Hauser, S. L. *et al.* Intensive immunosuppression in progressive multiple sclerosis. A randomized, three-arm study of high-dose intravenous cyclophosphamide, plasma exchange, and ACTH. *N Engl J Med* **308**, 173-180, doi:10.1056/NEJM198301273080401 (1983).

38 Goodkin, D. E., Hertsgaard, D. & Seminary, J. Upper extremity function in multiple sclerosis: improving assessment sensitivity with box-and-block and nine-hole peg tests. *Arch Phys Med Rehabil* **69**, 850-854 (1988).

39 Cutter, G. R. *et al.* Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain* **122 (Pt 5)**, 871-882 (1999).

40 Lanzillo, R. *et al.* No evidence for an effect on brain atrophy rate of atorvastatin add-on to interferon beta1b therapy in relapsing-remitting multiple sclerosis (the ARIANNA study). *Mult Scler* **22**, 1163-1173, doi:10.1177/1352458515611222 (2016).

41 Panitch, H., Miller, A., Paty, D., Weinshenker, B. & North American Study Group on Interferon beta-1b in Secondary Progressive, M. S. Interferon beta-1b in secondary progressive MS: results from a 3-year controlled study. *Neurology* **63**, 1788-1795 (2004).

42 Rao, S. & Cognitive Function Group, N. *A manual for the Brief Repeatable Battery of neuropsychological tests in multiple sclerosis*. (National Multiple Sclerosis Society, New York, 1990).

43 Cohen, J. A. *et al.* Benefit of interferon beta-1a on MSFC progression in secondary progressive MS. *Neurology* **59**, 679-687 (2002).

44 Freedman, M. S. *et al.* A phase III study evaluating the efficacy and safety of MBP8298 in secondary progressive MS. *Neurology* **77**, 1551-1560, doi:10.1212/WNL.0b013e318233b240 (2011).

45 Kalincik, T. *et al.* Defining reliable disability outcomes in multiple sclerosis. *Brain* **138**, 3287-3298, doi:10.1093/brain/awv258 (2015).

46 Kragt, J. J. *et al.* Responsiveness and predictive value of EDSS and MSFC in primary progressive MS. *Neurology* **70**, 1084-1091, doi:10.1212/01.wnl.0000288179.86056.e1 (2008).

47 Calabresi, P. A. *et al.* Safety and efficacy of fingolimod in patients with relapsing-remitting multiple sclerosis (FREEDOMS II): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Neurol* **13**, 545-556, doi:10.1016/S1474-4422(14)70049-3 (2014).

48 Bosma, L. V. *et al.* Progression on the Multiple Sclerosis Functional Composite in multiple sclerosis: what is the optimal cut-off for the three components? *Mult Scler* **16**, 862-867, doi:10.1177/1352458510370464 (2010).

49 Tourbah, A. *et al.* MD1003 (high-dose biotin) for the treatment of progressive multiple sclerosis: A randomised, double-blind, placebo-controlled study. *Mult Scler* **22**, 1719-1731, doi:10.1177/1352458516667568 (2016).

50 Chataway, J. *et al.* Effect of high-dose simvastatin on brain atrophy and disability in secondary progressive multiple sclerosis (MS-STAT): a randomised, placebo-controlled, phase 2 trial. *Lancet* **383**, 2213-2221, doi:10.1016/S0140-6736(13)62242-4 (2014).

51 ClinicalTrials.gov. *US National Library of Medicine at the National Institutes of Health*, <http://www.clinicaltrials.gov/ (access date: July 2017)> (

52 Zhang, J., Waubant, E., Cutter, G., Wolinsky, J. & Leppert, D. Composite end points to assess delay of disability progression by MS treatments. *Mult Scler* **20**, 1494-1501, doi:10.1177/1352458514527180 (2014).

53 Prieto-Merino, D., Smeeth, L., Staa, T. P. & Roberts, I. Dangers of non-specific composite outcome measures in clinical trials. *BMJ* **347**, f6782, doi:10.1136/bmj.f6782 (2013).

54 Lublin, F. *et al.* Oral fingolimod in primary progressive multiple sclerosis (INFORMS): a phase 3, randomised, double-blind, placebo-controlled trial. *Lancet* **387**, 1075-1084, doi:10.1016/S0140-6736(15)01314-8 (2016).

55 Hommes, O. R. *et al.* Intravenous immunoglobulin in secondary progressive multiple sclerosis: randomised placebo-controlled trial. *Lancet* **364**, 1149-1156, doi:10.1016/S0140-6736(04)17101-8 (2004).

56 Speight, J. & Barendse, S. M. FDA guidance on patient reported outcomes. *BMJ* **340**, c2921, doi:10.1136/bmj.c2921 (2010).

57 Hobart, J. & Cano, S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* **13**, iii, ix-x, 1-177, doi:10.3310/hta13120 (2009).

58 Doward, L. C., McKenna, S. P., Meads, D. M., Twiss, J. & Eckert, B. J. The development of patient-reported outcome indices for multiple sclerosis (PRIMUS). *Mult Scler* **15**, 1092-1102, doi:10.1177/1352458509106513 (2009).

59 Vickrey, B. G., Hays, R. D., Harooni, R., Myers, L. W. & Ellison, G. W. A health-related quality of life measure for multiple sclerosis. *Qual Life Res* **4**, 187-206 (1995).

60 Krupp, L. B., LaRocca, N. G., Muir-Nash, J. & Steinberg, A. D. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol* **46**, 1121-1123 (1989).

61 Frith, J. & Newton, J. Fatigue Impact Scale. *Occup Med (Lond)* **60**, 159, doi:10.1093/occmed/kqp180 (2010).

62 Hobart, J. C., Riazi, A., Lamping, D. L., Fitzpatrick, R. & Thompson, A. J. Measuring the impact of MS on walking ability: the 12-Item MS Walking Scale (MSWS-12). *Neurology* **60**, 31-36 (2003).

63 Hobart, J., Cano, S., Ingram, W. & Zajicek, J. A new path for the MS Walking Scale: MSWS-12 version 2. . *Mult Scler.* **18**, 334–335 (2012).

64 Andersen, O. *et al.* Multicentre, randomised, double blind, placebo controlled, phase III study of weekly, low dose, subcutaneous interferon beta-1a in secondary progressive multiple sclerosis. *J Neurol Neurosurg Psychiatry* **75**, 706-710 (2004).

65 Zajicek, J. *et al.* Effect of dronabinol on progression in progressive multiple sclerosis (CUPID): a randomised, placebo-controlled trial. *Lancet Neurol* **12**, 857-865, doi:10.1016/S1474-4422(13)70159-5 (2013).

66 Hobart, J. Prolonged-release fampridine for multiple sclerosis: was the effect on walking ability clinically significant? . *Mult Scler* **16(Suppl 10):S172.** (2010).

67 Goodman, A. D. *et al.* Sustained-release oral fampridine in multiple sclerosis: a randomised, double-blind, controlled trial. *Lancet* **373**, 732-738, doi:10.1016/S0140-6736(09)60442-6 (2009).

68 Novotna, A. *et al.* A randomized, double-blind, placebo-controlled, parallel-group, enriched-design study of nabiximols* (Sativex((R))), as add-on therapy, in subjects with refractory spasticity caused by multiple sclerosis. *Eur J Neurol* **18**, 1122-1131, doi:10.1111/j.1468-1331.2010.03328.x (2011).

69 Zajicek, J. *et al.* Cannabinoids for treatment of spasticity and other symptoms related to multiple sclerosis (CAMS study): multicentre randomised placebo-controlled trial. *Lancet* **362**, 1517-1526, doi:10.1016/S0140-6736(03)14738-1 (2003).

70 Learmonth, Y. C., Motl, R. W., Sandroff, B. M., Pula, J. H. & Cadavid, D. Validation of patient determined disease steps (PDDS) scale scores in persons with multiple sclerosis. *BMC Neurol* **13**, 37, doi:10.1186/1471-2377-13-37 (2013).

71 Barkhof, F. *et al.* MRI monitoring of immunomodulation in relapse-onset multiple sclerosis trials. *Nat Rev Neurol* **8**, 13-21, doi:10.1038/nrneurol.2011.190 (2011).

72 Cotton, F., Weiner, H. L., Jolesz, F. A. & Guttmann, C. R. MRI contrast uptake in new lesions in relapsing-remitting MS followed at weekly intervals. *Neurology* **60**, 640-646 (2003).

73 Filippi, M., Preziosa, P. & Rocca, M. A. Magnetic resonance outcome measures in multiple sclerosis trials: time to rethink? *Curr Opin Neurol* **27**, 290-299, doi:10.1097/WCO.0000000000000095 (2014).

74 Mitjana, R. *et al.* Diagnostic value of brain chronic black holes on T1-weighted MR images in clinically isolated syndromes. *Mult Scler* **20**, 1471-1477, doi:10.1177/1352458514526083 (2014).

75 Comi, G. *et al.* Effect of early interferon treatment on conversion to definite multiple sclerosis: a randomised study. *Lancet* **357**, 1576-1582 (2001).

76 Secondary Progressive Efficacy Clinical Trial of Recombinant Interferon-Beta-1a in, M. S. S. G. Randomized controlled trial of interferon- beta-1a in secondary progressive MS: Clinical results. *Neurology* **56**, 1496-1504 (2001).

77 Li, D. K., Zhao, G. J., Paty, D. W. & University of British Columbia, M. S. M. R. I. A. R. G. T. S. S. G. Randomized controlled trial of interferon-beta-1a in secondary progressive MS: MRI results. *Neurology* **56**, 1505-1513 (2001).

78 Tintore, M. *et al.* Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain* **138**, 1863-1874, doi:10.1093/brain/awv105 (2015).

79 Sormani, M. P. *et al.* Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: a meta-analytic approach. *Ann Neurol* **65**, 268-275, doi:10.1002/ana.21606 (2009).

80 Sormani, M. P. & Bruzzi, P. MRI lesions as a surrogate for relapses in multiple sclerosis: a meta-analysis of randomised trials. *Lancet Neurol* **12**, 669-676, doi:10.1016/S1474-4422(13)70103-0 (2013).

81 Cohen, J. *et al.* Equivalence of Generic Glatiramer Acetate in Multiple Sclerosis: A Randomized Clinical Trial. *JAMA Neurol* **72**, 1433-1441, doi:10.1001/jamaneurol.2015.2154 (2015).

82 Moraal, B. *et al.* Long-interval T2-weighted subtraction magnetic resonance imaging: a powerful new outcome measure in multiple sclerosis trials. *Ann Neurol* **67**, 667-675, doi:10.1002/ana.21958 (2010).

83 Tedeschi, E. *et al.* In vivo dentate nucleus MRI relaxometry correlates with previous administration of Gadolinium-based contrast agents. *Eur Radiol* **26**, 4577-4584, doi:10.1007/s00330-016-4245-2 (2016).

84 Fisniku, L. K. *et al.* Gray matter atrophy is related to long-term disability in multiple sclerosis. *Ann Neurol* **64**, 247-254, doi:10.1002/ana.21423 (2008).

85 Miller, D. H., Barkhof, F., Frank, J. A., Parker, G. J. & Thompson, A. J. Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. *Brain* **125**, 1676-1695 (2002).

86 Sormani, M. P., Arnold, D. L. & De Stefano, N. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. *Ann Neurol* **75**, 43-49, doi:10.1002/ana.24018 (2014).

87 Barkhof, F., Giovannoni, G., Hartung, H.-P., Cree, B., Uccelli, A., Sormani, M.P., Krieger, S., Uitdehaag, B., Vollmer, T., Montalban, X., Steinerman, J., Sasson, N., Gorfine, T. and Knappertz, V. ARPEGGIO: A randomized, placebo-controlled study to evaluate oral laquinimod in patients with primary progressive multiple sclerosis (PPMS). *Neurology* **84** (2015).

88 Comi, G. *et al.* Effect of glatiramer acetate on conversion to clinically definite multiple sclerosis in patients with clinically isolated syndrome (PreCISe study): a randomised, double-blind, placebo-controlled trial. *Lancet* **374**, 1503-1511, doi:10.1016/S0140-6736(09)61259-9 (2009).

89 Kapoor, R. *et al.* Lamotrigine for neuroprotection in secondary progressive multiple sclerosis: a randomised, double-blind, placebo-controlled, parallel-group trial. *Lancet Neurol* **9**, 681-688, doi:10.1016/S1474-4422(10)70131-9 (2010).

90 Rudick, R. A., Fisher, E., Lee, J. C., Simon, J. & Jacobs, L. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. Multiple Sclerosis Collaborative Research Group. *Neurology* **53**, 1698-1704 (1999).

91 Smith, S. M., De Stefano, N., Jenkinson, M. & Matthews, P. M. Normalized accurate measurement of longitudinal brain change. *J Comput Assist Tomogr* **25**, 466-475 (2001).

92 Smith, S. M. *et al.* Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* **17**, 479-489 (2002).

93 Miller, D. H. *et al.* MRI outcomes in a placebo-controlled trial of natalizumab in relapsing MS. *Neurology* **68**, 1390-1401, doi:10.1212/01.wnl.0000260064.77700.fd (2007).

94 Montalban, X. *et al.* A single-center, randomized, double-blind, placebo-controlled study of interferon beta-1b on primary progressive and transitional multiple sclerosis. *Mult Scler* **15**, 1195-1205, doi:10.1177/1352458509106937 (2009).

95 Cohen, J. A. *et al.* Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis. *N Engl J Med* **362**, 402-415, doi:10.1056/NEJMoa0907839 (2010).

96 Vollmer, T. L. *et al.* A randomized placebo-controlled phase III trial of oral laquinimod for multiple sclerosis. *J Neurol* **261**, 773-783, doi:10.1007/s00415-014-7264-4 (2014).

97 Vidal-Jordana, A. *et al.* Brain Volume Loss During the First Year of Interferon-Beta Treatment in Multiple Sclerosis: Baseline Inflammation and Regional Brain Volume Dynamics. *J Neuroimaging* **26**, 532-538, doi:10.1111/jon.12337 (2016).

98 Molyneux, P. D. *et al.* The effect of interferon beta-1b treatment on MRI measures of cerebral atrophy in secondary progressive multiple sclerosis. European Study Group on Interferon beta-1b in secondary progressive multiple sclerosis. *Brain* **123 (Pt 11)**, 2256-2263 (2000).

99 Nakamura, K. *et al.* Diurnal fluctuations in brain volume: Statistical analyses of MRI from large populations. *Neuroimage* **118**, 126-132, doi:10.1016/j.neuroimage.2015.05.077 (2015).

100 Nunes, P. V. *et al.* Factors associated with brain volume in major depression in older adults without dementia: results from a large autopsy study. *Int J Geriatr Psychiatry*, doi:10.1002/gps.4649 (2017).

101 Lublin, F. D. *et al.* Randomized study combining interferon and glatiramer acetate in multiple sclerosis. *Ann Neurol* **73**, 327-340, doi:10.1002/ana.23863 (2013).

102 O'Connor, P. *et al.* Randomized trial of oral teriflunomide for relapsing multiple sclerosis. *N Engl J Med* **365**, 1293-1303, doi:10.1056/NEJMoa1014656 (2011).

103 Vidal-Jordana, A. *et al.* Early brain pseudoatrophy while on natalizumab therapy is due to white matter volume changes. *Mult Scler* **19**, 1175-1181, doi:10.1177/1352458512473190 (2013).

104 Vidal-Jordana, A., Sastre-Garriga, J., Rovira, A. & Montalban, X. Treating relapsing-remitting multiple sclerosis: therapy effects on brain atrophy. *J Neurol* **262**, 2617-2626, doi:10.1007/s00415-015-7798-0 (2015).

105 Fisher, E., Lee, J. C., Nakamura, K. & Rudick, R. A. Gray matter atrophy in multiple sclerosis: a longitudinal study. *Ann Neurol* **64**, 255-265, doi:10.1002/ana.21436 (2008).

106 Brownlee, W. J. *et al.* Association of asymptomatic spinal cord lesions and atrophy with disability 5 years after a clinically isolated syndrome. *Mult Scler*, doi:10.1177/1352458516663034 (2016).

107 Lin, X., Tench, C. R., Turner, B., Blumhardt, L. D. & Constantinescu, C. S. Spinal cord atrophy and disability in multiple sclerosis over four years: application of a reproducible automated technique in monitoring disease progression in a cohort of the interferon beta-1a (Rebif) treatment trial. *J Neurol Neurosurg Psychiatry* **74**, 1090-1094 (2003).

108 De Stefano, N. *et al.* Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. *Neurology* **74**, 1868-1876, doi:10.1212/WNL.0b013e3181e24136 (2010).

109 Lukas, C. *et al.* Cervical spinal cord volume loss is related to clinical disability progression in multiple sclerosis. *J Neurol Neurosurg Psychiatry* **86**, 410-418, doi:10.1136/jnnp-2014-308021 (2015).

110 Cawley, N., Tur, C. ; Prados, F.; Plantone, D.; Kearney, H.; Abdel – Aziz, K.; Ourselin, S.; Gandini Wheeler – Kingshott, C.A.M.; Miller, D.H.; Thompson, A.J.; Ciccarelli, O. Spinal cord atrophy as a primary outcome measure in phase II trials of progressive multiple sclerosis. *Mult Scler Journal (under review)* (2016).

111 Gold, R. *et al.* Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. *N Engl J Med* **367**, 1098-1107, doi:10.1056/NEJMoa1114287 (2012).

112 Schmierer, K., Scaravilli, F., Altmann, D. R., Barker, G. J. & Miller, D. H. Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Ann Neurol* **56**, 407-415, doi:10.1002/ana.20202 (2004).

113 Arnold, D. L. *et al.* Magnetization transfer ratio in the delayed-release dimethyl fumarate DEFINE study. *J Neurol* **261**, 2429-2437, doi:10.1007/s00415-014-7504-7 (2014).

114 Narayanan, S. *et al.* Axonal metabolic recovery in multiple sclerosis patients treated with interferon beta-1b. *J Neurol* **248**, 979-986 (2001).

115 Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* **61**, 1000-1016, doi:10.1016/j.neuroimage.2012.03.072 (2012).

116 Paling, D. *et al.* Sodium accumulation is associated with disability and a progressive course in multiple sclerosis. *Brain* **136**, 2305-2317, doi:10.1093/brain/awt149 (2013).

117 Sergott, R. C. *et al.* ATON: results from a Phase II randomized trial of the B-cell-targeting agent atacicept in patients with optic neuritis. *J Neurol Sci* **351**, 174-178, doi:10.1016/j.jns.2015.02.019 (2015).

118 Balcer, L. J. Clinical trials to clinical use: using vision as a model for multiple sclerosis and beyond. *J Neuroophthalmol* **34 Suppl**, S18-23, doi:10.1097/WNO.0000000000000163 (2014).

119 Saidha, S. *et al.* Optical coherence tomography reflects brain atrophy in multiple sclerosis: A four-year study. *Ann Neurol* **78**, 801-813, doi:10.1002/ana.24487 (2015).

120 Raftopoulos, R. *et al.* Phenytoin for neuroprotection in patients with acute optic neuritis: a randomised, placebo-controlled, phase 2 trial. *Lancet Neurol* **15**, 259-269, doi:10.1016/S1474-4422(16)00004-1 (2016).

121 Cambron, M. *et al.* Fluoxetine in progressive multiple sclerosis (FLUOX-PMS): study protocol for a randomized controlled trial. *Trials* **15**, 37, doi:10.1186/1745-6215-15-37 (2014).

122 Rice, C. M. *et al.* Assessment of bone marrow-derived Cellular Therapy in progressive Multiple Sclerosis (ACTiMuS): study protocol for a randomised controlled trial. *Trials* **16**, 463, doi:10.1186/s13063-015-0953-1 (2015).

123 Polman, C. H. *et al.* A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med* **354**, 899-910, doi:10.1056/NEJMoa044397 (2006).

124 Burman, J. *et al.* Autologous haematopoietic stem cell transplantation for aggressive multiple sclerosis: the Swedish experience. *J Neurol Neurosurg Psychiatry* **85**, 1116-1121, doi:10.1136/jnnp-2013-307207 (2014).

125 Atkins, H. L. *et al.* Immunoablation and autologous haemopoietic stem-cell transplantation for aggressive multiple sclerosis: a multicentre single-group phase 2 trial. *Lancet* **388**, 576-585, doi:10.1016/S0140-6736(16)30169-6 (2016).

126 Sormani, M. P., Muraro, P. A., Saccardi, R. & Mancardi, G. NEDA status in highly active MS can be more easily obtained with autologous hematopoietic stem cell transplantation than other drugs. *Mult Scler*, doi:10.1177/1352458516645670 (2016).

127 Kappos, L. *et al.* Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis. *Mult Scler* **22**, 1297-1305, doi:10.1177/1352458515616701 (2016).

128 Nash, R. A. *et al.* High-dose immunosuppressive therapy and autologous hematopoietic cell transplantation for relapsing-remitting multiple sclerosis (HALT-MS): a 3-year interim report. *JAMA Neurol* **72**, 159-169, doi:10.1001/jamaneurol.2014.3780 (2015).

129 Assessment., E. N. f. H. T. *Endpoints used for REA of pharmaceuticals: composite endpoints*, <http://www.eunethta.eu/outputs/methodological-guideline-rea-pharmaceuticals-composite-endpoints> (2013).

130 Kalincik, T. *et al.* Treatment effectiveness of alemtuzumab compared with natalizumab, fingolimod, and interferon beta in relapsing-remitting multiple sclerosis: a cohort study. *Lancet Neurol* **16**, 271-281, doi:10.1016/S1474-4422(17)30007-8 (2017).

131 Lorscheider, J. *et al.* Anti-inflammatory disease-modifying treatment and short-term disability progression in SPMS. *Neurology*, doi:10.1212/WNL.0000000000004330 (2017).

132 Sormani, M. P. *et al.* Combined MRI lesions and relapses as a surrogate for disability in multiple sclerosis. *Neurology* **77**, 1684-1690, doi:10.1212/WNL.0b013e31823648b9 (2011).

133 Sormani, M. P. *et al.* Assessing response to interferon-beta in a multicenter dataset of patients with MS. *Neurology* **87**, 134-140, doi:10.1212/WNL.0000000000002830 (2016).

134 Rudick, R. *et al.* Clinical outcomes assessment in multiple sclerosis. *Ann Neurol* **40**, 469-479, doi:10.1002/ana.410400321 (1996).
135 Gronwall, D. M. Paced auditory serial-addition task: a measure of recovery from concussion. *Percept Mot Skills* **44**, 367-373, doi:10.2466/pms.1977.44.2.367 (1977).
136 Fragoso, Y. D. *et al.* Real-life experience with fampridine (Fampyra(R)) for patients with multiple sclerosis and gait disorders. *NeuroRehabilitation* **39**, 301-304, doi:10.3233/NRE-161361 (2016).
137 Barker-Collo, S. L. Within session practice effects on the PASAT in clients with multiple sclerosis. *Arch Clin Neuropsychol* **20**, 145-152, doi:10.1016/j.acn.2004.03.007 (2005).
138 Solari, A., Radice, D., Manneschi, L., Motti, L. & Montanari, E. The multiple sclerosis functional composite: different practice effects in the three test components. *J Neurol Sci* **228**, 71-74, doi:10.1016/j.jns.2004.09.033 (2005).
139 Brooks, J. B. *et al.* Paced auditory serial addition test (PASAT): a very difficult test even for individuals with high intellectual capability. *Arq Neuropsiquiatr* **69**, 482-484 (2011).
140 Langdon, D. W. *et al.* Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS). *Mult Scler* **18**, 891-898, doi:10.1177/1352458511431076 (2012).
141 Smith, A. *Symbol digit modalities test.* . (Western Psychological Services, 1991).
142 Amato, M. P. *et al.* The Rao's Brief Repeatable Battery and Stroop Test: normative values with age, education and gender corrections in an Italian population. *Mult Scler* **12**, 787-793 (2006).
143 Tur, C. *et al.* Grey matter damage and overall cognitive impairment in primary progressive multiple sclerosis. *Mult Scler* **17**, 1324-1332, doi:10.1177/1352458511410341 (2011).
144 Sonder, J. M., Burggraaff, J., Knol, D. L., Polman, C. H. & Uitdehaag, B. M. Comparing long-term results of PASAT and SDMT scores in relation to neuropsychological testing in multiple sclerosis. *Mult Scler* **20**, 481-488, doi:10.1177/1352458513501570 (2014).
145 Jacques, F. H. *et al.* Cognitive evolution in natalizumab-treated multiple sclerosis patients. *Mult Scler J Exp Transl Clin* **2**, 2055217316657116, doi:10.1177/2055217316657116 (2016).
146 Fisk, J. D., Pontefract, A., Ritvo, P. G., Archibald, C. J. & Murray, T. J. The impact of fatigue on patients with multiple sclerosis. *Can J Neurol Sci* **21**, 9-14 (1994).
147 Kos, D., Nagels, G., D'Hooghe, M. B., Duportail, M. & Kerckhofs, E. A rapid screening tool for fatigue impact in multiple sclerosis. *BMC Neurol* **6**, 27, doi:10.1186/1471-2377-6-27 (2006).
148 Wattjes, M. P. *et al.* Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis--establishing disease prognosis and monitoring patients. *Nat Rev Neurol* **11**, 597-606, doi:10.1038/nrneurol.2015.157 (2015).
149 Traboulsee, A. *et al.* Revised Recommendations of the Consortium of MS Centers Task Force for a Standardized MRI Protocol and Clinical Guidelines for

the Diagnosis and Follow-Up of Multiple Sclerosis. *AJNR Am J Neuroradiol* **37**, 394-401, doi:10.3174/ajnr.A4539 (2016).

150 Scolding, N. *et al.* Association of British Neurologists: revised (2015) guidelines for prescribing disease-modifying treatments in multiple sclerosis. *Pract Neurol* **15**, 273-279, doi:10.1136/practneurol-2015-001139 (2015).

151 Rovira, A. *et al.* Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis-clinical implementation in the diagnostic process. *Nat Rev Neurol* **11**, 471-482, doi:10.1038/nrneurol.2015.106 (2015).

152 Pareto, D. *et al.* Lesion filling effect in regional brain volume estimations: a study in multiple sclerosis patients with low lesion load. *Neuroradiology* **58**, 467-474, doi:10.1007/s00234-016-1654-5 (2016).

153 Kanda, T. *et al.* Gadolinium deposition in the brain. *Magn Reson Imaging*, doi:10.1016/j.mri.2016.08.024 (2016).

154 Gulani, V. *et al.* Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol* **16**, 564-570, doi:10.1016/S1474-4422(17)30158-8 (2017).

155 McGuigan, C. *et al.* Stratification and monitoring of natalizumab-associated progressive multifocal leukoencephalopathy risk: recommendations from an expert group. *J Neurol Neurosurg Psychiatry* **87**, 117-125, doi:10.1136/jnnp-2015-311100 (2016).

156 Chard, D. Brain atrophy measurements should be used to guide therapy monitoring in MS - Commentary. *Mult Scler* **22**, 1526-1527, doi:10.1177/1352458516656061 (2016).

157 Barkhof, F. Brain atrophy measurements should be used to guide therapy monitoring in MS - NO. *Mult Scler* **22**, 1524-1526, doi:10.1177/1352458516649452 (2016).

158 Zivadinov, R., Dwyer, M. G. & Bergsland, N. Brain atrophy measurements should be used to guide therapy monitoring in MS - YES. *Mult Scler* **22**, 1522-1524, doi:10.1177/1352458516649253 (2016).

159 Smeets, D. *et al.* Reliable measurements of brain atrophy in individual patients with multiple sclerosis. *Brain Behav* **6**, e00518, doi:10.1002/brb3.518 (2016).

160 Dwyer, M. G. *et al.* Neurological software tool for reliable atrophy measurement (NeuroSTREAM) of the lateral ventricles on clinical-quality T2-FLAIR MRI scans in multiple sclerosis. *Neuroimage Clin* **15**, 769-779, doi:10.1016/j.nicl.2017.06.022 (2017).

161 Vrenken, H. *et al.* Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis. *J Neurol* **260**, 2458-2471, doi:10.1007/s00415-012-6762-5 (2013).

162 University of California, S. F. M. S. E. T. *et al.* Long-term evolution of multiple sclerosis disability in the treatment era. *Ann Neurol* **80**, 499-510, doi:10.1002/ana.24747 (2016).

163 Rothwell, P. M., McDowell, Z., Wong, C. K. & Dorman, P. J. Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. *BMJ* **314**, 1580-1583 (1997).

164 Jacobs, L. D. *et al.* Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis. CHAMPS Study Group. *N Engl J Med* **343**, 898-904, doi:10.1056/NEJM200009283431301 (2000).

165 Leist, T. P. *et al.* Effect of oral cladribine on time to conversion to clinically definite multiple sclerosis in patients with a first demyelinating event (ORACLE MS): a phase 3 randomised trial. *Lancet Neurol* **13**, 257-267, doi:10.1016/S1474-4422(14)70005-5 (2014).

166 Giovannoni, G. *et al.* A placebo-controlled trial of oral cladribine for relapsing multiple sclerosis. *N Engl J Med* **362**, 416-426, doi:10.1056/NEJMoa0902533 (2010).

167 Miller, D. H. *et al.* Effects of delayed-release dimethyl fumarate on MRI measures in the phase 3 CONFIRM study. *Neurology* **84**, 1145-1152, doi:10.1212/WNL.0000000000001360 (2015).

168 Massacesi, L. *et al.* Azathioprine versus beta interferons for relapsing-remitting multiple sclerosis: a multicentre randomized non-inferiority trial. *PLoS One* **9**, e113371, doi:10.1371/journal.pone.0113371 (2014).

169 Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. European Study Group on interferon beta-1b in secondary progressive MS. *Lancet* **352**, 1491-1497 (1998).

170 Panitch, H. *et al.* Randomized, comparative study of interferon beta-1a treatment regimens in MS: The EVIDENCE Trial. *Neurology* **59**, 1496-1506 (2002).

171 Khan, O. *et al.* Three times weekly glatiramer acetate in relapsing-remitting multiple sclerosis. *Ann Neurol* **73**, 705-713, doi:10.1002/ana.23938 (2013).

172 Mikol, D. D. *et al.* Comparison of subcutaneous interferon beta-1a with glatiramer acetate in patients with relapsing multiple sclerosis (the REbif vs Glatiramer Acetate in Relapsing MS Disease [REGARD] study): a multicentre, randomised, parallel, open-label trial. *Lancet Neurol* **7**, 903-914, doi:10.1016/S1474-4422(08)70200-X (2008).

173 Sorensen, P. S. *et al.* Simvastatin as add-on therapy to interferon beta-1a for relapsing-remitting multiple sclerosis (SIMCOMBIN study): a placebo-controlled randomised phase 4 trial. *Lancet Neurol* **10**, 691-701, doi:10.1016/S1474-4422(11)70144-2 (2011).

174 Kappos, L. *et al.* Daclizumab HYP versus Interferon Beta-1a in Relapsing Multiple Sclerosis. *N Engl J Med* **373**, 1418-1428, doi:10.1056/NEJMoa1501481 (2015).

175 Comi, G., Filippi, M. & Wolinsky, J. S. European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer acetate on magnetic resonance imaging--measured disease activity and burden in patients with relapsing multiple sclerosis. European/Canadian Glatiramer Acetate Study Group. *Ann Neurol* **49**, 290-297 (2001).

176 Comi, G. *et al.* Phase III dose-comparison study of glatiramer acetate for multiple sclerosis. *Ann Neurol* **69**, 75-82, doi:10.1002/ana.22316 (2011).

177 Noseworthy, J. H. *et al.* Linomide in relapsing and secondary progressive MS: part I: trial design and clinical results. North American Linomide Investigators. *Neurology* **54**, 1726-1733 (2000).

178 Jacobs, L. D. *et al.* Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). *Ann Neurol* **39**, 285-294, doi:10.1002/ana.410390304 (1996).

179 Rudick, R. A. *et al.* Natalizumab plus interferon beta-1a for relapsing multiple sclerosis. *N Engl J Med* **354**, 911-923, doi:10.1056/NEJMoa044396 (2006).
180 Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. The IFNB Multiple Sclerosis Study Group. *Neurology* **43**, 655-661 (1993).
181 Confavreux, C. *et al.* Oral teriflunomide for patients with relapsing multiple sclerosis (TOWER): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Neurol* **13**, 247-256, doi:10.1016/S1474-4422(13)70308-9 (2014).
182 Wolinsky, J. S. *et al.* Glatiramer acetate in primary progressive multiple sclerosis: results of a multinational, multicenter, double-blind, placebo-controlled trial. *Ann Neurol* **61**, 14-24, doi:10.1002/ana.21079 (2007).
183 Wolinsky, J. S. *et al.* Linomide in relapsing and secondary progressive MS: part II: MRI results. MRI Analysis Center of the University of Texas-Houston, Health Science Center, and the North American Linomide Investigators. *Neurology* **54**, 1734-1741 (2000).
184 Balto, J. M., Kinnett-Hopkins, D. L. & Motl, R. W. Accuracy and precision of smartphone applications and commercially available motion sensors in multiple sclerosis. *Multiple Sclerosis Journal - Experimental, Translational and Clinical*, doi:DOI: 10.1177/ 2055217316634754 (2016).
185 Schwartzbach, C. J. *et al.* Lesion remyelinating activity of GSK239512 versus placebo in patients with relapsing-remitting multiple sclerosis: a randomised, single-blind, phase II study. *J Neurol* **264**, 304-315, doi:10.1007/s00415-016-8341-7 (2017).
186 Bodini, B. *et al.* Dynamic imaging of individual remyelination profiles in multiple sclerosis. *Ann Neurol*, doi:10.1002/ana.24620 (2016).
187 Sethi, V. *et al.* Slowly eroding lesions in multiple sclerosis. *Mult Scler* **23**, 464-472, doi:10.1177/1352458516655403 (2017).
188 Ching, A. S. *et al.* Current paradigm of the 18-kDa translocator protein (TSPO) as a molecular target for PET imaging in neuroinflammation and neurodegenerative diseases. *Insights Imaging* **3**, 111-119, doi:10.1007/s13244-011-0128-x (2012).
189 Poutiainen, P., Jaronen, M., Quintana, F. J. & Brownell, A. L. Precision Medicine in Multiple Sclerosis: Future of PET Imaging of Inflammation and Reactive Astrocytes. *Front Mol Neurosci* **9**, 85, doi:10.3389/fnmol.2016.00085 (2016).
190 Herranz, E. *et al.* Neuroinflammatory component of gray matter pathology in multiple sclerosis. *Ann Neurol* **80**, 776-790, doi:10.1002/ana.24791 (2016).
191 Horton, L. *et al.* Effect of 4-aminopyridine on vision in multiple sclerosis patients with optic neuropathy. *Neurology* **80**, 1862-1866, doi:10.1212/WNL.0b013e3182929fd5 (2013).
192 Zarbin, M. A. *et al.* Ophthalmic evaluations in clinical studies of fingolimod (FTY720) in multiple sclerosis. *Ophthalmology* **120**, 1432-1439, doi:10.1016/j.ophtha.2012.12.040 (2013).
193 Fuhr, P., Borggrefe-Chappuis, A., Schindler, C. & Kappos, L. Visual and motor evoked potentials in the course of multiple sclerosis. *Brain* **124**, 2162-2168 (2001).

194 Jung, P., Beyerle, A. & Ziemann, U. Multimodal evoked potentials measure and predict disability progression in early relapsing-remitting multiple sclerosis. *Mult Scler* **14**, 553-556, doi:10.1177/1352458507085758 (2008).

195 Pelayo, R. *et al.* Do multimodal evoked potentials add information to MRI in clinically isolated syndromes? *Mult Scler* **16**, 55-61, doi:10.1177/1352458509352666 (2010).

196 Cadavid, D., Balcer, L., Galetta, S., Aktas, O., Ziemssen, T., Vanopdenbosch, L., Butzkueven, H., Ziemssen, F., Massacesi, L., Chai, Y., Xu, L., Freeman, S. Efficacy Analysis of the Anti-LINGO-1 Monoclonal Antibody BIIB033 in Acute Optic Neuritis: the RENEW Trial *Neurology* **84** (2015).

Supplementary Tables

Supplementary Table 1: Clinical outcome measures in phase III trials in relapsing-remitting (RR) MS

Original clinical outcome	Derived outcome measures	Trial	Condition (no. of patients randomised)	Drug, effect (vs. placebo/ another active arm)	Duration of the trial
Relapses	**Mean annualised relapse rate (a)**	The IFNB Multiple Sclerosis Study Group, Neurology 1993, phase III	RRMS (n=372)	IFN beta-1b 1.6 MIU: 1.17, p (vs. placebo) = 0.0101; IFN beta-1b 8 MIU: 0.84, p (vs. placebo) = 0.0001; p (vs. 1.6 MIU) =0.0086; Placebo: 1.27	24 months
		Johnson et al., Neurology 1995, phase III (The Copolymer 1 Multiple Sclerosis Study)	RRMS (n=251)	Glatiramer acetate 20mg SC/day: 0.59; Placebo: 0.84, p=0.007	24 months
		Jacobs et al., Ann Neurol 1996, phase III (MSCRG study)	Relapsing MS (n=301)	IFN beta-1a 30mcg IM/week: 0.61; Placebo: 0.9, p=0.03	104 weeks
		PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)	RRMS (n=560)	Placebo: 1.28; IFN beta-1a 22mcg SC tiw: 0.91, p<0.005 (vs. placebo); IFN beta-1a 44mcg SC tiw: 0.865, p<0.005 (vs. placebo); **(s)**	24 months
		Noseworthy et al., Neurology 2000, phase III (linomide study)	RMS (n=715)	The study was of insufficient duration for any of the primary or secondary outcome measures to reach significance	Early termination for safety issues (initially planned: 36 months)
		Comi et al., Ann Neurol 2001, phase	RRMS (n=249)	Glatiramer Acetate vs.	9 months

		III (European/Canadian Glatiramer Acetate Study)		Glatiramer acetate 20mg SC/day: 0.81; Placebo: 1.21, p=0.012	
		Polman et al., NEJM 2006, phase III (AFFIRM study)	RRMS (n=627)	Natalizumab 300mg/4 weeks: 0.23 (0.19 to 0.28); Placebo: 0.73 (0.62 to 0.87), p<0.001	24 months
		Panitch et al., Neurology 2002; Schwid et al., Clinical Therapeutics 2007, phase 4 – post-commercialisation (EVIDENCE study)	RRMS (n=677)	IFN beta-1a IM 30mcg/week: 0.65; IFN beta-1a SC 44mcg tiw: 0.54, p=0.033	24 months (0-12m: comparative phase; 12-24m: cross-over phase) **(n)**
		Rudick et al., NEJM 2006, phase III (SENTINEL study)	RRMS (n=1171)	Natalizumab 300mg/4 weeks + IFN beta-1a IM 30mcg/week: 0.34 (0.29 to 0.39); IFN beta-1a IM 30mcg/week: 0.75 (0.67 to 0.84), p=0.001	24 months
		O'Connor et al., Lancet Neurol 2009, phase III (BEYOND study)	RRMS (n=2244)	IFN beta-1b 500mcg SC eod: 0.33; IFN beta-1b 250mcg SC eod: 0.36; GA 20mg SC/day: 0.34, p-values (all comparisons) > 0.05	24 months
		Cohen et al., NEJM 2010, phase III (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 0.5mg/day: 0.16 (0.12 to 0.21), p (vs. IFN) <0.001; Fingolimod 1.25mg/day: 0.20 (0.16 to 0.26), p (vs. IFN) <0.001; IFN beta-1a IM 30mcg/week: 0.33 (95% CI 0.26 to 0.42);	12 months
		Kappos et al., NEJM	RRMS	Fingolimod	24 months

		2010, phase III (FREEDOMS study)	(n=1272)	0.5mg/day: 0.18 (0.15 to 0.22), p (vs. placebo) <0.001; Fingolimod 1.25mg/day: 0.16 (0.13 to 0.19) , p (vs. placebo) <0.001; Placebo: 0.40 (95% CI 0.34 to 0.47);	
		Giovannoni et al., NEJM 2010, phase III (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/Kg: 0.14 (0.12 to 0.17), p (vs. placebo) <0.001; Cladribine 5.25mg/Kg: 0.15 (0.12 to 0.17), p (vs. placebo) <0.001; Placebo: 0.33 (95% CI 0.29 to 0.38);	96 weeks
		Comi at el., Ann Neurol 2011, phase III (FORTE study)	RRMS (n=1155)	GA 20mg SC/day: 0.33 (SD 0.81); GA 40mg SC/day: 0.35 (SD 0.99), p=0.486	12 months
		O'Connor et al., NEJM 2011, phase III (TEMSO study)	Relapsing MS (n=1088)	Teriflunomide 7mg PO/day: 0.37 (0.32–0.43), p (vs. placebo) <0.001; Teriflunomide 14mg PO/day: 0.37 (0.31–0.44), p (vs. placebo) <0.001; Placebo: 0.54 (0.47–0.62)	108 weeks
		Sorensen et al., Lancet Neurology 2011, phase 4 (SIMCOMBIN study)	RRMS (n=307)	IFN beta-1a 30mcg IM/week + simvastatin 80mg/day: 0.188 (95% CI 0.126 to 0.281); IFN beta-1a	12 months after last patient was included

				30mcg IM/week + Placebo: 0.144 (95% CI 0.092 to 0.227), p = 0.35	
		Cohen et al., Lancet 2012, phase III (CARE-MS I study)	RRMS previously untreated (n=581)	Alemtuzumab 12mg IV/day x 5 days: 0.18 (0.13 to 0.23); IFN beta-1a 44mcg SC tiw: 0.39 (95% CI: 0.29 to 0.53), p<0.0001	24 months
		Coles et al., Lancet 2012, phase III (CARE-MS II study)	RRMS previously treated (n=840)	Alemtuzumab 12mg IV/day x 5 days: 0.26 (95% CI 0.21 to 0.33); IFN beta 1a 44mcg SC tiw: 0.52 (95% CI 0.41 to 0.66), p<0.0001	24 months
		Comi et al., NEJM 2012, phase III (ALLEGRO study)	RRMS (n=1106)	Laquinimod 0.6mg OD: 0.30 (SE 0.02), p (vs. placebo) =0.002; Placebo: 0.39 (SE 0.03);	24 months
		Fox et al., NEJM 2012, phase III (CONFIRM study)	RRMS (n=1417)	BG-12 240mg BD: 0.22 (95% CI 0.18 to 0.28), p (vs. placebo) <0.001; BG-12 240mg TDS: 0.20 (95% CI 0.16 to 0.25), p (vs. placebo) <0.001; GA 40mg SC/day: 0.29 (95% CI 0.23 to 0.35), p (vs. placebo) <0.05; Placebo: 0.40 (95% CI 0.33 to 0.49);	24 months
		Gold et al., NEJM 2012, phase III (DEFINE study)	RRMS (n=1234)	BG-12 240mg BD: 0.17 (95% CI 0.14 to 0.21), p (vs. placebo) <0.001; BG-12 240mg TDS: 0.19 (95% CI 0.15 to 0.23), p	24 months

				(vs. placebo) <0.001; Placebo: 0.36 (95% CI 0.30 to 0.44);	
		Khan et al., Ann Neurol 2013, phase III (GALA study)	RRMS (n=1404)	GA 40mg SC tiw: 0.331 (95% CI 0.280 to 0.392) vs. placebo: 0.505 (0.418 to 0.609), p<0.0001	12 months
		Lublin et al., Ann Neurol 2013, phase III (CombiRx study)	RRMS (n=1008)	IFN beta-1a 30mcg SC/week + GA 20mg SC/day: 0.23 vs. IFN: 0.32, p=0.001; IFN+GA: 0.23 vs. GA: 0.23, p=0.44; IFN vs. GA: p=0.008	36 months after last patient was included
		Calabresi et al. Lancet Neurol 2014, phase III (ADVANCE study)	RRMS (n=1516)	Peginterferon beta-1a 125mcg SC/2 weeks vs. placebo: 0.256 (0.206–0.318) vs. 0.397 (0.328–0.481), p=0.0007; Peginterferon beta-1a 125mcg SC/4 weeks vs. placebo: 0.288 (0·234–0.355) vs. 0.397 (0.328–0.481), p=0.0114	24 months (but primary endpoint: 48 weeks, which is the placebo-controlled phase)
		Calabresi et al., Lancet Neurol 2014, phase III (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 0.5mg: 0·21 (0·17–0·25); placebo: 0.40 (95% CI 0.34–0.48), p<0·0001	24 months
		Confavreux et al., Lancet Neurol 2014, phase III (TOWER study)	RRMS (n=1169)	Teriflunomide 7mg: 0·39 (0·33–0·46); p (vs. placebo) =0·0183 Teriflunomide 14mg: 0·32 (0·27–0·38); p (vs. placebo)	48 weeks after the last patient was included (MRI results not published)

				=0·0001 Placebo: 0·50 (95% CI 0·43–0·58)	
		Massacesi et al., PLoS ONE 2014, phase III	RRMS (n=150)	Azathioprine (target dose: 3 mg/kg/d) vs. BIFN beta (1a or 1b): 0.26 (95% CI: 0.19–0.37) vs. 0.39 (95% CI: 0.30–0.51), p=0.07	24 months
		Mikol et al., Lancet Neurol 2014, phase III (REGARD study)	RRMS (n=764)	IFN beta-1a 44mcg SC tiw: 0.30, vs. Glatiramer acetate 20mg SC/day: 0.29; p = 0.828	96 weeks
		Vermersch et al., MSJ 2014, phase III (TENERE study)	Relapsing MS (n=324)	Teriflunomide 7mg: 0.41 (0.27 to 0.64), p (vs. IFN) =0.03; Teriflunomide 14mg: 0.26 (0.15 to 0.44), p (vs. IFN) =0.59; IFN beta-1a: 0.22 (0.11 to 0.42);	48 weeks after the last patient was included
		Vollmer et al., J Neurol 2014, phase III (BRAVO study)	RRMS (n=1331)	Laquinimod 0.6mg: 0.28 (0.03); IFN-beta 30 mcg IM: 0.26 (0.02); Placebo: 0.34 (0.03); p (Laq vs. placebo)=0.075; p (IFN vs. placebo)=0.007	24 months
		Cohen et al., JAMA Neurol 2015, phase III (GATE study)	RRMS (n=796)	Generic GA 20mg/d vs. brand GA 20mg/d vs. placebo: 0.31 (0.20 to 0.48) vs. 0.40 (0.26 to 0.62) vs. 0.38 (0.22 to 0.66) (ns)	9 months
		Kappos et al., New Engl J Med 2015, phase III (DECIDE	RRMS (n=1841)	Daclizumab HYP 150mg SC/4 weeks vs. IFN	144 weeks

		study)		beta-1a 30µg IM/week: 0.22 vs. 0.39 (p<0.001)	
		Lanzillo et al., Mult Scler Journal 2015, phase III (ARIANNA study)	RRMS (n=154)	Beta-IFN 1b eod SC + atorvastatin 40mg PO/day vs. Beta-IFN 1b eod SC + placebo: 0.39 vs. 0.32, (p>0.05)	24 months
	Mean annualised severe relapse rate (i)	The IFNB Multiple Sclerosis Study Group, Neurology 1993, phase III	RRMS (n=372)	There was a twofold reduction in the frequency of moderate and severe attacks in the IFN beta-1b 8 MIU (probably vs. placebo – not specified in abstract); p-value not specified.	24 months
		PRISMS study group (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)	RRMS (n=560)	IFN beta-1a 22mcg SC tiw: 0.355, p<0.005 (vs. placebo); IFN beta-1a 44mcg SC tiw: 0.31, p<0.005 (vs. placebo); Placebo: 0.495; **(s)**	24 months
		O'Connor et al., Lancet Neurol 2009, phase III (BEYOND study)	RRMS (n=2244)	GA 20mg SC/day at 2 years FU: 0.18; IFN beta-1b 250mcg SC EmTheOD at 2 years FU: 0.19; IFN beta-1b 500mcg SC EOD at 2 years FU: 0.18, p values (all comparisons) > 0.05	24 months
		Comi et al., NEJM 2012, phase III (ALLEGRO study)	RRMS (n=1106)	Placebo: 0.33 (SE 0.02); Laquinimod 0.6mg OD: 0.24 (SE 0.02), p (vs.	24 months

				placebo) <0.001;	
		Khan et al., Ann Neurol 2013, phase III (GALA study)	RRMS (n=1404)	GA 40mg sc tiw: 0.301 (95% CI 0.252 to 0.359) vs. placebo: 0.466 (0.383 to 0.568), p<0.0001	12 months
	% patients with at least 1 relapse (a) (l)	O'Connor et al., Lancet Neurol 2009, phase III (BEYOND study) **(m)**	RRMS (n=2244)	GA 20mg SC/day at 2 years FU: 27%; IFN beta-1b 250mcg SC eod at 2 years FU: 27%; IFN beta-1b 500mcg SC eod at 2 years FU: 26%, p-values (all comparisons) > 0.05	24 months
		Fox et al., NEJM 2012, phase III (CONFIRM study)	RRMS (n=1417)	Placebo: 41%; BG-12 240mg BD: 29%, p (vs. placebo) ≤0.01; BG-12 240mg TDS: 24%, p (vs. placebo) <0.001; GA 40mg SC/day: 32%, p (vs. placebo) ≤0.01;	24 months
		Gold et al., NEJM 2012, phase III (DEFINE study)	RRMS (n=1234)	Placebo: 46%; BG-12 240mg BD: 27%, p (vs. placebo) <0.001; BG-12 240mg TDS: 26%, p (vs. placebo) <0.001;	24 months
		Lublin et al., Ann Neurol 2013, phase III (CombiRx study)	RRMS (n=1008)	IFN+GA: 38.9% vs. IFN: 44.4%, p=0.19 IFN+GA: 38.9% vs. GA: 35.9%, p=0.21 IFN: 44.4% vs. GA: 35.9%, p=0.14	36 months after last patient was included
		Calabresi et al. Lancet Neurol 2014, phase III (ADVANCE study)	RRMS (n=1516)	Peginterferon beta-1a 125µg/2 weeks SC vs. placebo:	24 months (but primary endpoint:

				0.187 (0.0178) vs. 0.291 (0.0206), p=0.0003; Peginterferon beta-1a 125µg/4 weeks SC vs. placebo: 0.222 (0.0191) vs. 0.291 (0.0206), p=0.02	48 weeks, which is the placebo-controlled phase)
		Massacesi et al., PLoS ONE 2014, phase III	RRMS (n=150)	Azathioprine (target dose: 3 mg/kg/d) vs. beta-IFN (1a or 1b): 35.5% vs. 47.8%, p=0.22 (ns)	24 months
		Vermersch et al., MSJ 2014, phase III (TENERE study)	Relapsing MS (n=324)	IFNβ-1a: 15.4%, p (vs Teriflunomide 7mg) = 0.03, p (vs Teriflunomide 14mg) = 0.6; Teriflunomide 7mg: 42.2%; Teriflunomide 14mg: 23.4%;	48 weeks after the last patient was included
	% relapse-free patients at the end of FU	The IFNB Multiple Sclerosis Study Group, Neurology 1993, phase III	RRMS (n=372)	IFN beta-1b 8 MIU: 29%; Placebo: 14.5%, p=0.007	24 months
		Johnson et al., Neurology 1995, phase III (The Copolymer 1 Multiple Sclerosis Study)	RRMS (n=251)	Glatiramer acetate 20mg SC/day: 33.6%; Placebo: 27.0%, p=0.098	24 months
		PRISMS study group (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)	RRMS (n=560)	IFN beta-1a 22mcg SC tiw: 27%, p≤0.05 (vs. placebo); IFN beta-1a 44mcg SC tiw: 32%, p<0.005 (vs. placebo); Placebo: 16%	24 months
		Panitch et al., Neurology 2002; Schwid et al., Clinical Therapeutics 2007, phase 4 – post-	RRMS (n=677)	IFN beta-1a SC 44mcg tiw: 56%; IFN beta-1a IM 30mcg/week: 48%, p=0.023	24 months (0-12m: comparative phase; 12-24m: cross-over phase)

		commercialisation (EVIDENCE study)			**(n)**
		Polman et al., NEJM 2006, phase III (AFFIRM study)	RRMS (n=627)	Natalizumab 300mg/4 weeks: 72%, p (vs. placebo) <0.05; Placebo: 46% **(o)**	24 months
		Rudick et al., NEJM 2006, phase III (SENTINEL study)	RRMS (n=1171)	Natalizumab 300mg/4 weeks + IFN beta-1a IM 30mcg/week: 61%, p (vs. IFN) <0.05; IFN beta-1a IM 30mcg/week: 37% **(o)**	24 months
		Cohen et al., NEJM 2010, phase III (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 0.5mg/day: 82.6% (79.0 to 86.3), p (vs. IFN) <0.001; Fingolimod 1.25mg/day: 79.8% (75.9 to 83.7), p (vs. IFN) <0.001; IFN beta-1a IM 30mcg/week: 69.3% (95% CI 64.8 to 73.8)	12 months
		O'Connor et al., Lancet Neurol 2009, phase III (BEYOND study)	RRMS (n=2244)	IFN beta-1b 500mcg SC eod at 2 years FU: 60%; IFN beta-1b 250mcg SC eod at 2 years FU: 58%; GA 20mg SC/day at 2 years FU: 59%, p values (all comparisons) > 0.05	24 months
		Kappos et al., NEJM 2010, phase III (FREEDOMS study)	RRMS (n=1272)	Fingolimod 0.5mg/day: 70.4% (66.0 to 74.8), p (vs. placebo) <0.001 Fingolimod 1.25mg/day: 74.7% (70.4 to	24 months

				78.9), p (vs. placebo) <0.001 Placebo: 45.6% (95% CI 40.7 to 50.6)	
		Giovannoni et al., NEJM 2010, phase III (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/Kg: 79.7%, p (vs. placebo) <0.001; Cladribine 5.25mg/Kg: 78.9%, p (vs. placebo) <0.001; Placebo: 60.9%	96 weeks
		Comi at el., Ann Neurol 2011, phase III (FORTE study)	RRMS (n=1155)	GA 20mg SC/day: 77.6% (SD 17.4); GA 40mg SC/day: 77.0% (SD 17.7), p=0.999	12 months
		O'Connor et al., NEJM 2011, phase III (TEMSO study)	Relapsing MS (n=1088)	Teriflunomide 7mg PO/day: 53.7% (48.3–59.1), p (vs. placebo) =0.01; Teriflunomide 14mg PO/day: 56.5% (51.0–62.0) , p (vs. placebo) =0.003; Placebo: 45.6% (95% CI: 40.2–51.0)	108 weeks
		Sorensen et al., Lancet Neurology 2011, phase 4 (SIMCOMBIN study)	RRMS (n=307)	IFN beta-1a 30mcg IM/week + simvastatin 80mg/day: 75%; IFN beta-1a 30mcg IM/week + Placebo: 81%, p = 0.512	12 months after last patient was included
		Cohen et al., Lancet 2012, phase III (CARE-MS I study)	RRMS previously untreated (n=581)	Alemtuzumab 12mg IV/day x 5 days: 77.6% (72.9 to 81.6); IFN beta 1a 44mcg SC tiw: 58.7% (95% CI: 51.1 to 65.5), p<0.0001	24 months

		Coles et al., Lancet 2012, phase III (CARE-MS II study)	RRMS previously treated (n=840)	Alemtuzumab 12mg IV/day x 5 days: 65.4% (95% CI 60.7 to 69.7); IFN beta 1a 44mcg SC tiw: 46.7% (95% CI 39.5 to 53.5), p<0.0001;	24 months
		Comi et al., NEJM 2012, phase III (ALLEGRO study)	RRMS (n=1106)	Laquinimod 0.6mg OD: 52.24%; Placebo: 62.90%, p (vs. placebo) <0.001;	24 months
		Khan et al., Ann Neurol 2013, phase III (GALA study)	RRMS (n=1404)	GA 40mg sc tiw: 77.0% vs. Placebo: 65.5%, p<0.0001	12 months
		Lublin et al., Ann Neurol 2013, phase III (CombiRx study)	RRMS (n=1008)	IFN+GA: 61.1% vs. IFN: 55.6%, p=0.19 IFN+GA: 61.1% vs. GA: 64.1%, p=0.21 IFN: 55.6% vs. GA: 64.1%, p=0.14	36 months after last patient was included
		Calabresi et al., Lancet Neurol 2014, phase III (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 0.5mg: 71.5% (66.6 to 76.4); Placebo: 52.7% (2.8; 47.2 to 58.2), p<0.0001	24 months
		Confavreux et al., Lancet Neurol 2014, phase III (TOWER study)	RRMS (n=1169)	Teriflunomide 7mg: 71.9% (67.3 to 76.5), p (vs. placebo) =0.016 Teriflunomide 14mg: 76.3% (71.7 to 81.0), p (vs. placebo) <0.0001 Placebo: 60.6% (95% CI: 55.5 to 65.6);	48 weeks after the last patient was included
		Massacesi et al., PLoS ONE 2014, phase III	RRMS (n=150)	Azathioprine (target dose: 3 mg/kg/d) vs. IFN beta (1a or 1b): 62.9% vs. 47.7%, p=0.22 (ns)	24 months

		Mikol et al., Lancet Neurol 2014, phase III (REGARD study)	RRMS (n=764)	IFN beta-1a 44mcg SC tiw: 62%; Glatiramer acetate 20mg SC/day: 62%, p=0.64;	96 weeks
		Vollmer et al., J Neurol 2014, phase III (BRAVO study)	RRMS (n=1331)	Laquinimod 0.6mg: 66%, Placebo: 61%, IFN-beta 30 mcg IM: 69%; p (Laq vs. placebo)=0.21; p (IFN vs. placebo)=0.023	24 months
		Cohen et al., JAMA Neurol 2015, phase III (GATE study)	RRMS (n=796)	Generic GA 20mg/d vs. brand GA 20mg/d vs. placebo: 79.3% vs. 73.9% vs. 73.8% (ns)	9 months
		Kappos et al., New Engl J Med 2015, phase III (DECIDE study)	RRMS (n=1841)	Daclizumab HYP 150mg/4 weeks vs. IFN beta-1a 30µg/week: 67% vs. 51%, p<0.05	144 weeks
		Lanzillo et al., Mult Scler Journal 2015, phase III (ARIANNA study)	RRMS (n=154)	IFN beta-1b 8 MIU eod SC + atorvastatin 40mg PO vs. IFN beta-1b MIU eod SC + placebo: 69% vs. 75% (ns)	24 months
	Time to first confirmed relapse	Johnson et al., Neurology 1995, phase III (The Copolymer 1 Multiple Sclerosis Study)	RRMS (n=251)	Median time: Glatiramer acetate 20mg SC/day: 287 days, vs. placebo: 198 days, p=0.097	24 months
		PRISMS study group (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)	RRMS (n=560)	Median time to relapse: delayed by 3 or 5 months, for IFN beta-1a 22mcg SC tiw or IFN beta-1a 44mcg SC tiw, vs. placebo, respectively (p<0.05);	24 months
		Panitch et al.,	RRMS	IFN beta-1a SC	24 months

		Neurology 2002; Schwid et al., Clinical Therapeutics 2007, phase 4 – post-commercialisation (EVIDENCE study)	(n=677)	44mcg tiw: 13.5 mo.; IFN beta-1a IM 30mcg/week: 6.7 mo.; HR (95% CI) 0.70 (0.56 to 0.88), p=0.002	(0-12m: comparative phase; 12-24m: cross-over phase) **(n)**
		O'Connor et al., Lancet Neurol 2009, phase III (BEYOND study)	RRMS (n=2244)	GA 20mg SC/day at 2 years FU: 271 days (25th percentile); IFN beta-1b 250mcg SC EOD at 2 years FU: 283 days (25th percentile); IFN beta-1b 500mcg SC EOD at 2 years FU: 348 days (25th percentile), p values (all comparisons) > 0.05	24 months
		Kappos et al., NEJM 2010, phase III (FREEDOMS study)	RRMS (n=1272)	Fingolimod 0.5mg/day vs. placebo: HR (95% CI) 0.48 (0.39 to 0.61), p<0.001 Fingolimod 1.25mg/day vs. placebo: HR (95% CI) 0.38 (0.30 to 0.48), p<0.001	24 months
		Giovannoni et al., NEJM 2010, phase III (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/Kg vs. placebo: HR (95% CI) 0.44 (0.34 to 0.58), p<0.001 Cladribine 5.25mg/Kg vs. placebo: HR (95% CI) 0.46 (0.36 to 0.60), p<0.001	96 weeks
		Sorensen et al., Lancet Neurology 2011, phase 4 (SIMCOMBIN study)	RRMS (n=307)	IFN beta-1a 30mcg IM/week + simvastatin 80mg/day vs. IFN beta-1a	12 months after last patient was included

				30mcg IM/week + Placebo: HR (95% CI) 1.21 (0.74 to 1.99), p=0.512	
		Fox et al., NEJM 2012, phase III (CONFIRM study)	RRMS (n=1417)	BG-12 240mg BD vs. placebo: HR (95% CI) 0.66 (0.51 to 0.86), p≤0.01; BG-12 240mg TDS vs. placebo: HR (95% CI) 0.55 (0.42 to 0.73), p<0.001; GA 40mg SC/day vs. placebo: HR (95% CI) 0.71 (0.55 to 0.92), p≤0.01;	24 months
		Gold et al., NEJM 2012, phase III (DEFINE study)	RRMS (n=1234)	BG-12 240mg BD vs. placebo: HR (95% CI) 0.51 (0.40 to 0.66), p<0.001; BG-12 240mg TDS vs. placebo: HR (95% CI) 0.50 (0.39 to 0.65), p<0.001;	24 months
		O'Connor et al., NEJM 2011, phase III (TEMSO study)	Relapsing MS (n=1088)	Teriflunomide 7mg PO/day vs. placebo: HR (95% CI) 0.76 (0.61–0.94), p=0.01; Teriflunomide 14mg PO/day vs. placebo: HR (95% CI) 0.72 (0.58–0.90), p=0.003;	108 weeks
		Khan et al., Ann Neurol 2013, phase III (GALA study)	RRMS (n=1404)	GA 40mg sc tiw vs. placebo: HR (95% CI) 0.606 (0.493 to 0.744), p<0.0001	12 months
		Lublin et al., Ann Neurol 2013, phase III (CombiRx study)	RRMS (n=1008)	HRs not specified, p=0.19	36 months after last patient was included

		Massacesi et al., PLoS ONE 2014, phase III	RRMS (n=150)	Azathioprine (target dose: 3 mg/kg/d) vs. IFN beta (1a or 1b) (hazard ratio [95%CI]): 0.66 (0.40–1.10) (ns)	24 months
		Mikol et al., Lancet Neurol 2014, phase III (REGARD study)	RRMS (n=764)	IFN beta-1a 44mcg SC tiw vs. glatiramer acetate 20mg SC/day, HR (95% CI) 0.94 (0.74–1.21), p=0.64;	96 weeks
	Time to failure, defined as the occurrence of the first confirmed relapse or to permanent treatment discontinuation for any cause	Vermersch et al., MSJ 2014, phase III (TENERE study)	Relapsing MS (n=324)	Teriflunomide 7mg vs. IFNβ-1a: HR (95% CI) 1.12 (0.75 to 1.67), p=0.52; Teriflunomide 14mg vs. IFNβ-1a: HR (95% CI) 0.86 (0.56 to 1.31), p=0.60	48 weeks after the last patient was included
	Relapse risk (assessed with the Andersen–Gill model for time to recurring events)	O'Connor et al., Lancet Neurol 2009, phase III (BEYOND study)	RRMS (n=2244)	IFN beta-1b 500mcg SC eod vs. IFN beta-1b 250mcg SC eod: HR (95% CI) 0·94 (0·82–1·08), p=0·20; IFN beta-1b 500mcg SC eod vs. GA 20mg SC/day: HR (95% CI)1·00 (0·83–1·19), p=0·48; IFN beta-1b 250mcg SC eod vs. GA 20mg SC/day: HR (95% CI) 1·06 (0·89–1·27), p=0·74;	24 months
EDSS score	**Change in EDSS score from baseline to follow-up (k)**	The IFNB Multiple Sclerosis Study Group, Neurology 1993, phase III	RRMS (n=372)	IFN beta-1b 1.6 MIU, IFN beta-1b 8 MIU or placebo: little changes (not significant – no further details given in the abstract);	24 months

		Johnson et al., Neurology 1995, phase III (The Copolymer 1 Multiple Sclerosis Study)	RRMS (n=251)	Glatiramer acetate 20mg SC/day: -0.05 (SE 1.13); Placebo: 0.21(SE 0.99), p=0.023	24 months
		PRISMS study group (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)	RRMS (n=560)	IFN beta-1a 22mcg SC tiw: 0.23 (SD 1.3), p≤0.05 (vs. placebo); IFN beta-1a 44mcg SC tiw: 0.24 (SD 1.1), p≤0.05 (vs. placebo); Placebo: 0.48 (SD 1.3);	24 months
		Cohen et al., NEJM 2010, phase III (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 0.5mg/day: -0.08 (SD 0.79), p (vs. IFN) = 0.06 Fingolimod 1.25mg/day: -0.11 (SD 0.90), p (vs. IFN) = 0.02 IFN beta-1a IM 30mcg/week: 0.01 (SD 0.78)	12 months
		Kappos et al., NEJM 2010, phase III (FREEDOMS study)	RRMS (n=1272)	Fingolimod 0.5mg/day: 0.00 (SD 0.88), p (vs. placebo) = 0.002; Fingolimod 1.25mg/day: -0.03 (SD 0.88), p (vs. placebo) = 0.002; Placebo: 0.13 (SD 0.94);	24 months
		Cohen et al., Lancet 2012, phase III (CARE-MS I study)	RRMS previously untreated (n=581)	Alemtuzumab 12mg IV/day x 5 days: -0.14 (95% CI -0.25 to -0.02) IFN beta 1a 44mcg SC tiw: -0.14 (95% CI -0.29 to 0.01), p=0.97	24 months
		Coles et al., Lancet	RRMS	Alemtuzumab	24 months

		2012, phase III (CARE-MS II study)	previously treated (n=840)	12mg IV/day x 5 days: -0.17 (95% CI -0.29 to -0.05); IFN beta 1a 44mcg SC tiw: 0.24 (95% CI 0.07 to 0.41), p<0.0001	
		Calabresi et al., Lancet Neurol 2014, phase III (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 0.5mg PO: 0·046 (SD: 1.02); Placebo: 0.055 (SD: 1.20), p=0.945	24 months
		Confavreux et al., Lancet Neurol 2014, phase III (TOWER study)	RRMS (n=1169)	Teriflunomide 7mg PO: 0.04 (0.05), p (vs. placebo) = 0.4819; Teriflunomide 14mg PO: -0.05 (0.05), p (vs. placebo) = 0.0429; Placebo: 0.09 (0.05);	48 weeks after the last patient was included
		Massacesi et al., PLoS ONE 2014, phase III	RRMS (n=150)	Azathioprine (target dose: 3 mg/kg/day PO) vs. IFN beta (1a or 1b SC) (mean change [95%CI]): 20.08 (20.3 to 0.16) vs. 0.22 (20.03 to 0.47), p=0.08	24 months
		Cohen et al., JAMA Neurol 2015, phase III (GATE study)	RRMS (n=796)	Generic GA 20mg/d: (mean change [range]) −0.11 (−0.22 to 0.00); Brand GA 20mg/d: (mean change [range]) −0.08 (−0.19 to 0.03); Placebo: (mean change [range]): −0.02 (−0.17 to 0.14); p-values (all comparisons) >0.05	9 months
		Lanzillo et al., Mult	RRMS	IFN beta-1b 8	24 months

		Scler Journal 2015, phase III (ARIANNA study)	(n=154)	MIU SC eod + atorvastatin 40mg/d: 0.3 vs. IFN beta-1b 8 MIU SC eod + placebo: 0.2, p>0.05	
	Time to 3-month CDP (g)	Jacobs et al., Ann Neurol 1996, phase III (MSCRG study)	Relapsing MS (n=301)	IFN beta-1a 30mcg IM/week vs. placebo: HR <1, p=0.02 **(v)**	104 weeks
		PRISMS study group (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)	RRMS (n=560)	IFN beta-1a 22mcg SC tiw: 18.5 months (first quartile), risk ratio (95% CI) 0.68 (0.48 to 0.98): p (vs. placebo) <0.05; IFN beta-1a 44mcg SC tiw: 21.3 months (first quartile), risk ratio (95% CI) 0.42 (0.18 to 0.99), p (vs. placebo) <0.05; Placebo: 11.9 months (first quartile) **(u)**	24 months
		Noseworthy et al., Neurology 2000, phase III (linomide study)	RMS (n=715)	The study was of insufficient duration for any of the primary or secondary outcome measures to reach significance	Early termination for safety issues (initially planned: 36 months)
		O'Connor et al., Lancet Neurol 2009, phase III (BEYOND study)	RRMS (n=2244)	IFN beta-1b 500mcg SC EOD at 2 years FU: 190 days (10^{th} percentile); IFN beta-1b 250mcg SC EOD at 2 years FU: 274 days (10^{th} percentile); GA 20mg SC/day at 2 years FU: 268 days (10^{th} percentile), p	24 months

				values (all comparisons) > 0.05	
		Kappos et al., NEJM 2010, phase III (FREEDOMS study)	RRMS (n=1272)	Fingolimod 0.5mg/day vs. placebo: HR (95% CI) 0.70 (0.52 to 0.96), p = 0.02 Fingolimod 1.25mg/day vs. placebo: HR (95% CI) 0.68 (0.50 to 0.93), p = 0.02	24 months
		Giovannoni et al., NEJM 2010, phase III (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/Kg vs. placebo: HR (95% CI) 0.67 (0.48 to 0.93), p<0.001 Cladribine 5.25mg/Kg vs. placebo: HR (95% CI) 0.69 (0.49 to 0.96), p<0.001	96 weeks
		O'Connor et al., NEJM 2011, phase III (TEMSO study)	Relapsing MS (n=1088)	Teriflunomide 7mg PO/day vs. placebo: HR (95% CI) 0.76 (0.56–1.05), p=0.08 Teriflunomide 14mg PO/day vs. placebo: HR (95% CI) 0.70 (0.51–0.97); p=0.03	108 weeks
		Sorensen et al., Lancet Neurology 2011, phase 4 (SIMCOMBIN study)	RRMS (n=307)	IFN beta-1a 30mcg IM/week + simvastatin 80mg/day vs. mThe IFN beta-1a 30mcg IM/week + Placebo: HR (95% CI) 1.01 (0.63 to 1.64), p=0.953	12 months after last patient was included
		Comi et al., NEJM 2012, phase III (ALLEGRO study)	RRMS (n=1106)	Laquinimod 0.6mg OD vs. placebo: HR (95% CI) 0.64	24 months

				(0.45 to 0.91), p=0.01	
		Fox et al., NEJM 2012, phase III (CONFIRM study)	RRMS (n=1417)	BG-12 240mg BD vs. placebo: HR (95% CI) 0.79 (0.52 to 1.19), p>0.05; BG-12 240mg TDS vs. placebo: HR (95% CI) 0.76 (0.50 to 1.16), p>0.05 GA 40mg SC/day vs. placebo: HR (95% CI) 0.93 (0.63 to 1.37), p>0.05	24 months
		Gold et al., NEJM 2012, phase III (DEFINE study)	RRMS (n=1234)	BG-12 240mg BD vs. placebo: HR (95% CI) 0.62 (0.44 to 0.87), p=0.005; BG-12 240mg TDS vs. placebo: HR (95% CI) 0.66 (0.48 to 0.92), p=0.01;	24 months
		Confavreux et al., Lancet Neurol 2014, phase III (TOWER study)	RRMS (n=1169)	Teriflunomide 7mg vs. placebo: HR (95% CI) 0.95 (0.68 to 1.35), p= 0.7620; Teriflunomide 14mg vs. placebo: HR (95% CI) 0.68 (0.47 to 1.00), p=0.0442	48 weeks after the last patient was included
		Vollmer et al., J Neurol 2014, phase III (BRAVO study)	RRMS (n=1331)	Laquinimod 0.6mg vs. placebo: HR (95% CI) 0.69 (0.46–1.02), p=0.063; IFN-beta 30 mcg IM vs. placebo: HR (95% CI0.74 (0.51–1.09), p=0.13	24 months
	% patients with 3-month CDP	Johnson et al., Neurology 1995,	RRMS (n=251)	Glatiramer acetate 20mg	24 months

		phase III (The Copolymer 1 Multiple Sclerosis Study)		SC/day: 20.8%; Placebo: 28.8%, p=0.037	
		Jacobs et al., Ann Neurol 1996, phase III (MSCRG study)	Relapsing MS (n=301)	IFN beta-1a 30mcg IM/week: 21.9%; Placebo: 34.9%, p<0.05 **(v)**	104 weeks
		Noseworthy et al., Neurology 2000, phase III (linomide study)	RMS (n=715)	The study was of insufficient duration for any of the primary or secondary outcome measures to reach significance	Early termination for safety issues (initially planned: 36 months)
		Polman et al., NEJM 2006, phase III (AFFIRM study)	RRMS (n=627)	Natalizumab 300mg/4 weeks: 17%, p (vs. placebo) <0.001; Placebo: 29%	24 months
		Rudick et al., NEJM 2006, phase III (SENTINEL study)	RRMS (n=1171)	Natalizumab 300mg/4 weeks + IFN beta-1a IM 30mcg/week: 23%; IFN beta-1a IM 30mcg/week: 29%, p=0.02	24 months
		O'Connor et al., Lancet Neurol 2009, phase III (BEYOND study)	RRMS (n=2244)	IFN beta-1b 500mcg SC EOD at 2 years FU: 22% IFN beta-1b 250mcg SC EOD at 2 years FU: 21%; GA 20mg SC/day at 2 years FU: 20%, p values (all comparisons) > 0.05	24 months
		O'Connor et al., NEJM 2011, phase III (TEMSO study)	Relapsing MS (n=1088)	Teriflunomide 7mg PO/day: 21.7 (17.1–26.3), p (vs. placebo) = 0.08; Teriflunomide 14mg PO/day:	108 weeks

				20.2 (15.6–24.7), p (vs. placebo) = 0.03; Placebo: 27.3 (22.3–32.3)	
		Sorensen et al., Lancet Neurology 2011, phase 4 (SIMCOMBIN study)	RRMS (n=307)	IFN beta-1a 30mcg IM/week + Placebo: 24%; IFN beta-1a 30mcg IM/week + simvastatin 80mg/day: 28%, p=0.953	12 months after last patient was included
		Comi et al., NEJM 2012, phase III (ALLEGRO study)	RRMS (n=1106)	Laquinimod 0.6mg OD: 11.1%, p (vs. placebo) =0.01; Placebo: 15.7%	24 months
		Fox et al., NEJM 2012, phase III (CONFIRM study)	RRMS (n=1417)	BG-12 240mg BD: 13%, p (vs. placebo) >0.05; BG-12 240mg TDS: 13%, p (vs. placebo) >0.05; GA 40mg SC/day: 16%, p (vs. placebo) >0.05; Placebo: 17%	24 months
		Gold et al., NEJM 2012, phase III (DEFINE study)	RRMS (n=1234)	BG-12 240mg BD: 16%, p (vs. placebo) = 0.005; BG-12 240mg TDS: 18%, p (vs. placebo) = 0.01; Placebo: 27%	24 months
		Calabresi et al. Lancet Neurol 2014, phase III (ADVANCE study)	RRMS (n=1516)	Peginterferon beta-1a 125µg/2 weeks SC vs. placebo: 0.068 (0.0119) vs. 0.105 (0.0142), p=0.0383; Peginterferon beta-1a 125µg/4 weeks SC vs. placebo: 0.068 (0.0119) vs. 0.105 (0.0142), p=0.0380 **(e)**	24 months (but primary endpoint: 48 weeks, which is the placebo-controlled phase)
		Massacesi et al.,	RRMS	Azathioprine	24 months

		PLoS ONE 2014, phase III	(n=150)	(target dose: 3 mg/kg/d) vs. IFN beta (1a or 1b SC): 1.8% vs. 8%, p=0.19	
		Vollmer et al., J Neurol 2014, phase III (BRAVO study)	RRMS (n=1331)	Laquinimod 0.6mg: 10%; IFN-beta 30 mcg IM: 11%; Placebo: 13%; p (Laq vs. placebo)=0.063; p (IFN vs. placebo)=0.13	24 months
		Kappos et al., New Engl J Med 2015, phase III (DECIDE study)	RRMS (n=1841)	Daclizumab HYP 150mg/4 weeks vs. IFN beta-1a 30mcg/week: 16% vs. 20% (p=0.16)	144 weeks
	Time to 6-month CDP	Kappos et al., NEJM 2010, phase III (FREEDOMS study)	RRMS (n=1272)	Fingolimod 0.5mg/day vs. placebo: HR (95% CI) 0.63 (0.44 0.90), p = 0.01 Fingolimod 1.25mg/day vs. placebo: HR (95% CI) 0.60 (0.41 to 0.86), p = 0.006	24 months
		Sorensen et al., Lancet Neurology 2011, phase 4 (SIMCOMBIN study)	RRMS (n=307)	IFN beta-1a 30mcg IM/week + simvastatin 80mg/day vs. IFN beta-1a 30mcg IM/week + Placebo: HR 0.991, p=0.986	12 months after last patient was included
		Cohen et al., Lancet 2012, phase III (CARE-MS I study)	RRMS previously untreated (n=581)	IFN beta 1a 44mcg SC tiw vs. Alemtuzumab 12mg IV/day x 5 days: HR (95% CI) 0.70 (0.40 to 1.23), p=0.22	24 months
		Coles et al., Lancet 2012, phase III (CARE-MS II study)	RRMS previously treated (n=840)	IFN beta 1a 44mcg SC tiw vs. Alemtuzumab 12mg IV/day x 5 days: HR (95%	24 months

				CI) 0.58 (0.38 to 0.87), p=0.0084	
		Comi et al., NEJM 2012, phase III (ALLEGRO study)	RRMS (n=1106)	Laquinimod 0.6mg OD vs. placebo: HR (95% CI) 0.51 (0.34 to 0.79), p=0.002	24 months
		Vollmer et al., J Neurol 2014, phase III (BRAVO study)	RRMS (n=1331)	Laquinimod 0.6mg vs. placebo: HR (95% CI) 0.610 (0.38 to 0.98), p=0.042; IFN-beta 30 mcg IM vs. placebo: HR (95% CI) 0.73 (0.47–1.14), p=0.17	24 months
	% patients with 6-month CDP	Cohen et al., Lancet 2012, phase III (CARE-MS I study)	RRMS previously untreated (n=581)	Alemtuzumab 12mg IV/day x 5 days: 8.00% (95% CI 5.66 to 11.24); IFN beta-1a 44mcg SC tiw: 11.12% (95% CI 7.32 to 16.71), p=0.22	24 months
		Coles et al., Lancet 2012, phase III (CARE-MS II study)	RRMS previously treated (n=840)	Alemtuzumab 12mg IV/day x 5 days: 12.71% (95% CI 9.89 to 16.27); IFN beta-1a 44mcg SC tiw: 21.13% (95% CI 15.95 to 27.68), p=0.0084	24 months
		Lublin et al., Ann Neurol 2013, phase III (CombiRx study)	RRMS (n=1008)	IFN+GA: 23.9% vs. IFN: 21.6%, p>0.05 IFN+GA: 23.9% vs. GA: 24.8%, p>0.05 IFN: 21.6% vs. GA: 24.8%, p>0.05	36 months after last patient was included
		Vollmer et al., J Neurol 2014, phase III (BRAVO study)	RRMS (n=1331)	Laquinimod 0.6mg: 7%; IFN-beta 30 mcg IM: 8%; Placebo: 10%; p	24 months

				(Laq vs. placebo)=0.042; p (IFN vs. placebo)=0.17	
		Lanzillo et al., Mult Scler Journal 2015, phase III (ARIANNA study)	RRMS (n=154)	IFN beta-1b SC eod + atorvastatin 40mg PO/day vs. IFN beta-1b SC eod + placebo: 7.9 vs. 3.8, p>0.05	24 months
	% patients free from EDSS progression, confirmed at 3 months	Johnson et al., Neurology 1995, phase III (The Copolymer 1 Multiple Sclerosis Study)	RRMS (n=251)	Glatiramer acetate 20mg SC/day: 78.4%, Placebo: 75.4%, p>0.05	24 months (no MRI results published)
		PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)	RRMS (n=560)	IFN beta-1a 22mcg SC tiw: ~ 60%, p (vs. placebo) <0.05; IFN beta-1a 44mcg SC tiw: ~ 74%, p (vs. placebo) <0.05; Placebo: ~ 48%	24 months
		Cohen et al., NEJM 2010, phase III (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 0.5mg/day: 94.1% (91.8 to 96.3), p (vs. IFN) = 0.25; Fingolimod 1.25mg/day: 93.3% (90.9 to 95.8), p (vs. IFN) = 0.50; IFN beta-1a IM 30mcg/week: 92.1% (95% CI 89.4 to 94.7)	12 months
		Kappos et al., NEJM 2010, phase III (FREEDOMS study)	RRMS (n=1272)	Fingolimod 0.5mg/day: 82.3% (78.6 to 86.1), p (vs. placebo) = 0.03; Fingolimod 1.25mg/day: 83.4% (79.7 to 87.1), p (vs. placebo) = 0.01; Placebo: 75.9% (95% CI71.7 to 80.2)	24 months
		Giovannoni et al.,	RRMS	Cladribine	96 weeks

		NEJM 2010, phase III (CLARITY study)	(n=1326)	3.5mg/Kg: 85.7%, p (vs. placebo) =0.02; Cladribine 5.25mg/Kg: 84.9%, p (vs. placebo) =0.03; Placebo: 79.4%	
		Calabresi et al., Lancet Neurol 2014, phase III (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 0.5mg: 74.7% (69.9 to 79.5); Placebo: 71.0% (65.9 to 76.1), p=0.320	24 months
		Confavreux et al., Lancet Neurol 2014, phase III (TOWER study)	RRMS (n=1169)	Teriflunomide 7mg: 78.9% (73.9 to 83.9), p=0·7620; Teriflunomide 14mg: 84.2% (79.6 to 88.8), p=0.0442; Placebo: 80.3% (75.9 to 84.8)	48 weeks after the last patient was included
	% patients free of EDSS progression, confirmed at 6 months	Kappos et al., NEJM 2010, phase III (FREEDOMS study)	RRMS (n=1272)	Fingolimod 0.5mg/day: 87.5% (84.7 to 90.7), p (vs. placebo) = 0.01; Fingolimod 1.25mg/day: 88.5% (85.3 to 91.6), p (vs. placebo) = 0.004; Placebo: 81.0% (95% CI 77.1 to 84.9)	24 months
		Calabresi et al., Lancet Neurol 2014, phase III (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 0.5mg: 86.2% (82.3 to 90.0); Placebo: 82.2% (77.9 to 86.4), p=0.101	24 months
		Mikol et al., Lancet Neurol 2014, phase III (REGARD study)	RRMS (n=764)	IFN beta-1a 44mcg SC tiw: 11.7%; Glatiramer acetate 20mg SC/day: 8.7%, p=0.117	96 weeks
	% patients with improvement of EDSS after 24 months	Johnson et al., Neurology 1995, phase III (The Copolymer 1 Multiple Sclerosis	RRMS (n=251)	Glatiramer acetate 20mg SC/day: 24.8%; Placebo: 15.2%, p=0.037	24 months

		Study)			
	% patients with sustained EDSS reduction for 6 months	Coles et al., Lancet 2012, phase III (CARE-MS II study)	RRMS previously treated (n=840)	Alemtuzumab 12mg IV/day x 5 days: 28.82% (95% CI 24.18 to 34.13); IFN beta 1a 44mcg SC tiw: 12.93% (95% CI 8.34 to 19.77), p=0.0002	24 months
MSFC	**Score at FU**	Comi et al., NEJM 2012, phase III (ALLEGRO study)	RRMS (n=1106)	Laquinimod 0.6mg PO/day: 0.04 (-0.02 to 0.09); Placebo: 0.06 (0.00 to 0.11), p=0.59 **(j)**	24 months
	Change in MSFC z-score from baseline to follow-up (f) (k)	Cohen et al., NEJM 2010, phase III (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 0.5mg/day: 0.04 (SD 0.42), p (vs. IFN) = 0.02; Fingolimod 1.25mg/day: 0.08 (SD 0.46), p (vs. IFN) <0.001; IFN beta-1a IM 30mcg/week: -0.03 (SD 0.48)	12 months
		Kappos et al., NEJM 2010, phase III (FREEDOMS study)	RRMS (n=1272)	Fingolimod 0.5mg/day: 0.03 (SD 0.39), p (vs. placebo) = 0.01; Fingolimod 1.25mg/day: 0.01 (SD 0.40), p (vs. placebo) = 0.02; Placebo: -0.06 (SD 0.57)	24 months
		Cohen et al., Lancet 2012, phase III (CARE-MS I study)	RRMS previously untreated (n=581)	Alemtuzumab 12mg IV/day x 5 days: 0·15 (SD 0·52); IFN beta 1a 44mcg SC tiw: 0.07 (SD 0.45), p=0.01	24 months
		Coles et al., Lancet 2012, phase III (CARE-MS II study)	RRMS previously treated (n=840)	Alemtuzumab 12mg IV/day x 5 days: 0.08 (0.04 to 0.12); IFN beta 1a	24 months

				44mcg SC tiw: -0.04 (95% CI -0.10 to 0.02), p=0.002;	
		Lublin et al., Ann Neurol 2013, phase III (CombiRx study)	RRMS (n=1008)	IFN+GA: 0.1 (SD 0.5) vs. IFN: 0.1 (SD 0.5), p>0.05 IFN+GA: 0.1 (SD 0.5) vs. GA: 0.2 (SD 0.5), p>0.05 IFN: 0.1 (SD 0.5) vs. GA: 0.2 (SD 0.5), p>0.05	36 months after last patient was included
		Calabresi et al., Lancet Neurol 2014, phase III (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 0.5mg PO/day: 0·00 (0·60); Placebo: –0·07 (0·54), p=0·012	24 months
	% patients with decrease ≥20% in MSFC	Lanzillo et al., Mult Scler Journal 2015, phase III (ARIANNA study)	RRMS (n=154)	IFN beta-1b eod SC + atorvastatin 40mg PO/day: 0.08; IFN beta-1b eod SC + placebo: 0.09, p>0.05	24 months
Ambulation index	**Score at FU**	Johnson et al., Neurology 1995, phase III (The Copolymer 1 Multiple Sclerosis Study)	RRMS (n=251)	Glatiramer acetate 20mg SC/day: 0.27 (SE 0.94); Placebo: 0.28 (SE 0.93), p>0.05	24 months
		PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)	RRMS (n=560)	IFN beta-1a 44mcg SC tiw: better than placebo (p<0.05); no further details given	24 months
	% patients with 3-month CDP (t)	PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)	RRMS (n=560)	IFN beta-1a 22mcg SC tiw: 12%, p>0.05 (vs. placebo); IFN beta-1a 44mcg SC tiw: 7%, p≤0.05 (vs. placebo); Placebo: 13%	24 months
Arm index	**Change from baseline to FU**	PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in	RRMS (n=560)	IFN beta-1a 22mcg SC tiw, IFN beta-1a 44mcg SC tiw, placebo: no	24 months

		Multiple Sclerosis) Study Group, Lancet 1998, phase III (PRISMS study)		changes in any of the groups (no differences – no further details given)	
Rao's Brief Repeatable Battery	**% patients with change in cognitive impairment (c)**	Lanzillo et al., Mult Scler Journal 2015, phase III (ARIANNA study)	RRMS (n=154)	IFN beta-1b SC eod + atorvastatin 40mg/d: −37.1 IFN beta-1b SC eod + placebo: −35.2, p>0.05	24 months
No evidence of clinical activity (NECA)	**% of patients with no evidence of clinical activity (no relapses and no progression of disability)**	Cohen et al., Lancet 2012, phase III (CARE-MS I study)	RRMS previously untreated (n=581)	Alemtuzumab 12mg IV/day x 5 days: 74%; IFN beta 1a 44mcg SC tiw: 56%, p<0.0001	24 months
		Coles et al., Lancet 2012, phase III (CARE-MS II study)	RRMS previously treated (n=840)	Alemtuzumab 12mg IV/day x 5 days: 60%; IFN beta 1a 44mcg SC tiw: 41%, p<0.0001	24 months
		Lublin et al., Ann Neurol 2013, phase III (CombiRx study)	RRMS (n=1008)	IFN+GA: 45.4% vs. IFN: 46.9%, p=0.35; IFN+GA: 45.4% vs. GA: 47.4%, p=0.35; IFN: 46.9% vs. GA: 47.4%, p=0.92	36 months after last patient was included
Unidimensional Fatigue Impact Scale (FIS or UFIS)	**Change from baseline to FU**	O'Connor et al., NEJM 2011, phase III (TEMSO study)	Relapsing MS (n=1088)	Teriflunomide 7mg PO/day: 2.3 (SD 1.6), p (vs. placebo) = 0.39; Teriflunomide 14mg PO/day: 3.8 (SD 1.7), p (vs. placebo) = 0.83; Placebo: 4.3 (SD 1.7)	108 weeks
		Confavreux et al., Lancet Neurol 2014, phase III (TOWER study)	RRMS (n=1169)	Teriflunomide 7mg: 4.46 (1.66), p (vs. placebo) = 0.3686; Teriflunomide 14mg: 2.04 (1.68), p (vs. placebo) = 0.0429;	48 weeks after the last patient was included

				Placebo: 6.31(1.67);	
		Vermersch et al., MSJ 2014, phase III (TENERE study)	Relapsing MS (n=324)	Teriflunomide 7mg: 0.97 (2.96), p (vs. placebo) = 0.03; Teriflunomide 14mg: 4.10 (3.03), p (vs. placebo) = 0.18; Placebo: 9.10 (SE 3.21)	48 weeks after the last patient was included
MSIS-29	**% patients with worsening in MSIS-29 (global score)**	Kappos et al., New Engl J Med 2015, phase III (DECIDE study)	RRMS (n=1841)	Daclizumab HYP 150mg/4 weeks: 19% IFN beta-1a 30mcg IM/week: 23% **(d)**	144 weeks (this outcome was evaluated at 96 weeks)
SF-36	**Change in physical summary score from baseline to last FU**	Confavreux et al., Lancet Neurol 2014, phase III (TOWER study)	RRMS (n=1169)	Teriflunomide 7mg: -0.91 (0.44), p (vs. placebo) = 0.1772; Teriflunomide 14mg: -0.64 (0.44), p (vs. placebo) = 0.0687; Placebo: -1.63 (0.44)	48 weeks after the last patient was included
	Change in mental summary score from baseline to last FU	Confavreux et al., Lancet Neurol 2014, phase III (TOWER study)	RRMS (n=1169)	Teriflunomide 7mg: -1.70 (0.60), p (vs. placebo) = 0.1363; Teriflunomide 14mg: -1.09 (0.59), p (vs. placebo) = 0.0224; Placebo: -2.79 (0.59)	48 weeks after the last patient was included
TSQM	**Effectiveness domain, score at FU**	Vermersch et al., MSJ 2014, phase III (TENERE study)	Relapsing MS (n=324)	Teriflunomide 7mg: 67.25 (SE 2.70), p (vs. placebo) = 0.02; Teriflunomide 14mg: 63.13 (SE 2.75), p (vs. placebo) = 0.28; Placebo: 59.30 (SE 2.97)	48 weeks after the last patient was included
	Side-effects domain, score at FU	Vermersch et al., MSJ 2014, phase III (TENERE study)	Relapsing MS (n=324)	Teriflunomide 7mg: 95.29 (2.31), p (vs.	48 weeks after the last patient

				placebo) <0.0001; Teriflunomide 14mg: 93.15 (2.34), p (vs. placebo) = <0.0001; Placebo: 71.38 (SE 2.50)	was included
	Convenience domain, score at FU	Vermersch et al., MSJ 2014, phase III (TENERE study)	Relapsing MS (n=324)	Teriflunomide 7mg: 88.30 (1.97), p (vs. placebo) <0.0001; Teriflunomide 14mg: 89.85 (1.98), p (vs. placebo) <0.0001; Placebo: 61.90 (SE 2.11)	48 weeks after the last patient was included
	Global satisfaction domain, score at FU	Vermersch et al., MSJ 2014, phase III (TENERE study)	Relapsing MS (n=324)	Teriflunomide 7mg: 68.29 (2.77), p (vs. placebo) = 0.02; Teriflunomide 14mg: 68.82 (2.78), p (vs. placebo) = 0.02; Placebo: 60.98 (SE 2.94)	48 weeks after the last patient was included
No evidence of disease activity (NEDA)	**% of patients with no evidence of disease activity (no relapses + no progression of disability + no MRI activity (h))**	Cohen et al., Lancet 2012, phase III (CARE-MS I study)	RRMS previously untreated (n=581)	IFN beta 1a 44mcg SC tiw: 27% vs. Alemtuzumab 12mg IV/day x 5 days: 39%, p=0.006	24 months
		Coles et al., Lancet 2012, phase III (CARE-MS II study)	RRMS previously treated (n=840)	IFN beta 1a 44mcg SC tiw: 14% vs. Alemtuzumab 12mg IV/day x 5 days: 32%, p<0.0001	24 months
		Lublin et al., Ann Neurol 2013, phase III (CombiRx study)	RRMS (n=1008)	IFN+GA: 26.9% vs. IFN: 17.1%, p=0.002; IFN+GA: 26.9% vs. GA: 16.1%, p=0.001; IFN: 17.1% vs. GA: 16.1%, p=0.762	36 months after last patient was included

Table footnote:
(a) ARR refers to mean ARR per each group; it includes confirmed relapse rate, which includes rate of relapses with confirmed increase in EDSS (Voskuhl et al., Lancet Neurol 2016) and also adjusted mean relapse rate (Vollmer et al., J Neurol 2014)
(b) No detailed figures provided
(c) Cognitive impairment was defined on the number of failed tests, as mild (one to two tests failed) or moderate–severe (three or more tests failed)
(d) Defined as ≥7.5 points increase in MSIS-29
(e) CDP: Confirmed disability progression was defined as an increase of Expanded Disability Status Scale score of at least 1·0 point for patients with a baseline score of 1·0 or more, or an increase of at least 1·5 points for patients with a baseline score of 0, confirmed after 12 weeks. For the rest, EDSS increase of ≥1 point if EDSS ≤5.5; EDSS increase of ≥0.5 point if EDSS > 5.5;
(f) Includes adjusted MSFC z-score; also it may include values obtained at an early termination time point if this occurred after 12 months.
(g) Includes time to sustained accumulation of disability, which is considered as increase in 1 point in EDSS sustained for a minimum of 12 weeks (Confavreux et al., Lancet Neurol 2014, TOWER trial)
(h) No MRI activity includes: no new/enlarging lesions and no gadolinium-enhancing lesions
(i) Includes relapses requiring hospitalization/IV steroids (Comi et al., NEJM 2012, ALLEGRO study)
(j) Adjusting for baseline values of MSFC z-score, ANCOVA model
(k) Mean change reported, unless otherwise specified
(l) It includes 'at least 1 major relapse'
(m) The authors also estimated the proportion of patients with: i) at least one MS-related admission to hospital; ii) at least 1 MS-rekated steroid course
(n) The results shown refer to the comparative phase (0-12m) of the trial, where half of the patients were receiving IFN beta-1a IM 30mcg/week and the other half IFN beta-1a SC 44mcg tiw.
(o) p-value not specified
(p) this analysis refers to disability progression in both hands
(q) worsening in 9HPT is defined as deterioration greater or equal to 20%
(r) confirmed at 2 months
(s) mean number of relapses per patient during the trial/2 years (duration of trial)
(t) defined as 2-step increase (sustained for 3 months)
(u) in this context, this outcome measure (risk ratio or odds ratio) is equivalent to hazard ratio in the survival model
(v) timing for CDP not specified. Assumed 3 months
(w) this study looked at disability progression at the end of FU, so it is possible that just progression confirmed at just 3 months is also included here
(x) This refers to McDonald 2005 criteria

Abbreviations. BD: twice per day; CDP: confirmed disability progression; CI: confidence interval; eod: every other day; FU: follow-up; GA: glatiramer acetate; HR:

hazard ratio; IA & AHSCT: immunoablation and autologous haemopoietic stem-cell transplantation; IFN: interferon; IQR: interquartile range; MIU: million international units; MSCT: mesenchymal stem cell transplantation; MSFC: Multiple Sclerosis Functional Composite; MSIS-29: Multiple Sclerosis Impact Scale – 29 items; PO: per oral; RFSS: Regional Functional System Score; SC: subcutaneous; SF-36: Short Form 36 Health Survey (SF-36); SNRS: Scripps Neurological Rating Scale; TDS: three times per day; tiw: three times in a week; TSQM: Treatment Satisfaction Questionnaire for Medication, with domains for Effectiveness, Side-Effects, Convenience and Global Satisfaction

Table 2: Clinical outcome measures in phase III trials in clinically isolated syndromes (CIS)

Original clinical outcome	Derived outcome measures	Trial	Condition (no. of patients randomised)	Drug, effect (vs. placebo/ another active arm)	Duration of the trial
Relapses	**Time to CDMS**	Jacobs et al., NEJM 2000, phase III (CHAMPS study)	CIS (n=383)	IFN beta-1a 30mcg IM/week vs. placebo: rate ratio (95% CI) 0.56 (0.38 to 0.81), p=0002	Early termination: obvious superiority of IFN over placebo (initially planned: 36 months)
		Comi et al., Lancet 2001, phase III (ETOMS study)	CIS (n=308)	IFN beta-1a 22mcg SC/week: mean time (95% CI) 569 days (317 to infinity) (30^{th} percentile); Placebo: mean time (95% CI) 252 days (173 to 413) (30^{th} percentile), p= 0.034	24 months
		Kappos et al., Neurology 2006, phase III (BENEFIT study)	CIS (n=487)	IFN beta-1b 250mcg SC/eod: mean time: 618 days (25^{th} percentile), vs. placebo: mean time: 255 days (25^{th} percentile); HR (95% CI) 0.50 (0.36 to 0.70), p<0.0001	24 months
		Comi et al., Lancet 2009, phase III (PreCISe study)	CIS (n=481)	GA 20mg SC/day vs. placebo: HR (95% CI) 0.55 (0.40 to 0.77), p=0.0005	36 months
		Comi et al., Lancet Neurol 2012, phase III (REFLEX study)	CIS (n=517)	IFN beta-1a 44mcg SC/week vs. placebo: HR (95% CI) 0·53 (0·35 to 0·79), p = 0.0023; IFN beta-1a 44mcg SC tiw vs. placebo: HR (95% CI) 0·48 (0·31 to 0·73), p = 0.0004; IFN beta-1a 44mcg SC tiw vs.	108 weeks

				IFN beta-1a 44mcg SC/week: HR (95% CI) 0.90 (0.56 to 1.43), p = 0.7737	
		Leist et al., Lancet Neurol 2014, phase III (ORACLE MS study)	CIS (n=616)	Cladribine 5.25mg/Kg vs. placebo: HR (95%CI): 0·38, 95% CI 0·25–0·58, p<0·0001; Cladribine 3.5mg/Kg vs. placebo: HR (95%CI): 0·33, 0·21–0·51, p<0·0001	96 weeks
		Miller et al., Lancet Neurol 2014, phase III (TOPIC study)	CIS (n=618)	Teriflunomide 7mg vs. placebo: 0·628 (0·416–0·949), p=0.0271; Teriflunomide 14mg vs. placebo: 0·574 (0·379–0·869), p=0.0087	108 weeks
	% patients with CDMS	Jacobs et al., NEJM 2000, phase III (CHAMPS study)	CIS (=383)	IFN beta-1a 30mcg IM/week: 35%; Placebo: 50%, p=0002	Early termination: obvious superiority of IFN over placebo (initially planned: 36 months)
		Comi et al., Lancet 2001, phase III (ETOMS study)	CIS (n=308)	IFN beta-1a 22mcg SC/week: 34%; Placebo: 45% , p=0.047	24 months
		Kappos et al., Neurology 2006, phase III (BENEFIT study)	CIS (n=487)	IFN beta-1b 250mcg SC/eod: 28%; Placebo: 45%, p<0.00001	24 months
		Comi et al., Lancet Neurol 2012, phase III (REFLEX study)	CIS (n=517)	IFN beta-1a 44mcg SC/week: 21.6%, p (vs. placebo) = 0.0023; IFN beta-1a 44mcg SC tiw: 20.6%, p (vs. placebo) = 0.0004; Placebo: 37.5%	108 weeks
		Miller et al., Lancet	CIS (n=618)	Teriflunomide 7mg: 19%, p (vs.	108 weeks

		Neurol 2014, phase III (TOPIC study)		placebo) = 0.0271; Teriflunomide 14mg: 18%, p (vs. placebo) = 0.0087; Placebo: 28%	
	Time to McDonald MS (x)	Kappos et al., Neurology 2006, phase III (BENEFIT study)	CIS (n=487)	IFN beta-1b 250mcg SC/eod vs. placebo: HR (95% CI) 0.54 (0.43 to 0.67), p<0.00001	24 months
		Comi et al., Lancet Neurol 2012, phase III (REFLEX study)	CIS (n=517)	IFN beta-1a 44mcg SC/week vs. placebo: HR (95% CI) 0·69 (0·54–0·87), p = 0.0080; IFN beta-1a 44mcg SC tiw vs. placebo: HR (95% CI) 0·49 (0·38–0·64), p<0.0001; IFN beta-1a 44mcg SC tiw vs. IFN beta-1a 44mcg SC/week: HR (95% CI) 0·71 (0·54–0·91), p = 0.0087	108 weeks
		Miller et al., Lancet Neurol 2014, phase III (TOPIC study)	CIS (n=618)	Teriflunomide 7mg vs. placebo: 0·686 (0·540–0·871), p=0.0020; Teriflunomide 14mg vs. placebo: 0·651 (0·515–0·822), p=0.0003	108 weeks
	% patients with McDonald MS (x)	Kappos et al., Neurology 2006, phase III (BENEFIT study)	CIS (n=487)	IFN beta-1b 250mcg SC/eod: 69%; Placebo: 85%, p<0.00001	24 months
		Comi et al., Lancet Neurol 2012, phase III (REFLEX study)	CIS (n=517)	IFN beta-1a 44mcg SC/week: 75.5%, p (vs. placebo) = 0.0080; IFN beta-1a 44mcg SC tiw: 62.5%, p (vs. placebo) <0.0001; Placebo: 85.8%	108 weeks
		Miller et al., Lancet Neurol 2014, phase III (TOPIC	CIS (n=618)	Teriflunomide 7mg: 62%, p (vs. placebo) = 0.0020; Teriflunomide 14mg: 64%, p (vs.	108 weeks

		study)		placebo) = 0.0003; Placebo: 76%	
	Mean annualised relapse rate (a)	Comi et al., Lancet 2001, phase III (ETOMS study)	CIS (n=308)	IFN beta-1a 22mcg SC/week: 0.33; Placebo: 0.43, p=0.045	24 months
		Miller et al., Lancet Neurol 2014, phase III (TOPIC study)	CIS (n=618)	Teriflunomide 7mg: 0·190 (0·139–0·260), p (vs. placebo) = 0.0541; Teriflunomide 14mg: 0·194 (0·143–0·263), p (vs. placebo) = 0.0579; Placebo: 0·284 (0·214–0·378)	108 weeks
	% patients with at least 1 relapse (a) (l)	Comi et al., Lancet 2009, phase III (PreCISe study)	CIS (n=481)	Placebo: 42.9%; GA 20mg SC/day: 24.7%, p<0.0001	36 months
EDSS score	**Change in EDSS score from baseline to follow-up (k)**	Comi et al., Lancet 2001, phase III (ETOMS study)	CIS (n=308)	IFN beta-1a 22mcg SC/week: median (IQR) 0 (-1 to 0); Placebo: median (IQR) 0 (-1 to 0), p=0.521	24 months
		Miller et al., Lancet Neurol 2014, phase III (TOPIC study)	CIS (n=618)	Teriflunomide 7mg: –0·250 (SD 0·937), p (vs. placebo) = 0.0334; Teriflunomide 14mg: –0·265 (SD 0·849), p (vs. placebo) = 0.0443; Placebo: –0·056 (SD 0·955)	108 weeks
	Time to 3-month CDP (g)	Miller et al., Lancet Neurol 2014, phase III (TOPIC study)	CIS (n=618)	Teriflunomide 7mg PO vs. Placebo: HR 0.978 (0.521–1.835), p=0.9953; Teriflunomide 14mg PO vs. placebo: HR 0.701 (0.360–1.366), p=0.4244	108 weeks
	% patients with 3-month CDP	Comi et al., Lancet 2001, phase III (ETOMS study)	CIS (n=308)	IFN beta-1a 22mcg SC/week: 15%; Placebo: 20%, p-value not specified (probably not	24 months

				significant) **(w)**	
		Miller et al., Lancet Neurol 2014, phase III (TOPIC study)	CIS (n=618)	Teriflunomide 7mg: 10%, p (vs. placebo) = 0.9953; Teriflunomide 14mg: 7%, p (vs. placebo) = 0.4244; Placebo: 10%	108 weeks
SNRS	**Change from baseline to FU**	Comi et al., Lancet 2001, phase III (ETOMS study)	CIS (n=308)	IFN beta-1a 22mcg SC/week: median (IQR) 0 (-1 to 2); Placebo: median (IQR) 0 (-1 to 2), p=0.747	24 months
Unidimensional Fatigue Impact Scale (FIS or UFIS)	**Change from baseline to FU**	Miller et al., Lancet Neurol 2014, phase III (TOPIC study)	CIS (n=618)	Teriflunomide 7mg: –2·730 (SD 30·410), p (vs. placebo) = 0.9974; Teriflunomide 14mg: –4·487 (SD 32·519), p (vs. placebo) = 0.8492; Placebo: –3·535 (29·298);	108 weeks

Table footnote:

(a) ARR refers to mean ARR per each group; it includes confirmed relapse rate, which includes rate of relapses with confirmed increase in EDSS (Voskuhl et al., Lancet Neurol 2016) and also adjusted mean relapse rate (Vollmer et al., J Neurol 2014)

(b) No detailed figures provided

(c) Cognitive impairment was defined on the number of failed tests, as mild (one to two tests failed) or moderate–severe (three or more tests failed)

(d) Defined as ≥7.5 points increase in MSIS-29

(e) CDP: Confirmed disability progression was defined as an increase of Expanded Disability Status Scale score of at least 1·0 point for patients with a baseline score of 1·0 or more, or an increase of at least 1·5 points for patients with a baseline score of 0, confirmed after 12 weeks. For the rest, EDSS increase of ≥1 point if EDSS ≤5.5; EDSS increase of ≥0.5 point if EDSS > 5.5;

(f) Includes adjusted MSFC z-score; also it may include values obtained at an early termination time point if this occurred after 12 months.

(g) Includes time to sustained accumulation of disability, which is considered as increase in 1 point in EDSS sustained for a minimum of 12 weeks (Confavreux et al., Lancet Neurol 2014, TOWER trial)

(h) No MRI activity includes: no new/enlarging lesions and no gadolinium-enhancing lesions

(i) Includes relapses requiring hospitalization/IV steroids (Comi et al., NEJM 2012, ALLEGRO study)

(j) Adjusting for baseline values of MSFC z-score, ANCOVA model

(k) Mean change reported, unless otherwise specified

(l) It includes 'at least 1 major relapse'
(m) The authors also estimated the proportion of patients with: i) at least one MS-related admission to hospital; ii) at least 1 MS-rekated steroid course
(n) The results shown refer to the comparative phase (0-12m) of the trial, where half of the patients were receiving IFN beta-1a IM 30mcg/week and the other half IFN beta-1a SC 44mcg tiw.
(o) p-value not specified
(p) this analysis refers to disability progression in both hands
(q) worsening in 9HPT is defined as deterioration greater or equal to 20%
(r) confirmed at 2 months
(s) mean number of relapses per patient during the trial/2 years (duration of trial)
(t) defined as 2-step increase (sustained for 3 months)
(u) in this context, this outcome measure (risk ratio or odds ratio) is equivalent to hazard ratio in the survival model
(v) timing for CDP not specified. Assumed 3 months
(w) this study looked at disability progression at the end of FU, so it is possible that just progression confirmed at just 3 months is also included here
(x) This refers to McDonald 2005 criteria

Abbreviations. BD: twice per day; CDP: confirmed disability progression; CI: confidence interval; eod: every other day; FU: follow-up; GA: glatiramer acetate; HR: hazard ratio; IA & AHSCT: immunoablation and autologous haemopoietic stem-cell transplantation; IFN: interferon; IQR: interquartile range; MIU: million international units; MSCT: mesenchymal stem cell transplantation; MSFC: Multiple Sclerosis Functional Composite; MSIS-29: Multiple Sclerosis Impact Scale – 29 items; PO: per oral; RFSS: Regional Functional System Score; SC: subcutaneous; SF-36: Short Form 36 Health Survey (SF-36); SNRS: Scripps Neurological Rating Scale; TDS: three times per day; tiw: three times in a week; TSQM: Treatment Satisfaction Questionnaire for Medication, with domains for Effectiveness, Side-Effects, Convenience and Global Satisfaction

Table 3: Clinical outcome measures in phase III trials in progressive MS

Original clinical outcome	Derived outcome measures	Trial	Condition (no. of patients randomised)	Drug, effect (vs. placebo/ another active arm)	Duration of the trial
Relapses	**Mean annualised relapse rate (a)**	European Study Group on IFN beta-1b in SPMS, Lancet 1998, phase III (EUSPMS study)	SPMS (n=718)	IFN beta-1b 8 million IU eod: 0.44; Placebo: 0.64, p=0.0002	Early termination: obvious superiority of IFN vs. placebo (initially planned: 39 months)
		SPECTRIMS study group, Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 22mcg SC tiw: 0.50 (0.44 to 0.56), p (vs. placebo) <0.001; IFN beta-1a 44mcg SC tiw: 0.50 (0.45 to 0.56), p (vs. Placebo) <0.001; Placebo: 0.71 (0.65 to 0.78)	36 months
		Andersen et al., JNNP 2004, phase III (The Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week: 0.25; Placebo: 0.27, p=0.55	36 months
		Hommes et al., Lancet 2004, phase III (ESIMS study)	SPMS (n=318)	IVIG 1g/Kg/month: 0.46 Placebo: 0.46, p>0.05	24 months
		North American Study Group on IFN beta-1b in SPMS, Neurology 2004, phase III (NASPMS study)	SPMS (n=939)	Pooled IFN beta-1b (250mcg SC eod or 160mcg/m2 SC eod) vs. placebo: reduction of ARR in 36%, p<0.05	Early termination for futility (initially planned: 36 months)
		Freedman et al., Neurology 2011, phase III (MAESTRO study)	SPMS (n=612)	MBP8298 500mg IV/6 months: 0.13; Placebo: 0.14, p=0.633	24 months
	Mean annualised severe relapse rate (i)	SPECTRIMS study group, Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 22mcg SC tiw: 0.26 (0.22 to 0.31), p (vs. placebo) = 0.002; IFN beta-1a	36 months

				44mcg SC tiw: 0.27 (0.23 to 0.31), p (vs. placebo) = 0.003; Placebo: 0.39 (0.34 to 0.44);	
	% patients with at least 1 relapse (a) (l)	Hommes et al., Lancet 2004, phase III, (ESIMS study)	SPMS (n=318)	IVIG 1g/Kg/month: 48.4%, p=0.58 Placebo: 52.2%	24 months
	% relapse-free patients at the end of FU	Andersen et al., JNNP 2004, phase III (The Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week: 61%; Placebo: 62%, p=0.89	36 months
		North American Study Group on IFN beta-1b in SPMS, Neurology 2004, phase III (NASPMS study)	SPMS (n=939)	IFN beta-1b 250mcg SC eod: 71%, p (vs. Placebo) =0.018; Placebo: 62%	Early termination for futility (initially planned: 36 months)
	Time to first confirmed relapse	European Study Group on IFN beta-1b in SPMS, Lancet 1998, phase III (EUSPMS study)	SPMS (n=718)	Median time: Placebo: 403 days; IFN beta-1b 8 million IU eod: 644 days, p=0.0083	39 months
		SPECTRIMS study group, Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 22mcg SC tiw vs. placebo: HR = 0.87 (0.69 to 1.10), p=0.237; IFN beta-1a 44mcg SC tiw vs. placebo: HR 0.77 (0.61 to 0.98), p=0.034;	36 months
		Andersen et al., JNNP 2004, phase III (The Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week vs. placebo: no differences (not specified)	36 months
		North American Study Group on IFN beta-1b in SPMS, Neurology 2004, phase III (NASPMS study)	SPMS (n=939)	Placebo: 487 days (30^{th} percentile) IFN beta-1b 250mcg SC eod: 1051 days (30^{th} percentile), p=0.010	Early termination for futility (initially planned: 36 months)
	Time between first and second relapse	SPECTRIMS study group, Neurology 2001 (SPECTRIMS	SPMS (n=618)	IFN beta-1a 22mcg SC tiw vs. placebo: HR =	36 months

		study)		0.50 (0.37 to 0.69), p < 0.001; IFN beta-1a 44mcg SC tiw vs. placebo: HR = 0.60 (0.44 to 0.81), p = 0.001;	
	Mean annualised hospitalisation rate due to MS exacerbations	SPECTRIMS study group, Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 22mcg SC tiw: 0.14 (0.11 to 0.17), p (vs. placebo) = 0.006; IFN beta-1a 44mcg SC tiw: 0.15 (0.12 to 0.18), p (vs. placebo) = 0.005; Placebo: 0.22 (0.18 to 0.26);	36 months
	Mean annualised rate of steroid courses	SPECTRIMS study group, Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 22mcg SC tiw: 0.31 (0.27 to 0.36), p (vs. placebo) = 0.001; IFN beta-1a 44mcg SC tiw: 0.34 (0.30 to 0.39), p (vs. placebo) = 0.006; Placebo: 0.52 (0.46 to 0.58);	36 months
EDSS score	**Score at FU**	European Study Group on IFN beta-1b in SPMS, Lancet 1998, phase III (EUSPMS study)	SPMS (n=718)	IFN beta-1b 8 million IU eod: 5.57; placebo: 5.84, p=0.0750	39 months
	Change in EDSS score from baseline to follow-up (k)	European Study Group on IFN beta-1b in SPMS, Lancet 1998, phase III (EUSPMS study)	SPMS (n=718)	IFN beta-1b 8 million IU eod: 0.47; Placebo: 0.60, p=0.0299	39 months
		Cohen et al., Neurology 2002 (IMPACT study)	SPMS (n=436)	IFN beta-1a 60mcg/week IM vs. placebo: mean change 0.258 vs. 0.272, respectively, p=0.362	24 months
		Andersen et al., JNNP 2004, phase III (The Nordic	SPMS (n=371)	IFN beta-1a SC 22mcg/week vs. placebo: no	36 months

		SPMS study)		differences (no further details given)	
		Hommes et al., Lancet 2004, phase III, (ESIMS study)	SPMS (n=318)	IVIG 1g/Kg/month: median change (range): 0.5 (-3.0 to 5.0); Placebo: 0.5 (-3.0 to 5.0), p>0.05	24 months
		North American Study Group on IFN beta-1b in SPMS, Neurology 2004, phase III (NASPMS study)	SPMS (n=939)	Pooled IFN beta-1b (250mcg SC eod or 160mcg/m2 SC eod) vs. placebo: no difference (no further details given)	Early termination for futility (initially planned: 36 months)
		Wolinsky et al., Ann Neurol 2007, phase III (PROMiSe study)	PPMS (n=943)	GA 20mg SC/day: 0.58 (SD 1.00); Placebo: HR (95% CI): 0.61 (SD 1.13), p>0.05	Early termination for futility (initially planned: 36 months)
		Hawker et al., Ann Neurol 2009, phase II/3 (OLYMPUS study)	PPMS (n=439)	Rituximab 1000mg IV/24 weeks: 0.33 (1.0); Placebo: 0.45 (SD 1.0), p=0.34	96 weeks
		Freedman et al., Neurology 2011, phase III (MAESTRO study)	SPMS (n=612)	MBP8298 500mg IV/6 months: 0.22 (SE 0.06); Placebo: 0.17 (SE 0.06), p=0.465	24 months
	Time to EDSS 7.0	European Study Group on IFN beta-1b in SPMS, Lancet 1998, phase III (EUSPMS study)	SPMS (n=718)	IFN beta-1b 8 million IU eod vs. placebo: OR (95% CI) 0.66 (0.47 to 0.93), p=0.0133	39 months
	Time to 3-month CDP (g)	European Study Group on IFN beta-1b in SPMS, Lancet 1998, phase III (EUSPMS study)	SPMS (n=718)	IFN beta-1b 8 million IU eod vs. placebo: odds ratio of 0.65 (95% CI 0.52–0.83), p =0.0008 **(u)**	39 months
		SPECTRIMS study group, Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 22mcg SC tiw vs. placebo: HR 0.88, p = 0.305; IFN beta-1a 44mcg SC tiw vs. placebo: HR (95% CI) 0.83	36 months

				(0.65 to 1.07), p=0.146	
		Cohen et al., Neurology 2002 (IMPACT study)	SPMS (n=436)	IFN beta-1a 60mcg/week IM vs. placebo: HR (95% CI): 0.977 (0.679 to 1.407), p=0.90	24 months
		Hommes et al., Lancet 2004, phase III, (ESIMS study)	SPMS (n=318)	IVIG 1g/Kg/month vs. placebo: HR (95% CI) 1.11(0.80 to 1.53), p=0.53;	24 months
		Wolinsky et al., Ann Neurol 2007, phase III (PROMiSe study)	PPMS (n=943)	GA 20mg SC/day vs. placebo: HR (95% CI) 0.87 (0.71 to 1.07), p=0.1753	36 months
		Hawker et al., Ann Neurol 2009, phase II/3 (OLYMPUS study)	PPMS (n=439)	Rituximab 1000mg IV/24 weeks vs. placebo: HR (95% CI) 0.77 (0.55 to 1.09), p=0.1442	96 weeks
		Lublin et al., Lancet 2016, INFORMS study, phase III	PPMS (n=970)	Fingolimod 0.5mg/d vs. placebo: HR (95% CI) 0.88 (0.71 to 1.08), p=0.217	36 months
		Montalban et al., N Engl J Med. 2016 (ORATORIO study)	PPMS (n=732)	Ocrelizumab 600mg (300mg x2) /24 weeks IV vs. placebo: HR=0.76; p=0.0321	120 weeks
	% patients with 3-month CDP	European Study Group on IFN beta-1b in SPMS, Lancet 1998, phase III (EUSPMS study)	SPMS (n=718)	IFN beta-1b 8 million IU eod: 38.9%; Placebo: 49.7%, p =0.0048	39 months
		SPECTRIMS study group, Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 22mcg SC tiw: no differences vs. placebo (no more details reported); IFN beta-1a 44mcg SC tiw: no differences vs. placebo (no more details reported)	36 months
		Hommes et al., Lancet 2004,	SPMS (n=318)	IVIG 1g/Kg/month:	24 months

		phase III (ESIMS study)		48.4%; Placebo: 44%, p=0.53	
		Wolinsky et al., Ann Neurol 2007, phase III (PROMiSe study)	PPMS (n=943)	GA 20mg SC/day: 39.6%; Placebo: 45.2%, p>0.05	36 months
		Hawker et al., Ann Neurol 2009, phase II/3 (OLYMPUS study)	PPMS (n=439)	Placebo: 38.5%; Rituximab 1000mg IV/24 weeks: 30.2%, p=0.1442	96 weeks
		Lublin et al., Lancet 2016, INFORMS study, phase III	PPMS (n=970)	Fingolimod 0.5mg/d vs. placebo: 54.3% (47.16–61.45) vs. 58.7% (53.30–64.18), p>0.05	36 months
	Time to 6-month CDP	Andersen et al., JNNP 2004, phase III (The Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week vs. placebo: HR (95% CI) 1.13 (0.82 to 1.57), p=0.45	36 months
		North American Study Group on IFN beta-1b in SPMS, Neurology 2004, phase III (NASPMS study)	SPMS (n=939)	Pooled IFN beta-1b (250mcg SC eod or 160mcg/m2 SC eod) vs. placebo: no difference, p=0.712	Early termination for futility (initially planned: 36 months)
		Freedman et al., Neurology 2011, phase III (MAESTRO study)	SPMS (n=612)	MBP8298 500mg IV/6 months vs. placebo: HRs not reported, but not significant	24 months
		Zajicek et al., Lancet Neurol 2013, phase unspecified (CUPID study)	PPMS (n=191), SPMS (n=302) (randomised: n=498)	Dronabinol (max. dose: 28mg/day, titrated against bodyweight) vs. placebo: HR (95% CI) 0.92 (0.68 to 1.23), p=0.57	36 months
		Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg/d vs. placebo: HR (95% CI) no different from 1.0 (p>0.05, data not shown, no further details given)	36 months
		Montalban et al., N Engl J Med.	PPMS (n=732)	Ocrelizumab 600mg (300mg	120 weeks

		2016 (ORATORIO study)		x2) /24 weeks IV vs. placebo: HR= 0.75; p=0.0365	
	% patients with 6-month CDP	Andersen et al., JNNP 2004, phase III (The Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week: 41%; Placebo: 38%, p=0.45	36 months
		Hawker et al., Ann Neurol 2009, phase II/3 (OLYMPUS study)	PPMS (n=439)	Rituximab 1000mg IV/24 weeks: 27.3%; Placebo: 30.4%, p=0.59	96 weeks
		Freedman et al., Neurology 2011, phase III (MAESTRO study)	SPMS (n=612)	MBP8298 500mg IV/6 months: 30.7%; Placebo: 27.8%, p=0.527 (in patients DR2+ or DR4+)	24 months
		Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg/d vs. placebo: similar percentages (p>0.05, data not shown, no further details given)	36 months
	IDSS: Integrated Disability Status Score (IDSS, defined by area under an EDSS time-curve adjusted for baseline	SPECTRIMS study group, Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 22mcg SC tiw: no differences vs. placebo (no more details reported); IFN beta-1a 44mcg SC tiw: no differences vs. placebo (no more details reported)	36 months
TWT z-score	**Change in TWT z-score from baseline to FU**	Cohen et al., Neurology 2002 (IMPACT study)	SPMS (n=436)	IFN beta-1a 60mcg/week IM vs. placebo (SD): 0.979 (2.62) vs. 1.191 (3.13), p=0.378	24 months
		Hawker et al., Ann Neurol 2009, phase II/3 (OLYMPUS study)	PPMS (n=439)	Rituximab 1000mg IV/24 weeks: (median) -0.08; Placebo: (median) -0.14 (greater worsening than rituximab arm), p=0.015	96 weeks

		Freedman et al., Neurology 2011, phase III (MAESTRO study)	SPMS (n=612)	MBP8298 500mg IV/6 months: 0.99; Placebo: 1.57, p=0.096	24 months
	Time to 3-month CDP	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg PO/day vs. placebo: HR (95% CI) 0.94 (0.78 to 1.14), p=0.546;	36 months
	% patients with 3-month CDP	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg/d: 62.9% (57.10 to 68.62) Placebo: 70.0% (61.78 to 78.21), p=0.546	36 months
		Montalban et al., N Engl J Med. 2016 (ORATORIO study)	PPMS (n=732)	Ocrelizumab 600mg (300mg x2) /24 weeks IV vs. placebo: 39% vs. 55%, p=0.0404	120 weeks
	Time to 6-month CDP	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg PO/day vs. placebo: similar to 3-month CDP analysis (no further details given)	36 months
	% patients with 6-month CDP	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg PO/day vs. placebo: similar to 3-month CDP analysis (no further details given)	36 months
9HPT z-score	**Change in 9HPT z-score from baseline to FU**	Cohen et al., Neurology 2002 (IMPACT study)	SPMS (n=436)	IFN beta-1a 60mcg/week IM vs. placebo: 0.202 (SD 0.476) vs. 0.290 (SD 0.494), p=0.024	24 months
		Freedman et al., Neurology 2011, phase III (MAESTRO study)	SPMS (n=612)	MBP8298 500mg IV/6 months: -0.08; Placebo: -0.04, p=0.537	24 months
	Time to 3-month CDP	Hommes et al., Lancet 2004, phase III, (ESIMS study) **(q)**	SPMS (n=318)	IVIG 1g/Kg/month vs. placebo: HR (95% CI) 1.09 (0.75 to 1.59), p=0.67	24 months
		Lublin et al.,	PPMS (n=970)	Fingolimod 0.5mg	36 months

		Lancet 2016 (INFORMS study)		PO/day vs. placebo: HR (95% CI) 0·93 (0·71–1·22), p=0·607;	
	% patients with 3-month CDP	Hommes et al., Lancet 2004, phase III, (ESIMS study) **(q)**	SPMS (n=318)	IVIG 1g/Kg/month vs. placebo: 34.6%; Placebo: 33.3%, p=0.67 **(p)**	24 months
		Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg PO/day: 33.6% (26.11–41.08); Placebo: 41.3% (32.10–50.55), p= 0.607	36 months
	Time to 6-month CDP	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg PO/day vs. placebo: similar to 3-month CDP analysis (no further details given)	36 months
	% patients with 6-month CDP	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg PO/day vs. placebo: similar to 3-month CDP analysis (no further details given)	36 months
PASAT z-score	**Change from baseline to FU**	Cohen et al., Neurology 2002 (IMPACT study)	SPMS (n=436)	IFN beta-1a 60mcg/week IM vs. placebo: 0.094 (SD 0.498) vs. 0.004 (SD 0.473), p=0.061	24 months
		Freedman et al., Neurology 2011, phase III (MAESTRO study)	SPMS (n=612)	MBP8298 500mg IV/6 months: 0.24; Placebo: 0.17, p=0.393	24 months
MSFC	**Change in MSFC z-score from baseline to follow-up (f) (k)**	Cohen et al., Neurology 2002 (IMPACT study)	SPMS (n=436)	IFN beta-1a 60mcg/week IM vs. placebo: 0.362 (SD 1.41) vs. 0.495 (SD 1.58), p=0.033	24 months
		Wolinsky et al., Ann Neurol 2007, phase III (PROMiSe study)	PPMS (n=943)	GA 20mg SC/day vs. placebo: no differences between groups (no further details given)	36 months

		Hawker et al., Ann Neurol 2009, phase II/3 (OLYMPUS study)	PPMS (n=439)	Rituximab 1000mg IV/24 weeks: median change -0.06; Placebo: median change -0.10, p=0.089	96 weeks
		Freedman et al., Neurology 2011, phase III (MAESTRO study)	SPMS (n=612)	MBP8298 500mg IV/6 months: -0.28; Placebo: -0.46, p=0.137	24 months
		Zajicek et al., Lancet Neurol 2013, phase unspecified (CUPID study)	PPMS (n=191), SPMS (n=302) (randomised: n=498)	Dronabinol [max. dose: 28mg/day, titrated against bodyweight]: yearly change –0·17 (SD 0·28); Placebo: yearly change –0·16 (SD 0·30), p=0.72	36 months
RFSS	**Change from baseline to FU**	Andersen et al., JNNP 2004, phase III (The Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week vs. placebo: no differences (not specified)	36 months
	Time to an increase ≥ 2% in RFSS score	Andersen et al., JNNP 2004, phase III (The Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week vs. placebo: HR (95% CI) 0.93 (0.68 to 1.28), p=0.67	36 months
	% patients with an increase ≥ 2%	Andersen et al., JNNP 2004, phase III (The Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week: 44%; Placebo: 44%, p=0.45	36 months
Ambulation index	**Change from baseline to FU**	Andersen et al., JNNP 2004, phase III (Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week vs. placebo: no differences (not specified)	36 months
Arm index	**Change from baseline to FU**	Andersen et al., JNNP 2004, phase III (The Nordic SPMS study)	SPMS (n=371)	IFN beta-1a SC 22mcg/week vs. placebo: no differences (not specified)	36 months
Rao's Brief Repeatable Battery	**% patients with change in cognitive impairment (c)**	North American Study Group on IFN beta-1b in SPMS, Neurology 2004, phase III	SPMS (n=939)	Pooled IFN beta-1b (250mcg SC eod or 160mcg/m2 SC eod) vs. placebo:	Early termination for futility (initially planned: 36 months)

		(NASPMS study)		no difference (not specified)	
Composite progressive disability score	**Time to CDP, defined as presence of at least 1 out of the 3: -Increase in EDSS (0.5 if EDSS≤5.5; 1.0 if EDSS >6.0) -Increase in ≥20% in 9HPT -Increase ≥20% in TWT**	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg PO/day: 62.9% (57.10–68.62); Placebo: 70.0% (61.78–78.21), p>0.05	36 months
	% patients with at least one of the three situations (confirmed at 3m): -Increase in EDSS (0.5 if EDSS≤5.5; 1.0 if EDSS >6.0) -Increase in ≥20% in 9HPT -Increase ≥20% in TWT	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg PO/day: 62.9% (57.10–68.62); Placebo: 70.0% (61.78–78.21), p>0.05	36 months
	Time to 3-month CDP, using EDSS or 9HPT (q)	Hommes et al., Lancet 2004, phase III, (ESIMS study)	SPMS (n=318)	IVIG 1g/Kg/month vs. placebo: HR (95% CI) 1.12 (0.84 to 1.49), p=0.44	24 months
	% of patients with 3-month CDP, using EDSS or 9HPT (q)	Hommes et al., Lancet 2004, phase III, (ESIMS study)	SPMS (n=318)	Placebo: 57.9% IVIG 1g/Kg/month vs. placebo: 61.6%, p=0.44	24 months
Multiple Sclerosis Walking Scale (MSWS-12)	**Change from baseline to FU**	Zajicek et al., Lancet Neurol 2013, phase unspecified (CUPID study)	PPMS (n=191), SPMS (n=302) (received treatment: n=493; randomised: n=498)	Dronabinol [max. dose: 28mg/day, titrated against bodyweight]: yearly change 0.37 (SD 2.33); Placebo: yearly change 0.52 (2.68); p=0.74	36 months
MSIS-29	**Change from baseline to FU (physical score)**	Zajicek et al., Lancet Neurol 2013, phase unspecified (CUPID study)	PPMS (n=191), SPMS (n=302) (received treatment: n=493; randomised: n=498)	Dronabinol [max. dose: 28mg/day, titrated against bodyweight]: yearly change 0.62 (SD 3.29); Placebo: yearly change 1.03 (SD	36 months

				3.74); p=0.11	

Table footnote:
(a) ARR refers to mean ARR per each group; it includes confirmed relapse rate, which includes rate of relapses with confirmed increase in EDSS (Voskuhl et al., Lancet Neurol 2016) and also adjusted mean relapse rate (Vollmer et al., J Neurol 2014)
(b) No detailed figures provided
(c) Cognitive impairment was defined on the number of failed tests, as mild (one to two tests failed) or moderate–severe (three or more tests failed)
(d) Defined as ≥7.5 points increase in MSIS-29
(e) CDP: Confirmed disability progression was defined as an increase of Expanded Disability Status Scale score of at least 1·0 point for patients with a baseline score of 1·0 or more, or an increase of at least 1·5 points for patients with a baseline score of 0, confirmed after 12 weeks. For the rest, EDSS increase of ≥1 point if EDSS ≤5.5; EDSS increase of ≥0.5 point if EDSS > 5.5;
(f) Includes adjusted MSFC z-score; also it may include values obtained at an early termination time point if this occurred after 12 months.
(g) Includes time to sustained accumulation of disability, which is considered as increase in 1 point in EDSS sustained for a minimum of 12 weeks (Confavreux et al., Lancet Neurol 2014, TOWER trial)
(h) No MRI activity includes: no new/enlarging lesions and no gadolinium-enhancing lesions
(i) Includes relapses requiring hospitalization/IV steroids (Comi et al., NEJM 2012, ALLEGRO study)
(j) Adjusting for baseline values of MSFC z-score, ANCOVA model
(k) Mean change reported, unless otherwise specified
(l) It includes 'at least 1 major relapse'
(m) The authors also estimated the proportion of patients with: i) at least one MS-related admission to hospital; ii) at least 1 MS-rekated steroid course
(n) The results shown refer to the comparative phase (0-12m) of the trial, where half of the patients were receiving IFN beta-1a IM 30mcg/week and the other half IFN beta-1a SC 44mcg tiw.
(o) p-value not specified
(p) this analysis refers to disability progression in both hands
(q) worsening in 9HPT is defined as deterioration greater or equal to 20%
(r) confirmed at 2 months
(s) mean number of relapses per patient during the trial/2 years (duration of trial)
(t) defined as 2-step increase (sustained for 3 months)
(u) in this context, this outcome measure (risk ratio or odds ratio) is equivalent to hazard ratio in the survival model
(v) timing for CDP not specified. Assumed 3 months
(w) this study looked at disability progression at the end of FU, so it is possible that just progression confirmed at just 3 months is also included here
(x) This refers to McDonald 2005 criteria

Abbreviations. BD: twice per day; CDP: confirmed disability progression; CI: confidence interval; eod: every other day; FU: follow-up; GA: glatiramer acetate; HR: hazard ratio; IA & AHSCT: immunoablation and autologous haemopoietic stem-cell transplantation; IFN: interferon; IQR: interquartile range; MIU: million international units; MSCT: mesenchymal stem cell transplantation; MSFC: Multiple Sclerosis Functional Composite; MSIS-29: Multiple Sclerosis Impact Scale – 29 items; PO: per oral; RFSS: Regional Functional System Score; SC: subcutaneous; SF-36: Short Form 36 Health Survey (SF-36); SNRS: Scripps Neurological Rating Scale; TDS: three times per day; tiw: three times in a week; TSQM: Treatment Satisfaction Questionnaire for Medication, with domains for Effectiveness, Side-Effects, Convenience and Global Satisfaction

Table 4: Brain MRI outcome measures in phase III trials in relapsing-remitting MS

Brain MRI
Inclusion criteria: controlled phase III clinical trials
Exclusion criteria: incomplete data presentation (e.g. missing values); descriptive findings in absence of any statistical analysis; secondary analyses of clinical trials and extension studies evaluating the same clinical endpoints of the main trial in population subgroups or during longer observation time.

Original neuroimaging outcome	Derived outcome measures	Trial	Condition (no. of patients randomised)	Drug, effect (vs. placebo/ another active arm)	Duration of the trial
T2 lesions	**Number of new lesions**	The Interferon beta Multiple Sclerosis Study group; Paty et al., Neurology 1993 (Interferon beta Multiple Sclerosis Study Group)	RRMS (n=372)	Interferon beta-1b vs. Placebo, median new lesion rate 0.5 vs. 2.0 (p=0.0026)	24 months
		Comi et al., Ann Neurol 2001 (European/Canadian Glatiramer Acetate Study)	RRMS (n=249)	Glatiramer Acetate vs. Placebo, number of lesions 9.4 vs. 13.7 (p<0.003) after 9 months	9 months
		Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, number of lesions 1.1 vs. 5.8 after 1 year (p<0.001), 0.7 vs. 4.4 after 2 years (p<0.001), and 1.8 vs. 10.2 overall (p<0.001)	24 months
		O'Connor et al., Lancet Neurol 2009 (BEYOND study)	RRMS (n=2244)	Interferon beta-1a 500µg vs. 250µg vs. Glatiramer acetate, number of lesions 3.3 vs. 3.3 vs. 4.6 after 2 years (p=0.25; p=0.0009; p=0.011)	24 months
		Comi et al., Ann Neurol 2011 (FORTE study)	RRMS (n=980)	Glatiramer Acetate 20mg vs. 40mg, number of lesions 2.87 vs. 2.72 (ns) after 12 months	12 months
	Number of enlarging lesions	Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, number of lesions 0.1 vs. 0.4 after 1 year (p<0.001), 0.0 vs. 0.4 after 2 years (p<0.001), and 0.1 vs. 0.8 overall (p<0.001)	24 months
	Number of new or enlarging lesions	PRISMS Study Group, Lancet 1998; Li et al., Ann Neurol 1999 (PRISMS study)	RRMS (n=560)	Interferon beta-1a 44µg vs. 22µg vs. Placebo, percent difference compared to Placebo -67% and -78% (p<0.0001) after 2 years; median number of lesions per patient per scan 0.5 vs. -	24 months

				0.75 vs. 2.25 (p=0.0003; p<0.0001; p<0.0001) after 6 months; percent of scans with lesions 25% vs. 50% vs. 75% (p=0.0002; p<0.0001; p<0.0001) after 6 months; and percent of patients without lesions 31% vs. 19% vs. 8% (p=0.0009; p<0.0001; p<0.0001) after 6 months	
		Jacobs et al., New Eng J Med 2000 (CHAMPS study)	CIS (n=383)	Interferon beta-1a 30µg vs. Placebo, number of lesions 1.5 vs. 2.8 after 6 months (p=0.01), 2.1 vs. 4.0 after 12 months (p<0.001), 2.1 vs. 5.0 after 18 months (p<0.001)	Early termination: obvious superiority of IFNb over placebo (initially planned: 36 months)
		Comi et al., Lancet 2001 (ETOMS study)	CIS (n=309)	Interferon beta-1a 22µg vs. Placebo, median number of lesions per patient per scan 2.0 vs. 3.0 after 2 years (p<0.001)	24 months
		Panitch et al., Neurology 2002; Panitch et al., J Neurol Sci. 2005 (EVIDENCE study)	RRMS (n=677)	Interferon beta-1a 44µg vs. 30µg, number of lesions 0.9 vs. 1.4 (p<0.001), percent of scans with lesions 27% vs. 44% (p<0.001), percent of patients with no lesions 58% vs. 38% (p<0.001) after 16 months	24 months (0-12m: comparative phase; 12-24m: cross-over phase)
		Polman et al., New Eng J Neurol 2006 Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, number of lesions 1.2 vs. 6.1 after 1 year (p<0.001), 0.7 vs. 4.9 after 2 years (p<0.001), and 1.9 vs. 11.0 overall (p<0.001)	24 months
		Rudick et al., New Eng J Med 2006 (SENTINEL study)	RRMS (n=1171)	Natalizumab+Interferon beta-1a vs. Interferon beta-1a, number of lesions 0.9 vs. 5.4 after 2 years (p<0.001)	24 months
		Mikol et al., Lancet Neurol 2008 (REGARD study)	RRMS (n=764)	Interferon beta-1[a] 44 µg vs. Glatiramer acetate 20 mg, lesions per patient per scan 0.67 vs. 0.82 after 96 weeks (p=0.18); proportion of scans per patient with lesions 24.6% vs. 26.3% after 96 weeks (p=0.34); patients	96 weeks

				with no lesions 40% vs. 37% after 96 weeks (p=0.51)	
		Cohen et al., New Eng J Med 2010 (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 1.25mg and 0.5mg vs. Interferon beta-1a (30µg/week), number of lesions 1.5 (p<0.001) and 1.7 (p=0.004), vs. 2.6 after 12 months; percent of patients free of lesions 48.0% (p=0.37) and 54.8% (p=0.01), vs. 45.7% after 12 months	12 months
		Kappos et al., New Eng J Med 2010; Radue et al., Arch Neurol 2012 (FREEDOMS study)	RRMS (n=1272)	Fingolimod 1.25mg and 0.5mg vs. Placebo, number of lesions 1.1 (p<0.001) and 1.0 (p<0.011), vs. 3.6 after 6 months, 1.5 (p<0.001) and 1.6 (p<0.011), vs. 5.5 after 12 months, 1.1 (p<0.001) and 0.9 (p<0.011), vs. 4.3 between 13 and 24 months, 2.5 (p<0.001) and 2.5 (p<0.011), vs. 9.8 after 24 months; percent of patients lesion-free 58.7% (p<0.001) and 57.4% (p<0.001) vs. 26.4% after 12 months, 69.8% (p<0.001) and 72.8% (p<0.001) vs. 33.2% between 12 and 24 months, and 51.9% (p<0.001) and 50.5% (p<0.001) vs. 21.2% after 24 months	24 months
		Giovannoni et al., Lancet Neurol 2011; Comi et al., J Neurol 2013 (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/kg and Cladribine 5.25mg/kg vs. Placebo, proportion of patients lesion-free 61.8% (p<0.001) and 62.8% (p<0.001), vs. 27.6% after 96 weeks; relative reduction 73.4% (p<0.001) and 76.9% (p<0.001) after 96 weeks	96 weeks
		O'Connor et al., New Eng J Med 2011; Wolinsky et al., Mult Scler 2013 (TEMSO study)	RMS (1088)	Teriflunomide 14mg and 7mg vs. Placebo, mean difference from Placebo -0.089 (p=0.0003) and -0.053 (p=0.0317) after 108 weeks	108 weeks
		Sorensen et al., Lancet Neurol 2011	RRMS (n=307)	Interferon beta-1[a] 30 µg with vs. without	12 months

		(SIMCOMBIN study)		Simvastatin 80 mg, mean number of lesions 2.96 vs. 2.52 after 12 months (ns)	
		Cohen et al., Lancet 2012 (CARE-MS I)	RRMS (n=581)	Alemtuzumab 12mg vs. Interferon beta-1a 44 µg, proportion of patients with lesions 48% vs. 58% after 2 years (p=0.04)	24 months
		Coles et al., Lancet 2012 (CARE-MS II)	RRMS (n=840)	Alemtuzumab 12mg vs. Interferon beta-1a 44 µg, proportion of patients with lesions 46% vs. 68% after 2 years (p<0.0001)	24 months
		Comi et al., New Eng J Med. 2012 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, cumulative number of lesions 5.03 vs. 7.14 (p<0.001) at 12 and 24 months	24 months
		Fox et al., New Eng J Med. 2012 (CONFIRM study)	RRMS (n=682, MRI cohort)	Dimethyl Fumarate 240mg BID or TID or Glatiramer acetate vs. Placebo, number of lesions 5.1 (p<0.001), 4.7 (p<0.001), 8.0 (p<0.001), vs. 17.4 after 2 years	24 months
		Gold et al., New Eng J Med 2012; Arnold et al., J Neurol 2014 (DEFINE study)	RRMS (n=1234)	Dimethyl Fumarate 240mg BID and TID vs. Placebo, number of lesions 2.6 (p=0.01) and 4.4 (p=0.01) vs. 17.6 after 96 weeks; in a sub-cohort of 540 patients, 1.1 (p<0.0001) and 1.6 (p<0.0001) vs. 5.2 after 6 months, 1.6 (p<0.0001) and 2.6 (p<0.0001) vs. 10.3 after 1 year, and 2.6 (p<0.0001) and 4.4 (p<0.0001) vs. 17.0 after 2 years	24 months
		Khan et al., Ann Neurol 2013; Zivadinov et al., J Neurol 2015 (GALA study)	RRMS (n=1404)	Glatiramer acetate 40 mg vs. Placebo, cumulative number of lesions 3.650 vs. 5.592 after 6 and 12 months (p<0.0001)	12 months
		Calabresi et al., Lancet Neurol 2014; Arnold et al., BMC Neurol 2014 (ADVANCE study)	RRMS (n=1512)	Peginterferon beta-1a every 4 vs. 2 weeks vs. Placebo, number of lesions 4.6 vs. 2.2 vs. 5.8 (p<0.0001; p<0.0001; p=0.023) after 24 weeks, and 7.9 vs. 3.6 vs. 10.9 (p<0.0001; p<0.0001; p=0.0008) after 48 weeks	24 months
		Calabresi et al., Lancet Neurol 2014	RRMS (n=1083)	Fingolimod 1.25mg and 0.5mg vs. Placebo,	24 months

		(FREEDOMS II study)		number of lesions 1.6 (p<0.001) and 2.3 (p<0.001), vs. 8.9 after 24 months; percent of patients free of lesions 63% (p<0.001) and 50% (p<0.001), vs. 26% after 24 months	
		Massacesi et al., PloS One 2014 (EudraCT 2006-004937-13)	RRMS (n=150)	Azathioprine (3mg/kg/day) vs. Interferon, annualised number of lesions 0.76 vs. 0.69 after 2 years (p=0.75); and number of patients with new lesions (0, 1-2, ≥3) 27/11/12 vs. 21/18/8 after 2 years (p=0.41)	24 months
		Vollmer et al., J Neurol 2014 (BRAVO)	RRMS (n=1331)	Laquinimod or Interferon beta-1a 30 µg vs. Placebo, cumulative number of lesions 10.88 (p=0.078) or 6.37 (p<0.001) vs. 13.03 after 12 an 24 months	24 months
		Kappos et al., New Eng J Med 2015 (DECIDE study)	RRMS (n=1841)	Daclizumab vs. Interferon, number of lesions 2.14 vs. 3.81 (p<0.001) after 24 weeks; 4.3 vs. 9.4 (p<0.001) after 96 weeks	144 weeks
		Miller et al., Neurology 2015 (CONFIRM study)	RRMS (n=681)	Dimethyl Fumarate 240mg BID and TID vs. Glatiramer Acetate vs. Placebo, number of lesions 3.1 (p<0.0001), 2.8 (p<0.0001), and 4.6 (p<0.0001) vs. 9.5 after 1 year, 2.0 (p<0.0001), 1.9 (p<0.0001), and 3.4 (p<0.0001) vs. 8.0 between 1 and 2 years, and 5.1 (p<0.0001), 4.7 (p<0.0001), and 8.0 (p<0.0001) vs. 17.4 after 2 years	24 months
	Volume of T2 lesions	The Interferon beta Multiple Sclerosis Study group; Paty et al., Neurology 1993 (Interferon beta Multiple Sclerosis Study Group)	RRMS (n=327)	Interferon beta-1b vs. Placebo, median percent volume change -6.2% vs. 10.9% after 1 year (p<0.001), -0.9% vs. 16.5% after 2 years (p<0.001), -9.3% vs. 15.0 after 3 years (p=0.002)	24 months
		Jacobs et al., Ann Neurol 1996 (MSCRG study)	RRMS (n=300)	Interferon beta-1a 30µg vs. Placebo, median percent volume change -13.1% vs. -3.3% after 1	104 weeks

				year (P=0.02), and -13.2% vs. -6.5% after 2 years (p=0.36)	
		PRISMS Study Group, Lancet 1998; Li et al., Ann Neurol 1999 (PRISMS study)	RRMS (n=533)	Interferon beta-1a 44µg vs. 22µg vs. Placebo, median percent volume change -4.2% vs. -1.5% vs. 4.0% (p=0.0246; p=0.0001; p=0.0001) after 6 months, -4.5% vs. -3.5% vs. 6.4% (p=0.3809; p=0.0001; p=0.0001) after 12 months, -3.1% vs. -1.4% vs. 10.8% (p=0.0974; p=0.0001; p=0.0001) after 18 months, and -3.8% vs. -1.2% vs. 10.9% (p=0.0537; p=0.0001; p=0.0001) after 24 months	24 months
		Comi et al., Ann Neurol 2001 (European/Canadian Glatiramer Acetate Study)	RRMS (n=249)	Glatiramer Acetate vs. Placebo, volume change 3.0mL vs. 4.7mL (p=0.006) after 9 months	9 months
		Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, lesion volume 14303.7mm^3 vs. 15703.2mm^3 after 1 year (p=0.016), 14722.0mm^3 vs. 17853.1mm^3 lesions after 2 years (p<0.001), and 14722.0mm^3 vs. 17853.0mm^3 lesions overall (p<0.001)	24 months
		Mikol et al., Lancet Neurol 2008 (REGARD study)	RRMS (n=764)	Interferon beta-1[a] 44 µg vs. Glatiramer acetate 20 mg, volume change -2416.9mm^3 vs. -1583.5mm^3 after 96 weeks (p=0.26)	96 weeks
		O'Connor et al., Lancet Neurol 2009 (BEYOND study)	RRMS (n=2244)	Interferon beta-1a 500µg vs. 250µg vs. Glatiramer acetate, percent volume change 22.0% vs. 19.0% vs. 25.0% after 2 years (p=0.56; p=0.0008; p=0.0001)	24 months
		Cohen et al., New Eng J Med 2010 (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 1.25mg and 0.5mg vs. Interferon beta-1a (30µg/week), percent volume change 6.7% (p=0.48) and 9.9% (p=0.63), vs. 10.4% after 12 months	12 months
		Kappos et al., New	RRMS	Fingolimod 1.25mg and	24 months

		Eng J Med 2010; Radue et al., Arch Neurol 2012 (FREEDOMS study)	(n=1272)	0.5mg vs. Placebo, percent volume change 2.7% (p<0.001) and 3.4% (p<0.001), vs. 18.7% after 12 months, 1.6% (p<0.001) and 10.6% (p<0.001), vs. 33.8% after 24 months	
		Sorensen et al., Lancet Neurol 2011 (SIMCOMBIN study)	RRMS (n=307)	Interferon beta-1[a] 30 µg with vs. without Simvastatin 80 mg, volume change 0.033mL vs. 0.095mL after 12 months (p=0.612)	12 months
		O'Connor et al., New Eng J Med 2011; Wolinsky et al., Mult Scler 2013 (TEMSO study)	RMS (1088)	Teriflunomide 14mg and 7mg vs. Placebo, volume change 0.39mL (p<0.0001) and 0.81mL (p=0.04) vs. 1.67mL after 108 weeks	108 weeks
		Giovannoni et al., Lancet Neurol 2011; Comi et al., J Neurol 2013 (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/kg and Cladribine 5.25mg/kg vs. Placebo, relative reduction 24.0% (p<0.001) and 41.2% (p<0.001) after 96 weeks	96 weeks
		Cohen et al., Lancet 2012 (CARE-MS I)	RRMS (n=581)	Alemtuzumab 12mg vs. Interferon beta-1a 44 µg, median percent volume change -9.3% vs. -6.5% after 2 years (p=0.31)	24 months
		Coles et al., Lancet 2012 (CARE-MS II)	RRMS (n=840)	Alemtuzumab 12mg vs. Interferon beta-1a 44 µg, median percent volume change -1.27% vs. -1.23% after 2 years (p=0.14)	24 months
		Gold et al., New Eng J Med 2012; Arnold et al., J Neurol 2014 (DEFINE study)	RRMS (n=1234)	Dimethyl Fumarate 240mg BID and TID vs. Placebo, in a sub-cohort of 540 patients, median percent volume change -3.5% (p<0.001) and -1.7% (p<0.01) vs. 1.6% after 6 months, -5.8% (p<0.0001) and -3.7% (p<0.0001) vs. 6.5% after 1 year, and -6.2% (p<0.0001) and -1.9% (p<0.0001) vs. 20.1% after 2 years	24 months
		Lublin et al., Ann Neurol 2013 (CombiRx study)	RRMS (n=1008)	IFN beta-1a 30mcg SC/week + GA 20mg SC/day vs IFN beta-1a 30mcg SC/week vs GA 20mg SC/day: volume change -1.38mL vs. -0.25mL vs. 0.01mL	36 months

				(p=0.008; p=0.48) after 36 months	
		Calabresi et al., Lancet Neurol 2014 (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 1.25mg and 0.5mg vs. Placebo, median percent volume change -7.69% (p<0.001) and 13.74% (p<0.001), vs. 25.06% after 24 months	24 months
		Calabresi et al., Lancet Neurol 2014; Arnold et al., BMC Neurol 2014 (ADVANCE study)	RRMS (n=1512)	Peginterferon beta-1a every 4 and 2 weeks vs. Placebo, volume change 0.14cm^3 (p=0.0006) and -0.22cm^3 (p<0.0001) vs. 0.34cm^3 after 24 weeks, and 0.06cm^3 (p<0.0001) and -0.26cm^3 (p<0.0001) vs. 0.77cm^3 after 48 weeks	24 months
		Kappos et al., New Eng J Med 2015 (DECIDE study)	RRMS (n=1841)	Daclizumab vs. Interferon, median percent volume change -1.4% vs. 3.4% (p=0.02) after 24 weeks; 0.2% vs. 8.6% (p<0.001) after 96 weeks; volume of new or newly enlarged T2 lesions 217.0mm^3 vs. 463.1mm^3 (p<0.001) after 24 weeks, and 225.7mm^3 vs. 556.8mm^3 (p<0.001) after 96 weeks	144 weeks
		Miller et al., Neurology 2015 (CONFIRM study)	RRMS (n=681)	Dimethyl Fumarate 240mg BID and TID vs. Glatiramer Acetate vs. Placebo, median percent volume change -4.2% (p<0.0001), -0.3% (p<0.0001), and -3.4% (p<0.0001) vs. 4.8% after 1 year, and -7.4% (p<0.0001), -1.5% (p<0.0001), and -6.3% (p<0.0001) vs. 14.6% after 2 years	24 months
Gd-enhancing lesions	**Number of Gd-enhancing lesions**	The Interferon beta Multiple Sclerosis Study group; Paty et al., Neurology 1993 (Interferon beta Multiple Sclerosis Study)	RRMS (n=327)	Interferon beta-1b vs. Placebo, median percentage of scans with lesions 5.9% vs. 29.4% after 3 years (p=0.0062); median number of lesions per year 0.5 vs. 3.0 (p=0.0089)	24 months
		Jacobs et al., Ann Neurol 1996 (MSCRG study)	RRMS (n=300)	Interferon beta-1a 30µg vs. Placebo, number of lesions 1.04 vs. 1.59 after 1 year (p=0.02), and 0.80 vs. 1.65 after 2 years	104 weeks

				(p=0.05); scans with lesions 29.9% vs. 42.3% after 1 year (p=0.05)	
		Comi et al., Ann Neurol 2001 (European/Canadian Glatiramer Acetate Study)	RRMS (n=249)	Glatiramer Acetate vs. Placebo, mean cumulative number of lesions 36.8 vs. 26.0 (p=0.003) after 9 months; mean number of lesions per patient 2.9 vs. 4.1 (p<0.005) after 9 months; total number of new lesions 17.4 vs. 26 (p<0.003) after 9 months; mean percent of scans without lesions 28.7% vs. 35.8% (p=0.04) after 9 months	9 months
		Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, number of lesions 0.1 vs. 1.3 after 1 year (p<0.001), 0.1 vs. 1.2 after 2 years (p<0.001), and 0.2 vs. 2.4 overall (p<0.001)	24 months
		Rudick et al., New Eng J Med 2006 (SENTINEL study)	RRMS (n=1171)	Natalizumab + Interferon beta-1a vs. Interferon beta-1a, number of lesions 0.1 vs. 0.9 after 2 years (p<0.001)	24 months
		Mikol et al., Lancet Neurol 2008 (REGARD study)	RRMS (n=764)	Interferon beta-1[a] 44 µg vs. Glatiramer acetate 20 mg, lesions per patient per scan 0.24 vs. 0.41 after 96 weeks (p=0.0002); scans per patient with lesions 9.8% vs. 15.3% after 96 weeks (p=0.005)	96 weeks
		O'Connor et al., Lancet Neurol 2009 (BEYOND study)	RRMS (n=2244)	Interferon beta-1a 500µg vs. 250µg vs. Glatiramer acetate, number of lesions 1.0 vs. 0.9 vs. 1.2 after 2 years (p=0.80; p=0.07; p=0.12)	24 months
		Cohen et al., New Eng J Med 2010 (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 1.25mg and 0.5mg vs. Interferon beta-1a (30µg/week), number of lesions 0.14 (p<0.001) and 0.23 (p<0.001), vs. 0.51 after 12 months	12 months
		Kappos et al., New Eng J Med 2010; Radue et al., Arch Neurol 2012 (FREEDOMS study)	RRMS (n=1272)	Fingolimod 1.25mg and 0.5mg vs. Placebo, number of lesions 0.3 (p<0.001) and 0.2 (p<0.011), vs. 1.3 after 6 months, 0.3 (p<0.001)	24 months

				and 0.2 (p<0.011), vs. 1.1 after 12 months, 0.2 (p<0.001) and 0.2 (p<0.011), vs. 1.1 after 24 months	
		Comi et al., Ann Neurol 2011 (FORTE study)	RRMS (n=980)	Glatiramer Acetate 20mg vs. 40mg, number of lesions 0.68 vs. 0.54 (ns) after 12 months	12 months
		Giovannoni et al., Lancet Neurol 2011; Comi et al., J Neurol 2013 (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/kg and Cladribine 5.25mg/kg vs. Placebo, relative reduction 85.7% (p<0.001) and 87.9% (p<0.001) after 96 weeks	96 weeks
		O'Connor et al., New Eng J Med 2011; Wolinsky et al., Mult Scler 2013 (TEMSO study)	RMS (1088)	Teriflunomide 14mg and 7mg vs. Placebo, lesions per scan (relative risk reduction) 0.26 (80.4%) (p<0.0001) and 0.57 (57.2%) (p<0.0001), vs. 1.33 after 108 weeks	108 weeks
		Cohen et al., Lancet 2012 (CARE-MS I)	RRMS (n=581)	Alemtuzumab 12mg vs. Interferon beta-1a 44 µg, patients with lesions 7% vs. 19% (p<0.0001)	24 months
		Coles et al., Lancet 2012 (CARE-MS II)	RRMS (n=840)	Alemtuzumab 12mg vs. Interferon beta-1a 44 µg, patients with lesions 9% vs. 23% (p<0.0001)	24 months
		Comi et al., NEJM 2012; Filippi et al., J Neurol Neurosurg Psychiatry. 2014 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, cumulative number of lesions 1.33 vs. 2.12 (p<0.001) at 12 and 24 months	24 months
		Fox et al., New Eng J Med. 2012 (CONFIRM)	RRMS (n=682, MRI cohort)	Dimethyl Fumarate 240mg BID or TID or Glatiramer acetate vs. Placebo, number of lesions 0.5 (p<0.001), 0.4 (p<0.001), 0.7 (p<0.001), vs. 2.0 after 2 years	24 months
		Gold et al., New Eng J Med 2012; Arnold et al., J Neurol 2014 (DEFINE study)	RRMS (n=1234)	Dimethyl Fumarate 240mg BID and TID vs. Placebo, number of lesions 0.1 (p<0.001), 0,5 (p<0.001), vs. 1.8 after 96 weeks; in a sub-cohort of 540 patients, 0.1 (p<0.0001) and 0.3 (p<0.0001) vs. 1.5 after 6 months, 0.1 (p<0.0001) and 0.4 (p<0.0001) vs. 1.4 after 1 year, and 0.1 (p<0.0001) and 0.5 (p<0.0001) vs. 1.8 after 2	24 months

				years	
		Khan et al., Ann Neurol 2013; Zivadinov et al., J Neurol 2015 (GALA study)	RRMS (n=1404)	Glatiramer acetate 40 mg vs. Placebo, cumulative number of lesions 0.905 vs. 1.639 after 6 and 12 months (p<0.0001)	12 months
		Calabresi et al., Lancet Neurol 2014; Arnold et al., BMC Neurol 2014 (ADVANCE study)	RRMS (n=1512)	Peginterferon beta-1a every 4 vs. 2 weeks vs. Placebo, number of lesions 1.2 vs. 0.3 vs. 1.6 (p<0.0001; p<0.0001; p=0.099) after 24 weeks, and 0.9 vs. 0.2 vs. 1.4 (p<0.0001; p<0.0001; p=0.074) after 48 weeks	24 months
		Calabresi et al., Lancet Neurol 2014 (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 1.25mg and 0.5mg vs. Placebo, number of lesions 0.2 (p<0.001) and 0.4 (p<0.001), vs. 1.2 after 24 months	24 months
		Massacesi et al., PloS One 2014 (EudraCT 2006-004937-13)	RRMS (n=150)	Azatioprine (3mg/kg/day) vs. Interferon, number of lesions 0.2 vs. 0.4 after 2 years (p=0.52); and number of patients with lesions (0, 1-2, ≥3) 41/8/0 vs. 43/1/3 after 2 years (p=0.39)	24 months
		Vollmer et al., J Neurol 2014 (BRAVO)	RRMS (n=1331)	Laquinimod or Interferon beta-1a 30 µg vs. Placebo, cumulative number of lesions 1.84 (p=0.069) or 0.90 (p<0.001) vs. 2.34 after 12 an 24 months	24 months
		Cohen et al., JAMA Neurol 2015 (equivalence study) (GATE study)	RRMS (n=794)	Glatiramer acetate 20mg generic or brand version vs. Placebo, number of lesions 0.42 (p<0.001), or 0.38 (p<0.001), vs. 0.82 during months 7 through 9; ratio of generic drug to brand drug of 1.095	9 months
		Kappos et al., New Eng J Med 2015 (DECIDE study)	RRMS (n=1841)	Daclizumab vs. Interferon, number of lesions 0.5 vs. 0.8 (p<0.001) after 24 weeks	144 weeks
		Miller et al., Neurology 2015 (CONFIRM study)	RRMS (n=681)	Dimethyl Fumarate 240mg BID and TID vs. Glatiramer Acetate vs. Placebo, number of lesions 0.5 (p<0.0001), 0.5 (p<0.0001), and 1.6 (p<0.05) vs. 1.7 after 24 weeks, 0.4 (p<0.0001), 0.4 (p<0.0001), and 0.7	24 months

				(p<0.0001) vs. 2.2 after 1 year, and 0.5 (p<0.0001), 0.4 (p<0.001), and 0.8 (p<0.001) vs. 2.0 after 2 years	
	Proportion on patients with Gd-enhancing lesions	Mikol et al., Lancet Neurol 2008 (REGARD study)	RRMS (n=764)	Interferon beta-1[a] 44 µg vs. Glatiramer acetate 20 mg, patients with no lesions 81% vs. 67% after 96 weeks (p=0.0005)	96 weeks
		Kappos et al., New Eng J Med 2010; Radue et al., Arch Neurol 2012 (FREEDOMS study)	RRMS (n=1272)	Fingolimod 1.25mg and 0.5mg vs. Placebo, percent of patients free of lesions -87.8% (p<0.001) and 88.3 (p<0.001), vs. 64.3% after 12 months, 89.8% (p<0.001) and 89.7% (p<0.001), vs. 65.1% after 24 months	24 months
		Cohen et al., New Eng J Med 2010 (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 1.25mg and 0.5mg vs. Interferon beta-1a (30µg/week), percent of patients free of lesions 91.2% (p<0.001) and 90.1% (p<0.001), vs. 80.8% after 12 months	12 months
		O'Connor et al., New Eng J Med 2011; Wolinsky et al., Mult Scler 2013 (TEMSO study)	RMS (1088)	Teriflunomide 14mg and 7mg vs. Placebo, percent of patients free of lesions 64.1% (p<0.001) and 51.4% (p<0.001) vs. 39.0% after 108 weeks	108 weeks
		Giovannoni et al., Lancet Neurol 2011; Comi et al., J Neurol 2013 (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/kg and Cladribine 5.25mg/kg vs. Placebo, percent of patients free of lesions 87.2% (p<0.001) and 91.4% (p<0.001), vs. 78.9% after 96 weeks	96 weeks
		Calabresi et al., Lancet Neurol 2014 (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 1.25mg and 0.5mg vs. Placebo, percent of patients free of lesions 96% (p<0.001) and 87% (p<0.001), vs. 65% after 24 months	24 months
		Lanzillo et al., Mult Scler 2016 (ARIANNA study)	RRMS (n=154)	Interferon beta-1b with or without Atorvastatin 40 mg, percent of patients with lesions 8% vs. 18% after 2 years (p=0.20)	24 months
	Volume of Gd-enhancing lesions	Jacobs et al., Ann Neurol 1996 (MSCRG study)	RRMS (n=300)	Interferon beta-1a 30µg vs. Placebo, lesion volume 70.0mm^3 vs. 96.5mm^3 after 1 year (p=0.02), and	104 weeks

				38.3mm^3 vs. 48.5mm^3 after 2 years (p=0.03)	
		Comi et al., Ann Neurol 2001 (European/Canadian Glatiramer Acetate Study)	RRMS (n=249)	Glatiramer Acetate vs. Placebo, volume change -245.3µL vs. -105.1µL (p=0.01) after 9 months	9 months
		Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, lesion volume of 21mm^3 vs. 207mm^3 after 1 year (p<0.001), 32mm^3 vs. 192mm^3 after 2 years (p<0.001); volume change -343mm^3 vs. -126mm^3 after 1 year (p<0.001), and -332mm^3 vs. -141mm^3 after 2 years (p<0.001)	24 months
		Mikol et al., Lancet Neurol 2008 (REGARD study)	RRMS (n=764)	Interferon beta-1a 44µg vs. Glatiramer acetate 20 mg, volume change -164.3mm^3 vs. -162.6mm^3 after 96 weeks (p=0.42)	96 weeks
		O'Connor et al., Lancet Neurol 2009 (BEYOND study)	RRMS (n=2244)	Interferon beta-1a 500µg vs. 250µg vs. Glatiramer acetate, cumulative volume 0.11cm^3 vs. 0.12cm^3 vs. 0.14cm^3 after 2 years (p=0.87; p=0.028; p=0.017)	24 months
		Cohen et al., New Eng J Med 2010 (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 1.25mg and 0.5mg vs. Interferon beta-1a (30µg/week), lesion volume 19.54mm^3 (p<0.001) and 22.61mm^3 (p<0.001), vs. 50.68mm^3 after 12 months	12 months
		Gold et al., New Eng J Med 2012; Arnold et al., J Neurol 2014 (DEFINE study)	RRMS (n=1234)	Dimethyl Fumarate 240mg BID and TID vs. Placebo, in a sub-cohort of 540 patients, median volume change -203.2mm^3 (p<0.01) and -118.7mm^3 (p<0.05) vs. -1.8mm^3 after 6 months, -160.9mm^3 (p<0.01) and -110.2mm^3 (p<0.01) vs. -12.6mm^3 after 1 year, and -152.7mm^3 (p<0.0001) and -57.8mm^3 (p<0.0001) vs. 15.1mm^3 after 2 years	24 months
		Miller et al., Neurology 2015 (CONFIRM study)	RRMS (n=681)	Dimethyl Fumarate 240mg BID and TID vs. Glatiramer Acetate vs. Placebo, mean lesion volume 46.0mm^3	24 months

				(p<0.0001), 30.9mm^3 (p<0.0001), and 162.5mm^3 (p=0.0544) vs. 143.6mm^3 after 24 weeks, 27.0mm^3 (p<0.0001), 56.2mm^3 (p<0.0001), and 77.0mm^3 (p=0.0544) vs. 189.5mm^3 after 1 year, and 35.9mm^3 (p<0.0001), 42.6mm^3 (p<0.0001), and 45.6mm^3 (p<0.0001) vs. 141.8mm^3 after 2 years	
T1 lesions	**Number of new T1 lesions**	Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, number of lesions 0.6 vs. 2.3 after 1 year (p<0.001), 0.4 vs. 2.3 lesions after 2 years (p<0.001), and 1.1 vs. 4.6 overall (p<0.001)	24 months
		Mikol et al., Lancet Neurol 2008 (REGARD study)	RRMS (n=764)	Interferon beta-1[a] 44 µg vs. Glatiramer acetate 20 mg, lesions per patient per scan 0.23 vs. 0.24 after 96 weeks (p=0.15); scans per patient with lesions 10.5% vs. 12.4% after 96 weeks (p=0.12); patients with no lesions 75% vs. 70% after 96 weeks (p=0.29)	96 weeks
		O'Connor et al., New Eng J Med 2011; Wolinsky et al., Mult Scler 2013 (TEMSO study)	RMS (1088)	Teriflunomide 14mg and 7mg vs. Placebo, mean difference from Placebo -0.030 (p=0.0161) and -0.016 (p=0.1916) after 108 weeks	108 weeks
		Comi et al., New Eng J Med 2012; Filippi et al., J Neurol Neurosurg Psychiatry. 2014 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, cumulative number of lesions 1.61 vs. 2.23 (p=0.004) after 24 months	24 months
		Fox et al., New Eng J Med. 2012 (CONFIRM study)	RRMS (n=682, MRI cohort)	Dimethyl Fumarate 240mg BID or TID or Glatiramer acetate vs. Placebo, number of lesions 3.0 (p<0.001), 2.4 (p<0.001), 4.1 (p=0.002), vs. 7.0 after 2 years	24 months
		Calabresi et al., Lancet Neurol 2014; Arnold et al., BMC Neurol 2014 (ADVANCE study)	RRMS (n=1512)	Peginterferon beta-1a every 4 vs. every 2 weeks, vs. Placebo, number of lesions 2.0 vs. 1.2 vs. 2.1 (p<0.0001; p<0.0001; p=0.23) after 24 weeks, and 3.1 vs. 1.8 vs. 3.8 (p<0.0001; p<0.0001;	24 months

				p=0.082) after 48 weeks	
		Kappos et al., New Eng J Med 2015 (DECIDE study)	RRMS (n=1841)	Daclizumab vs. Interferon, number of lesions 1.22 vs. 1.94 (p<0.001) after 24 weeks; 2.13 vs. 4.43 (p<0.001) after 96 weeks	144 weeks
	Number of new non-enhancing T1 lesions	Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, number of lesions 0.6 vs. 1.9 after 1 year (p<0.001), 0.4 vs. 1.9 after 2 years (p<0.001), and 1.0 vs. 3.8 overall (p<0.001)	24 months
		Giovannoni et al., Lancet Neurol 2011; Comi et al., J Neurol 2013 (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/kg and Cladribine 5.25mg/kg vs. Placebo, relative reduction 2.9% (p<0.001) and 8.2% (p<0.001) after 96 weeks	96 weeks
		Gold et al., New Eng J Med 2012; Arnold et al., J Neurol 2014 (DEFINE study)	RRMS (n=1234)	Dimethyl Fumarate 240mg BID and TID vs. Placebo, in a sub-cohort of 540 patients, number of lesions 0.8 (p<0.0001) and 1.0 (p<0.001) vs. 1.9 after 6 months, 1.1 (p<0.0001) and 1.4 (p<0.0001) vs. 3.5 after 1 year, and 1.5 (p<0.0001) and 2.1 (p<0.0001) vs. 5.6 after 2 years	24 months
		Khan et al., Ann Neurol 2013; Zivadinov et al., J Neurol 2015 (GALA study)	RRMS (n=1404)	Glatiramer Acetate 40mg vs. Placebo, number of lesions 0.31 vs. 0.45 (p=0.0258) between 6 and 12 months; proportion of new active lesions converting to T1 lesions 15.8% vs. 19.8% (p=0.0060) between 6 and 12 months	12 months
		Miller et al., Neurology 2015 (CONFIRM study)	RRMS (n=681)	Dimethyl Fumarate 240mg BID and TID vs. Glatiramer Acetate vs. Placebo, number of lesions 2.2 (p<0.001), 1.5 (p<0.0001), and 2.6 (p<0.05) vs. 3.7 after 1 year, 1.0 (p<0.0001), 0.9 (p<0.0001), and 1.5 (p<0.001) vs. 3.3 between 1 and 2 years, and 3.0 (p<0.0001), 2.4 (p<0.0001), and 4.1 (p<0.01) vs. 7.0 after 2 years	24 months

	Volume of T1 lesions	Comi et al., Ann Neurol 2001 (European/Canadian Glatiramer Acetate Study)	RRMS (n=249)	Glatiramer Acetate vs. Placebo, volume change 0.8mL vs. 1.3mL (p=0.14) after 9 months	9 months
		Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, volume after 1 (p=0.004) and 2 years (p<0.001); volume change of -1508mm^3 vs. 548mm^3 overall (p<0.001); percent change -23.5% vs. -1.5% overall (p<0.001)	24 months
		Mikol et al., Lancet Neurol 2008 (REGARD study)	RRMS (n=764)	Interferon beta-1^a 44 µg vs. Glatiramer acetate 20 mg, volume change -667.0 mm^3 vs. -377.3mm^3 after 96 weeks (p=0.29)	96 weeks
		O'Connor et al., Lancet Neurol 2009 (BEYOND study)	RRMS (n=2244)	Interferon beta-1a 500µg vs. 250µg vs. Glatiramer acetate, percent volume change 36.0% vs. 23.1% vs. 40.6% after 2 years (p=0.18; p=0.54; p=0.68)	24 months
		Cohen et al., New Eng J Med 2010 (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 1.25mg and 0.5mg vs. Interferon beta-1a (30µg/week), percent volume change 34.7% (p=0.09) and 24.1% (p=0.17), vs. 15.0% after 12 months	12 months
		Kappos et al., New Eng J Med 2010; Radue et al., Arch Neurol 2012 (FREEDOMS study)	RRMS (n=1272)	Fingolimod 1.25mg and 0.5mg vs. Placebo, volume change 30mm^3 (p<0.001) and 33mm^3 (p=0.008), vs. 173mm^3 after 24 months; percent volume change 12.2% (p=0.02) and 8.8% (p=0.01), vs. 50.7% after 24 months	24 months
		O'Connor et al., New Eng J Med 2011; Wolinsky et al., Mult Scler 2013 (TEMSO study)	RMS (n=1088)	Teriflunomide 14mg and 7mg vs. Placebo, volume change 0.33mL (p=0.02) and 0.50mL (p=0.19) vs. 0.53mL after 108 weeks	108 weeks
		Sorensen et al., Lancet Neurol 2011 (SIMCOMBIN study)	RRMS (n=307)	Interferon beta-1^a 30 µg with vs. without Simvastatin 80 mg, volume change -0.011mL vs. 0.019mL after 12 months (p=0.547)	12 months
		Gold et al., New Eng J Med 2012; Arnold et al., J Neurol 2014 (DEFINE study)	RRMS (n=1234)	Dimethyl Fumarate 240mg BID and TID vs. Placebo, in a sub-cohort of 540 patients, median	24 months

				percent volume change 1.5% (ns) and 2.5% (ns) vs. 4.3% after 6 months, 5.4% (p<0.05) and 4.7% (ns) vs. 11.6% after 1 year, and 8.4% (p<0.0001) and 12.7% (p<0.01) vs. 26.9% after 2 years	
		Calabresi et al., Lancet Neurol 2014; Arnold et al., BMC Neurol 2014 (ADVANCE study)	RRMS (n=1512)	Peginterferon beta-1a every 4 and 2 weeks vs. Placebo, volume change 0.31cm^3 (p<0.0001) and -0.18cm^3 (p<0.0001) vs. 0.29cm^3 after 24 weeks, and 0.57cm^3 (p=0.018) and -0.32cm^3 (p<0.0001) vs. 0.54cm^3 after 48 weeks	24 months
		Calabresi et al., Lancet Neurol 2014 (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 1.25mg and 0.5mg vs. Placebo, percent volume change -4.69% (p=0.205) and 12.64% (p=0.372), vs. 26.42% after 24 months	24 months
		Kappos et al., New Eng J Med 2015 (DECIDE study)	RRMS (n=1841)	Daclizumab vs. Interferon, percent volume change 10.5% vs. 14.1% (p<0.001) after 24 weeks; 22.8% vs. 33.4% (p<0.001) after 96 weeks	144 weeks
		Miller et al., Neurology 2015 (CONFIRM study)	RRMS (n=681)	Dimethyl Fumarate 240mg BID and TID vs. Glatiramer Acetate vs. Placebo, median percent volume change 1.5% (p=0.2587), 2.8% (p=0.6540), and 2.5% (p=0.2741) vs. 7.9% after 1 year, and 10.7% (p<0.001), 8.5% (p<0.01), and 8.6% (p<0.01) vs. 19.5% after 2 years	24 months
	Permanent black holes (PBH)	Comi et al., NEJM 2012; Filippi et al., J Neurol Neurosurg Psychiatry. 2014 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo: Number of PBH from Gd+ lesions: 1.0 vs. 2.1 (p=0.001); Number of PBH from new T2 lesions: 0.87 vs. 1.67 (p=0.009); Number of PBH from Gd+ lesions and new T2 lesions: 1.20 vs. 2.34 (p<0.001); Proportion of Gd+ lesions converting to PBH: 21% vs. 29% (p=0.117); Proportion of new T2 lesions converting	24 months

				to PBH: 23% vs. 26% (p=0.572); Proportion of Gd+ lesions and new T2 lesions converting to PBH: 23% vs. 28% (p=0.260);	
	T1/T2 lesion volume ratio	Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, ratio 0.270 vs. 0.311 after 2 years (p=0.002 adjusting for the baseline ratio); changes in the ratio -0.058 vs. vs. -0.03 (p=0.002 adjusting for the baseline ratio)	24 months
Combined measures	**Combined unique active lesions**	Mikol et al., Lancet Neurol 2008 (REGARD study)	RRMS (n=764)	Interferon beta-1a 44 µg vs. Glatiramer acetate 20 mg, lesions per patient per scan 0.91 vs. 1.22 after 96 weeks (p=0.010); scans per patient with lesions 26.4% vs. 32.3% after 96 weeks (p=0.009); patients with no lesions 38% vs. 31% after 96 weeks (p=0.125)	96 weeks
		Kappos et al., New Eng J Med 2010; Radue et al., Arch Neurol 2012 (FREEDOMS study)	RRMS (n=1272)	Fingolimod 1.25mg and 0.5mg vs. Placebo, percent of patients lesion-free 58.7=2% (p<0.001) and 57.4% (p<0.001) vs. 27.1% after 12 months, 69.6% (p<0.001) and 73.1% (p<0.001) vs. 33.1% between 12 and 24 months, and 52.0% (p<0.001) and 50.7% (p<0.001) vs. 21.0% after 24 months	24 months
		Giovannoni et al., Lancet Neurol 2011; Comi et al., J Neurol 2013 (CLARITY study)	RRMS (n=1326)	Cladribine 3.5mg/kg and Cladribine 5.25mg/kg vs. Placebo, proportion of patients with MRI lesion activity-free 60.0% (p<0.001) and 61.2% (p<0.001), vs. 25.5% after 96 weeks; relative reduction: 0.43 (p<0.001) and 0.38 (p<0.001) vs. 1.72 after 96 weeks	96 weeks
		O'Connor et al., New Eng J Med 2011; Wolinsky et al., Mult Scler 2013 (TEMSO study)	RMS (1088)	Teriflunomide 14mg and 7mg vs. Placebo, lesions per scan (percent reduction vs Placebo) 0.75 (69.4%) (p<0.0001) and 1.29 (47.7%) (p<0.0001) vs. 2.46 after	108 weeks

				108 weeks	
		Comi et al., Lancet Neurol 2012 (REFLEX study)	CIS (n=517)	Interferon beta-1a three times a week vs. once a week vs. Placebo, number of lesions per patient per scan 0.60 vs. 1.23 vs. 2.70 (p<0.0001; p<0.0001; p=0.0015) after 2 years	108 weeks
		Lublin et al., Ann Neurol 2013 (CombiRx study)	RRMS (n=1008)	IFN beta-1a 30mcg SC/week + GA 20mg SC/day vs IFN beta-1a 30mcg SC/week vs GA 20mg SC/day: percent of patients free of lesions 49.2% vs. 32.2% vs. 32.5% (p<0.0001; p=0.95) after 36 months	36 months
		Calabresi et al., Lancet Neurol 2014; Arnold et al., BMC Neurol 2014 (ADVANCE study)	RRMS (n=1512)	Peginterferon beta-1a every 4 vs. 2 weeks vs. Placebo, percent of patients without MRI activity 24.9% vs. 40.9% vs. 19.1% (p<0.0001; p<0.0001; p=0.0318) after 48 weeks, 34.2% vs. 46.4% vs. 26.2% (p=0.0002; p<0.0001; p=0.0078) after 24 weeks, and 39.8% vs. 65.4% vs. 31.5% (p<0.0001; p<0.0001; p=0.0080) between 24 and 48 weeks; mean number of lesions 7.3 (p<0.001), and 3.7 (p<0.001) vs. 11.2 after 1 year	24 months
		Calabresi et al., Lancet Neurol 2014 (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 1.25mg and 0.5mg vs. Placebo, percent of patient free of MRI activity 63% (p<0.001) and 50% (p<0.001), vs. 26% after 24 months	24 months
		Massacesi et al., PloS One 2014 (EudraCT 2006-004937-13)	RRMS (n=150)	Azathioprine (3mg/kg/day) vs. Interferon, annualised number of lesions 0.78 vs. 0.70 after 2 years (p=0.53)	24 months
	Z4 score (Sum of Z-scores for volumes of Gd+ lesion volume, T2 lesions, T1	Noseworthy et al., Neurology 2000, phase III; Wolinsky et al., Neurology 2000 (Linomide study)	RMS (n=715)	Linomide vs. Placebo, Z4 score -0.05 vs. 0.13 (p<0.0006) after 6 months	Early termination for safety issues (initially planned: 36 months)

	lesions and CSF)				
		O'Connor et al., New Eng J Med 2011; Wolinsky et al., Mult Scler 2013 (TEMSO study)	RMS (1088)	Teriflunomide 14mg and 7mg vs. Placebo, mean Z4 score difference from Placebo -0.512 (p<0.0002) and -0.333 (p=0.0008) after 108 weeks	108 weeks
Brain atrophy	**Brain parenchymal fraction**	Polman et al., New Eng J Neurol 2006; Miller et al., Neurology 2007 (AFFIRM study)	RRMS (n=627)	Natalizumab vs. Placebo, percent volume change -0.56% vs. -0.40% after 1 year (p=0.002), -0.43% vs. -0.24% after 2 years (p=0.004), and -0.80 vs. -0.82 overall (ns)	24 months
		Mikol et al., Lancet Neurol 2008 (REGARD study)	RRMS (n=764)	Interferon beta-1a 44 µg vs. Glatiramer acetate 20 mg, percent volume change -1.240% vs. -1.073% after 96 weeks (p=0.018)	96 weeks
		O'Connor et al., Lancet Neurol 2009 (BEYOND study)	RRMS (n=2244)	Interferon beta-1a 500µg vs. 250µg vs. Glatiramer acetate, percent volume change -0.64% vs. -0.65% vs. -0.61% after 2 years (p=0.74; p=0.33; p=0.46)	24 months
		Cohen et al., New Eng J Med 2010 (TRANSFORMS study)	RRMS (n=1292)	Fingolimod 1.25mg and 0.5mg vs. Interferon beta-1a (30µg/week), percent volume change -0.30% (p<0.001) and -0.31% (p<0.001), vs. -0.45% after 12 months	12 months
		Kappos et al., New Eng J Med 2010; Radue et al., Arch Neurol 2012 (FREEDOMS study)	RRMS (n=1272)	Fingolimod 1.25mg and 0.5mg vs. Placebo, percent volume change (relative reduction compared with Placebo) (p=0.006) and -0.22% (39.2%) (p=0.003) vs. -0.34% after 6 months, -0.44% (22.7%) (p=0.03) and -0.50% (32.3%) (p=0.001) vs. -0.65% after 12 months, -0.42% (36.8%) (p=0.002) and -0.37% (44.7%) (p<0.001) vs. -0.67% between 12 and 24 months, -0.89% (35.5%) (p<0.001) and -0.84% (32.2%) (p<0.001) vs. -1.31% after 24 months	24 months
		Comi et al., Ann Neurol 2011	RRMS (n=980)	Glatiramer Acetate 20mg vs. 40mg, percent volume	12 months

		(FORTE study)		change -0.58% vs. -0.53% (ns) after 12 months	
		Sorensen et al., Lancet Neurol 2011 (SIMCOMBIN study)	RRMS (n=307)	Interferon beta-1[a] 30 µg with vs. without Simvastatin 80 mg, volume change -0.0099mL vs. -0.00080mL after 12 months (p=0.370)	12 months
		Cohen et al., Lancet 2012 (CARE-MS I)	RRMS (n=563)	Alemtuzumab 12mg vs. Interferon beta-1a 44 µg, median percent volume change -0.867% vs. -1.488% after 2 years (p<0.0001)	24 months
		Coles et al., Lancet 2012 (CARE-MS II)	RRMS (n=840)	Alemtuzumab 12mg vs. Interferon beta-1a 44 µg, median percent volume change -0.615% vs. -0.810% after 2 years (p=0.01)	24 months
		Comi et al., New Eng J Med. 2012 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, percent volume change -0.87% vs. -1.30% (p<0.001) after 24 months	24 months
		Gold et al., New Eng J Med 2012; Arnold et al., J Neurol 2014 (DEFINE study)	RRMS (n=1234)	Dimethyl Fumarate 240mg BID and TID vs. Placebo, in a sub-cohort of 540 patients, median percent volume change -0.64% (p<0.05) and -0.77% (ns) vs. -0.81% after 6 months, -0.46% (p<0.05) and -0.55% (ns) vs. -0.66% between 6 months and 2 years	24 months
		Khan et al., Ann Neurol 2013 Zivadinov et al., J Neurol 2015 (GALA study)	RRMS (n=1404)	Glatiramer acetate 40 mg vs. Placebo, percent volume change -0.706% vs. -0.645% after 12 months (p=0.2058)	12 months
		Calabresi et al., Lancet Neurol 2014; Arnold et al., BMC Neurol 2014 (ADVANCE study)	RRMS (n=1512)	Peginterferon beta-1a every 4 vs. 2 weeks vs. Placebo, mean percent volume change -0.671% (p=0.3747), and -0.721% (p=0.0841), vs. -0.621% after 1 year	24 months
		Calabresi et al., Lancet Neurol 2014 (FREEDOMS II study)	RRMS (n=1083)	Fingolimod 1.25mg and 0.5mg vs. Placebo, percent volume change -0.128% (p<0.001) and -0.228% (p=0.012), vs. -0.375% after 6 months; -0.354% (p<0.001) and -0.377% (p=0.0004), vs. -0.629% after 12 months; -	24 months

				0.285% (p<0.001) and -0.486% (p=0.013), vs. -0.678% after 24 months	
		Vollmer et al., J Neurol 2014 (BRAVO)	RRMS (n=1331)	Laquinimod or Interferon beta-1a 30 µg vs. Placebo, percent volume change -0.75% (p<0.001) or -1.14% (p=0.14) vs. -1.03% after 24 months	24 months
		Miller et al., Neurology 2015 (CONFIRM study)	RRMS (n=681)	Dimethyl Fumarate 240mg BID and TID vs. Glatiramer Acetate vs. Placebo, median percent volume change -0.320% (p=0.6645), -0.450% (p=0.9299), and -0.580% (p=0.2593) vs. -0.440% after 1 year, -0.400% (p=0.0359), -0.400% (p=0.0755), and -0.420% (p=0.0805) vs. -0.590% between 1 and 2 years, and -0.660% (p=0.0645), -0.750% (p=0.2636), and -0.960% (p=0.8802) vs. -0.945% after 2 years	24 months
		Lanzillo et al., MSJ 2016 (ARIANNA study)	RRMS (n=154)	Interferon beta-1b with or without Atorvastatin 40 mg, percent volume change -0.367% vs. -0.302% after 1 year (ns), -0.382% vs. -0.545% after 2 years (ns); percent annualized volume change -0.380% vs. -0.316% (p=0.920)	24 months
	Grey matter	O'Connor et al., New Eng J Med 2011; Wolinsky et al., MSJ 2013 (TEMSO study)	RMS (1088)	Teriflunomide 14mg and 7mg vs. Placebo, volume change –0.003mL (p=0.35) and -0.003mL (p=0.19) vs. -0.004mL after 108 weeks	108 weeks
		Filippi et al., J Neurol Neurosurg Psychiatry. 2013 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, median percent volume change -0.3% vs. -0.8% (p=0.004) after 12 months, -0.7% vs. -0.6% (p=0.664) between 12 and 24 months, and -0.9% vs. -1.2% (p=0.372) after 24 months	24 months
		Lublin et al., Ann Neurol 2013 (CombiRx study)	RRMS (n=1008)	IFN beta-1a 30mcg SC/week + GA 20mg SC/day vs IFN beta-1a 30mcg SC/week vs GA	36 months

				20mg SC/day: percent volume change -2.60% vs. -2.99% vs. -5.16% (ns; ns) after 36 months	
	White matter	O'Connor et al., New Eng J Med 2011; Wolinsky et al., Mult Scler 2013 (TEMSO study)	RMS (n=1088)	Teriflunomide 14mg and 7mg vs. Placebo, mean volume difference from Placebo -6.146mL (p=0.0002) and -3.106mL (p=0.0609) after 108 weeks	108 weeks
		Filippi et al., J Neurol Neurosurg Psychiatry. 2013 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, median percent volume change -0.0% vs. -0.4% (p=0.004) after 12 months, -0.2% vs. -0.2% (p=0.857) between 12 and 24 months, and -0.3% vs. -0.5% (p=0.327) after 24 months	24 months
		Lublin et al., Ann Neurol 2013 (CombiRx study)	RRMS (n=1008)	IFN beta-1a 30mcg SC/week + GA 20mg SC/day vs IFN beta-1a 30mcg SC/week vs GA 20mg SC/day: volume change −1.73mL (SD 22.63) vs. −0.71mL (17.01) −1.72mL (15.66); differences were not statistically significant	36 months
	CSF	Lublin et al., Ann Neurol 2013 (CombiRx study)	RRMS (n=1008)	IFN beta-1a 30mcg SC/week + GA 20mg SC/day vs IFN beta-1a 30mcg SC/week vs GA 20mg SC/day: percent volume change 0.60% vs. 0.51% vs. 0.57% (ns; ns) after 36 months	36 months
	Thalamus	Filippi et al., J Neurol Neurosurg Psychiatry. 2013 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, median percent volume change -0.6% vs. -1.0% (p=0.005) after 12 months, -0.7% vs. -0.9% (p=0.233) between 12 and 24 months, and -1.3% vs. -1.8% (p=0.003)	24 months
MTR	**Whole brain**	Gold et al., NEJM 2012; Arnold et al., J Neurol 2014 (DEFINE study)	RRMS (n=1234, but MRI cohort: n=540)	Dimethyl fumarate BID vs. TID vs. placebo: percent change: BID: 0.129%, p (vs. placebo) 0.0027; TID: 0.096%, p (vs. placebo) 0.0051; Placebo: -0.386% (reduction) after 24 months	24 months

		Calabresi et al., Lancet Neurol 2014; Arnold et al., BMC Neurol 2014 (ADVANCE study)	RRMS (n=1512)	Peginterferon beta-1a every 4 vs. 2 weeks vs. Placebo, percent change -0.432% (p=0.6873), and -0.129% (p=0.0438), vs. -0.382% after 1 year	24 months
		Filippi et al., J Neurol Neurosurg Psychiatry. 2014 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, signal change 0.31 vs. -0.09 (p=0.013) after 12 months, -0.08 vs. -0.18 (p=0.642) between 12 and 24 months, and 0.23 vs. -0.27 (p=0.015) after 24 months	24 months
		Miller et al., Neurology 2015 (CONFIRM study)	RRMS (n=681)	Dimethyl Fumarate 240mg BID and TID vs. Glatiramer Acetate vs. Placebo, percent change: -0.167 (ns), -0.008 (ns), and 0.010 (ns) vs. -0.419 after 2 years	24 months
	White matter	Filippi et al., J Neurol Neurosurg Psychiatry. 2014 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, signal change 0.32 vs. -0.09 (p=0.013) after 12 months, -0.05 vs. -0.18 (p=0.486) between 12 and 24 months, and 0.27 vs. -0.27 (p=0.011) after 24 months	24 months
	Grey matter	Filippi et al., J Neurol Neurosurg Psychiatry. 2014 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, signal change 0.30 vs. -0.11 (p=0.014) after 12 months, -0.16 vs. -0.22 (p=0.787) between 12 and 24 months, and 0.14 vs. -0.33 (p=0.034) after 24 months	24 months
	T2 lesions	Filippi et al., J Neurol Neurosurg Psychiatry. 2014 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, signal change 0.39 vs. 0.02 (p=0.239) after 12 months, 0.07 vs. -0.08 (p=0.651) between 12 and 24 months, and 0.46 vs. -0.07 (p=0.168) after 24 months	24 months
Proton MR Spectroscopy	**NAA/Cr value**	Filippi et al., J Neurol Neurosurg Psychiatry. 2014 (ALLEGRO study)	RRMS (n=1106)	Laquinimod vs. Placebo, signal change 0.047 vs. -0.176 (p=0.179) after 24 months	24 months

Abbreviations: Gd: gadolinium; MTR: magnetisation transfer ratio; NAA/Cr: N-acetyl aspartate-creatine ratio; RRMS: relapsing-remitting MS.

Table 5: Brain MRI outcome measures in phase III trials in CIS

Brain MRI
Inclusion criteria: controlled phase III clinical trials
Exclusion criteria: incomplete data presentation (e.g. missing values); descriptive findings in absence of any statistical analysis; secondary analyses of clinical trials and extension studies evaluating the same clinical endpoints of the main trial in population subgroups or during longer observation time.

Original neuroimaging outcome	Derived outcome measures	Trial	Condition (no. of patients randomised)	Drug, effect (vs. placebo/ another active arm)	Duration of the trial
T2 lesions	**Number of new lesions**	Kappos et al. Neurology 2006; Barkhof et al. Ann Neurol 2007 (BENEFIT study)	CIS (n=487)	Interferon beta-1b vs. Placebo, cumulative number of lesions 2.9 vs. 4.4 up to the conversion to MS (p<0.0001), 2.2 vs. 4.6 after 2 years (p<0.001)	24 months
		Comi et al. Lancet 2009 (PRECISE study)	CIS (n=481)	Glatiramer Acetate vs. Placebo, number of lesions 4.2 vs. 9.8 (p<0.0001) after 2.32 years	36 months
	Number of new or enlarging lesions	Jacobs et al. New Eng J Med 2000, Phase III (CHAMPS study)	CIS (n=383)	Interferon beta-1a 30µg vs. Placebo, number of lesions 1.5 vs. 2.8 after 6 months (p=0.01), 2.1 vs. 4.0 after 12 months (p<0.001), 2.1 vs. 5.0 after 18 months (p<0.001)	Early termination due to obvious superiority of IFN over placebo (initially planned: 36 months)
		Comi et al. Lancet 2001, phase III (ETOMS study)	CIS (n=309)	Interferon beta-1a 22µg vs. Placebo, median number of lesions per patient per scan 2.0 vs. 3.0 after 2 years (p<0.001)	24 months
		Leist et al. Lancet Neurol 2014 (ORACLE MS study)	CIS (n=616)	Cladribine 5.25 mg/Kg or 3.5 mg/Kg, vs. Placebo, median cumulative number of lesions 0.0 or 0.0 vs. 2.0 after 96 weeks (p<0.001)	96 weeks
	Lesion volume	Miller et al. Lancet Neurol 2014	CIS (n=618)	Teriflunomide 14mg vs. 7mg vs. Placebo, volume change -0.028mL (p=0.0374) vs. 0.023mL	108 weeks

		(TOPIC study)		(p=0.7789) vs. 0.044mL after 108 weeks	
	Volume of T2 lesions	Jacobs et al. New Eng J Med 2000 (CHAMPS study)	CIS (n=383)	Interferon beta-1a 30µg vs. Placebo, median volume change -123mm^3 vs. 40mm^3 after 6 months (p<0.001), 102mm^3 vs. 214mm^3 after 12 months (p=0.004), 28mm^3 vs. 313mm^3 after 18 months (p<0.001)	Early termination: obvious superiority of IFN over placebo (initially planned: 36 months)
		Comi et al. Lancet 2001 (ETOMS study)	CIS (n=309)	Interferon beta-1a 22µg vs. Placebo, median volume change -487mm^3 vs. -299mm^3 after 2 years (p=0.002); median percent volume change -13.0% vs. 8.8% after 2 years (p=0.002)	24 months
		Kappos et al. Neurology 2006; Barkhof et al. Ann Neurol 2007 (BENEFIT study)	CIS (n=487)	Interferon beta-1b vs. Placebo, volume change -888.5mm^3 vs. -431.6mm^3 up to the conversion to MS (p<0.05), -1.0cm^3 vs. -0.3cm^3 after 2 years (p=0.02)	24 months
		Miller et al. Lancet Neurol 2014 (TOPIC study)	CIS (n=618)	Teriflunomide 14mg vs. 7mg vs. Placebo, volume change -0.029mL (p=0.0503) vs. 0.022mL (p=0.7360) vs. 0.045mL after 108 weeks	108 weeks
Gd-enhancing lesions	**Number of Gd-enhancing lesions**	Jacobs et al. New Eng J Med 2000 (CHAMPS study)	CIS (n=383)	Interferon beta-1a 30µg vs. Placebo, number of lesions 0.9 vs. 1.5 after 6 months (p=0.03), 0.7 vs. 1.6 after 12 months (p=0.02), 0.4 vs. 1.4 after 18 months (p<0.001)	Early termination: obvious superiority of IFN over placebo (initially planned: 36 months)
		Comi et al. Lancet 2001 (ETOMS study)	CIS (n=309)	Interferon beta-1a 22µg vs. Placebo, median number of lesions per patient per scan 0.5 vs. 0.0 after 2 years (p=0.809)	24 months
		Kappos et al. Neurology 2006; Barkhof et al. Ann Neurol	CIS (n=487)	Interferon beta-1b vs. Placebo, cumulative number of lesions 1.9 vs. 4.3 up to conversion to MS (p<0.0001), 2.2 vs. 4.6 after 2 years (p<0.001); new lesions per scan 0.4 vs. 1.0	24 months

		2007 (BENEFIT study)		after 2 years (p<0.001)	
		Leist et al. Lancet Neurol 2014 (ORACLE MS study)	CIS (n=616)	Cladribine 5.25 mg/Kg or 3.5 mg/Kg, vs. Placebo, median cumulative number of lesions 0.0 or 0.0 vs. 2.0 after 96 weeks (p<0.001)	96 weeks
		Miller et al. Lancet Neurol 2014 (TOPIC study)	CIS (n=618)	Teriflunomide 14mg vs. 7mg vs. Placebo, number of lesions per scan 0.395 (p=0.0008) vs. 0.749 (p=0.4436) vs. 0.953 after 108 weeks	108 weeks
	Volume of Gd-enhancing lesions	Kappos et al. Neurology 2006; Barkhof et al. Ann Neurol 2007 (BENEFIT study)	CIS (n=487)	Interferon beta-1b vs. Placebo, cumulative volume of lesions 203.5mm^3 vs. 520.6mm^3 up to conversion to MS (p<0.0001), 0.2cm^3 vs. 0.5cm^3 after 2 years (p<0.001); volume of lesions per scan 0.1cm^3 vs. 0.1cm^3 after 2 years (p<0.001)	24 months
		Miller et al. Lancet Neurol 2014 (TOPIC study)	CIS (n=618)	Teriflunomide 14mg vs. 7mg vs. Placebo, volume change 0.034mL(p<0.0001) vs. 0.058mL (p=0.0077) vs. 0.079mL after 108 weeks	108 weeks
T1 lesions	**New T1 lesions**	Kappos et al. Neurology 2006; Barkhof et al. Ann Neurol 2007 (BENEFIT study)	CIS (n=487)	Interferon beta-1b vs. Placebo, cumulative number of lesions 0.2 vs. 0.3 after 2 years (p<0.001)	24 months
		Comi et al. Lancet 2009, phase III[10] (PRECISE study)	CIS (n=481)	Glatiramer Acetate vs. Placebo, cumulative number of lesions 1.7 vs. 3.6 (p<0.0001) after 2.32 years	36 months
	Volume of T1 lesions	Miller et al. Lancet Neurol 2014 (TOPIC study)	CIS (n=618)	Teriflunomide 14mg vs. 7mg vs. Placebo, volume change -0.016mL (p=0.0120) vs. 0.015mL (p=0.9100) vs. 0.014mL after 108 weeks	108 weeks
		Kappos et al.	CIS (n=487)	Interferon beta-1b vs. Placebo, change in volume	24 months

		Neurology 2006; Barkhof et al. Ann Neurol 2007 (BENEFIT study)		of lesions -0.0cm^3 vs. -0.1cm^3 after 2 years (p=0.29)	
Combined measures	**Combined unique lesions**	Comi et al. Lancet 2001 (ETOMS study)	CIS (n=309)	Interferon beta-1a 22µg vs. Placebo, proportion of patients without lesions 16% vs. 6% after 2 years (p=0.005)	24 months
		Kappos et al. Neurology 2006; Barkhof et al. Ann Neurol 2007 (BENEFIT study)	CIS (n=487)	Interferon beta-1b vs. Placebo, cumulative number of lesions 3.7 vs. 8.5 up to the conversion to MS (p<0.001), 5.7 vs. 10.3 after 2 years (p<0.001)	24 months
		Comi et al. Lancet Neurol 2012 (REFLEX study)	CIS (n=517)	Interferon beta-1a three times a week vs. once a week vs. Placebo, number of lesions per patient per scan 0.60 vs. 1.23 vs. 2.70 (p<0.0001; p<0.0001; p=0.0015) after 2 years	108 weeks
		Leist et al. Lancet Neurol 2014 (ORACLE MS study)	CIS (n=616)	Cladribine 5.25 mg/Kg or 3.5 mg/Kg, vs. Placebo, median cumulative number of lesions 1.0 or 1.0 vs. 4.0 after 96 weeks (p<0.001)	96 weeks
Brain atrophy	**Brain parenchymal fraction**	Comi et al. Lancet 2009 (PRECISE study)	CIS (n=481)	Glatiramer Acetate vs. Placebo, percent volume change -0.33% vs. -0.38% (ns)	36 months
		Miller et al. Lancet Neurol 2014 (TOPIC study)	CIS (n=618)	Teriflunomide 14mg vs. 7mg vs. Placebo, volume change -0.008mL (p=0.4495) vs. -0.002mL (p=0.4462) vs. -0.003mL after 108 weeks	108 weeks

Abbreviations: Gd: gadolinium; CIS: clinically isolated syndrome.

Table 6: Brain MRI outcome measures in phase III trials in progressive MS

Brain MRI
Inclusion criteria: controlled phase III clinical trials
Exclusion criteria: incomplete data presentation (e.g. missing values); descriptive findings in absence of any statistical analysis; secondary analyses of clinical trials and extension studies evaluating the same clinical endpoints of the main trial in population subgroups or during longer observation time.

Original neuroimaging outcome	Derived outcome measures	Trial	Condition (no. of patients randomised)	Drug, effect (vs. placebo/ another active arm)	Duration of the trial
T2 lesions	**Number of new or enlarging lesions**	Secondary Progressive Efficacy Clinical Trial of Recombinant Interferon-beta-1a in MS (SPECTRIMS) Study Group; Li et al., Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 44µg vs. IFN beta-1a 22µg vs. placebo: median number lesions per patient per scan: 0.17, 0.20 and 0.67, respectively, $p < 0.0001$ (all comparisons with placebo)	36 months
		Cohen et al., Neurology 2002 (IMPACT study)	SPMS (n=436)	IFN beta-1a 60mcg/week IM vs. placebo: mean number of lesions was reduced 45.6% in the IFN-1a group relative to the placebo group at month 24	24 months
		Hommes et al., Lancet Neurol 2004; Fazekas et al., Mult Scler 2005 (ESIMS study)	SPMS (n=612)	Intravenous Immunoglobulin vs. Placebo, number of lesions 2.67 vs. 3.44 after 1 year (ns), 2.45 vs. 3.01 after 2 years (ns), 4.94 vs. 6.44 overall (p=0.06)	24 months
		Freedman et al., Neurology 2011 (MAESTRO study)	SPMS (n=612)	MBP8298 vs. Placebo, cumulative number of lesions among $DR2^+$ or $DR4^+$ 1.9 vs. 1.8 after 12 months (p=0.034), among $DR2^-/DR4^-$ 1.7 vs. 2.0 after 12	24 months

				months (p=0.828)	
		Zajicek et al., Lancet Neurol 2013 (CUPID study)	PPMS (n=191), SPMS (n=302) (received treatment: n=493; randomised: n=498)	Dronabinol vs. Placebo, proportion of patients with lesions 37% vs. 40% after 3 years (p=0.70)	36 months
		Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg vs. Placebo, lesion number per year 0.13 vs. 0.50% (p<0.001); number of patients free of lesions 80% vs. 60% (p<0.001) after 36 months	36 months
	Volume of T2 lesions	European Study Group on Interferon beta-1b in Secondary Progressive MS, Lancet 1998 (EUSPMS study)	SPMS (n=718)	Interferon beta-1b vs. Placebo, percent lesion volume change -5% vs. 8% (p<0.0001) after 3 years	Early termination: obvious superiority of IFN vs. placebo (initially planned: 39 months)
		Secondary Progressive Efficacy Clinical Trial of Recombinant Interferon-beta-1a in MS (SPECTRIMS) Study Group; Li et al., Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 44µg vs. IFN beta-1a 22µg vs. placebo: Median change in burden of disease (in mm2, i.e. sum of lesional area per patient and scan, as an indirect measure of T2 lesion volume): -32 vs. -4 vs. +263, respectively, p<0.0001 for comparisons of both doses vs. placebo	36 months
		Cohen et al., Neurology 2002 (IMPACT study)	SPMS (n=436)	IFN beta-1a 60mcg/week IM vs. placebo: Median change in total T2-hyperintense lesion volume (from baseline) was reduced in the IFNb-1a group compared to the	24 months

				placebo group by 69.1% at month 24 ($p<0.001$)	
		Hommes et al., Lancet Neurol 2004; Fazekas et al., Mult Scler 2005 (ESIMS study)	SPMS (n=318)	Intravenous Immunoglobulin vs. Placebo, lesion volume 25.44cm^3 vs. 24.98cm^3 after 1 year (ns), 25.17cm^3 vs. 23.66cm^3 after 2 years (ns)	24 months
		The North American Study Group on Interferon beta-1b in Secondary Progressive MS Neurology 2004 (NASPMS study)	SPMS (n=939)	Interferon beta-1b 250µg or 160µg vs. Placebo, median percent change in annual lesion area 0.4% ($p<0.001$), 0.8% ($p<0.001$), vs. 10.9% after 3 years	Early termination for futility (initially planned: 36 months)
		Wolinsky et al., Ann Neurol 2007 (PROMISE study)	PPMS (n=943)	Glatiramer acetate vs. Placebo, percent volume change -39% after 1 year ($p=0.1716$), -71% after 2 years ($p=0.0026$), and -58% after 3 years ($p=0.1344$)	36 months
		Hawker et al., Ann Neurol 2009 (OLYMPUS study)	PPMS (n=439)	Rituximab vs. Placebo, volume change 2205mm^3 vs. 1507mm^3 ($p<0.001$) after 96 weeks	96 weeks
		Freedman et al., Neurology 2011 (MAESTRO study)	SPMS (n=612)	MBP8298 vs. Placebo, median volume change among $DR2^+$ or $DR4^+$ 417.5mm^3 vs. 491.5mm^3 after 24 months ($p=0.802$), among $DR2^-/DR4^-$ 684.8mm^3 vs. 738.0mm^3 after 24 months ($p=0.873$)	24 months
		Montalban et al., N Engl J Med.	PPMS (n=732)	Ocrelizumab 600mg (300mg	120 weeks

		2016 (ORATORIO study)		x2) /24 weeks IV vs. placebo: percent volume change: -3.4% vs. +7.4% (p<0.0001)	
Gd-enhancing lesions	**Number of Gd-enhancing lesions**	Hommes et al., Lancet Neurol 2004; Fazekas et al., MSJ 2005 (ESIMS study)	SPMS (=318)	Intravenous Immunoglobulin vs. Placebo, number of lesions 1.62 vs. 1.47 after 1 year (ns), 1.14 vs. 0.86 after 2 years (ns), 2.47 vs. 2.32 overall (ns); percent of enhancing scans 35.2% vs. 45.3% after 1 year (ns), 32.1% vs. 28.3% after 2 years (ns)	24 months
		The North American Study Group on Interferon beta-1b in Secondary Progressive MS Neurology 2004 (NASPMS study)	SPMS (n=939)	Interferon beta-1b 250µg or 160µg vs. Placebo, annual new active lesion rate 6.4 (p<0.001), 4.5 (p<0.001), vs. 18.7 after 3 years	Early termination for futility (initially planned: 36 months)
		Wolinsky et al., Ann Neurol 2007 (PROMISE study)	PPMS (n=943)	Glatiramer acetate vs. Placebo, percent change -89% after 1 year (p=0.0022), -47% after 2 years (p=0.0702), and -6% after 3 years (p=0=8387)	36 months
		Freedman et al., Neurology 2011 (MAESTRO study)	SPMS (n=612)	MBP8298 vs. Placebo, lesion change among $DR2^+$ or $DR4^+$ 1.1 vs. 0.8 after 12 months (p=0.427), among $DR2^-/DR4^-$ 0.9 vs. 1.0 after 12 months (p=0.765)	24 months
		Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg vs. Placebo, lesion number per scan 0.05 vs. 0.21 (p<0.001) after 36 months	36 months

	Number of patients with Gd-enhancing lesions	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg vs. Placebo, percentage of patients free of lesions 87% vs. 78% (p=0.006) after 36 months	36 months
T1 lesions	**New T1 lesions**	Zajicek et al., Lancet Neurol 2013 (CUPID study)	PPMS (n=191), SPMS (n=302) (randomised: n=498)	Dronabinol vs. Placebo, percentage of patients with lesions 34% vs. 33% after 3 years (p=0.87)	36 months
	New non-enhancing T1 lesions	Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg vs. Placebo, lesion number per year 0.09 vs. 0.24 (p<0.001); number of patients free of lesions 82% vs. 72% (p=0.003) after 36 months	36 months
	Volume of T1 lesions	Hommes et al., Lancet Neurol 2004; Fazekas et al., MSJ 2005 (ESIMS study)	SPMS (n=318)	Intravenous Immunoglobulin vs. Placebo, lesion volume 3.78mm^3 vs. 3.68mm^3 after 1 year (ns), 3.58mm^3 vs. 3.59mm^3 after 2 years (ns)	24 months
	T1/T2 lesion volume ratio	Hommes et al., Lancet Neurol 2004; Fazekas et al., MSJ 2005 (ESIMS study)	SPMS (n=318)	Intravenous Immunoglobulin vs. Placebo, ratio 0.136 vs. 0.131 after 1 year (ns), 0.123 vs. 0.136 after 2 years (ns)	24 months
Combined measures	**Combined unique active lesions**	Secondary Progressive Efficacy Clinical Trial of Recombinant Interferon-beta-1a in MS (SPECTRIMS) Study Group; Li et al., Neurology 2001 (SPECTRIMS study)	SPMS (n=618)	IFN beta-1a 44µg vs. IFN beta-1a 22µg vs. placebo: Median numbers of combined unique lesions: 0.11, 0.22 and 1.0, respectively, p = 0.005 (IFN beta-1a 22µg vs. placebo); p<0.0001 (IFN beta-1a 44µg vs. placebo).	36 months

Brain atrophy	**Brain parenchymal fraction**	Hawker et al., Ann Neurol 2009 (OLYMPUS study)	PPMS (n=439)	Rituximab vs. Placebo, volume change -9.9cm^3 vs. -10.8cm^3 (p=0.62) after 96 weeks	96 weeks
	Percent change	Hommes et al., Lancet Neurol 2004; Fazekas et al., MSJ 2005 (ESIMS study)	SPMS (n=318)	Intravenous Immunoglobulin vs. Placebo, percent change -0.30% vs. -0.13% after 1 year (ns), -0.11% vs. -0.06% after 2 years (ns)	24 months
		Freedman et al., Neurology 2011 (MAESTRO study)	SPMS (n=612)	MBP8298 vs. Placebo, percent change among DR2$^+$ or DR4$^+$ -1.21% vs. -0.78% after 24 months (p=0.440), among DR2$^-$/DR4$^-$ -1.23% vs. -0.62% after 24 months (p=0.942)	24 months
		Zajicek et al., Lancet Neurol 2013 (CUPID study)	PPMS (n=191), SPMS (n=302) (randomised: n=498)	Dronabinol vs. Placebo, yearly percent change -0.68% vs. -0.66% after 3 years (p=0.94)	36 months
		Lublin et al., Lancet 2016 (INFORMS study)	PPMS (n=970)	Fingolimod 0.5mg vs. Placebo, percent change -1.49% vs. -1.53% (p=0.673) after 36 months	36 months
		Montalban et al., N Engl J Med. 2016 (ORATORIO study)	PPMS (n=732)	Ocrelizumab 600mg (300mg x2) /24 weeks IV vs. placebo: rate of brain volume loss: -0.9% vs. -1.1% (p=0.0206)	120 weeks

Abbreviations: Gd: gadolinium; PPMS: primary progressive multiple sclerosis; SPMS: secondary progressive multiple sclerosis.

Table 7: Phase II and 3 trials which used spinal cord MRI outcomes

Original neuroimaging outcome	Trials	Condition (no. of patients randomised)	Drug, effect (vs. placebo/ another active arm)	Duration of the trial
Cervical cord area	Montalban et al. Mult Scler 2009, phase II	PPMS (n=49), transitional progressive MS (n=24)	Interferon beta-1b (250µg on alternate days) vs. Placebo, percent change in cord area -1.6% vs. -1.3% after 12 months (ns), -0.9% vs. -1.6% after 24 months (ns)	24 months
	Leary et al. Neurology 2003, phase II	PPMS (n=50)	Interferon beta-1a (30µg vs. 60µg per week) vs. Placebo, percent change in cord area -0.5% vs. -1.0% vs. 0.3% after 12 months (ns), -3.7% vs. 1.5% vs. -1.3% after 24 months (ns)	24 months
	Lin et al. J Neurol Neurosurg Psychiatry 2003, phase II	RRMS (n=20), SPMS (n=18)	Interferon beta-1a (44µg three times per week), percent change in cord area -1.0% vs. -1.7% after 6 months (ns), -1.5% vs. -2.8% after 12 months (ns), -1.8% vs. -2.9% after 18 months (ns), -4.5% vs. -5.7% after 48 months (ns)	48 months
	Frank et al. Mult Scler 2002, phase II	SPMS (n=6), RRMS (n=1)	RhIGF-1 (0.05 mg/kg twice a day), not reported (ns)	24 weeks
	Kapoor et al. Lancet Neurol 2010, phase II	SPMS (n=120)	Lamotrigine vs. Placebo, percent change in cord area -1.60% vs. -1.26% after 24 months (ns)	24 months
	Kalkers et al. Mult Scler 2002, phase II	PPMS (n=16)	Placebo for 12 months vs. Riluzole for following 12 months (2x50mg per day), percent change in cord area -2.0% vs. -0.2% (not reported)	24 months
	Yaldizli et al. ECTRIMS 2015, phase III (INFORMS study)	PPMS (n=823)	Fingolimod vs. Placebo, % change from baseline: percent change in cord area -2.04% vs. -2.44% after 24 months (ns)	24 months

Abbreviations: ns: not significant; PPMS: primary progressive multiple sclerosis; RRMS: relapsing-remitting multiple sclerosis; SPMS: secondary progressive multiple sclerosis

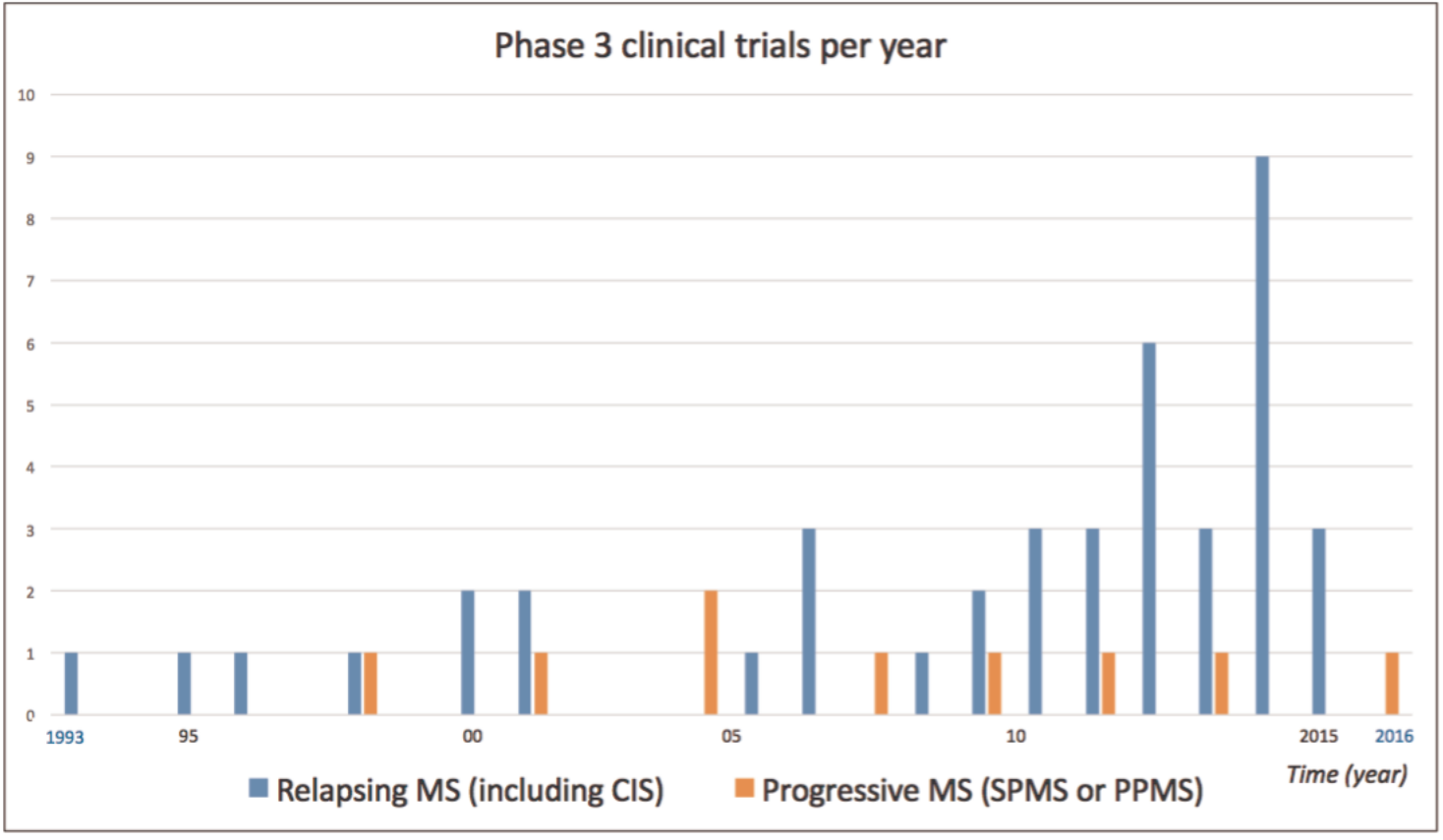
Table 8: Past and ongoing phase II and III trials which use OCT-related measures

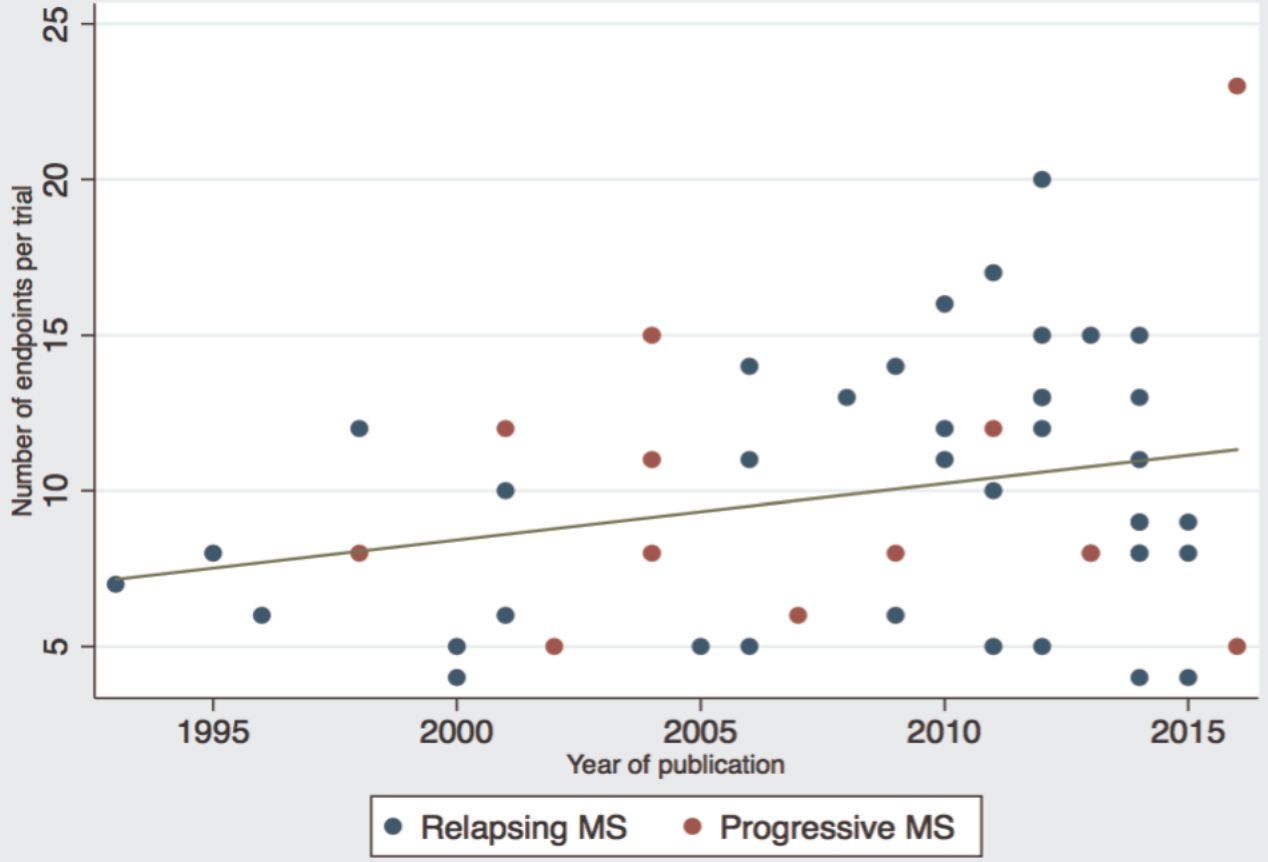
Optical Coherence Tomography

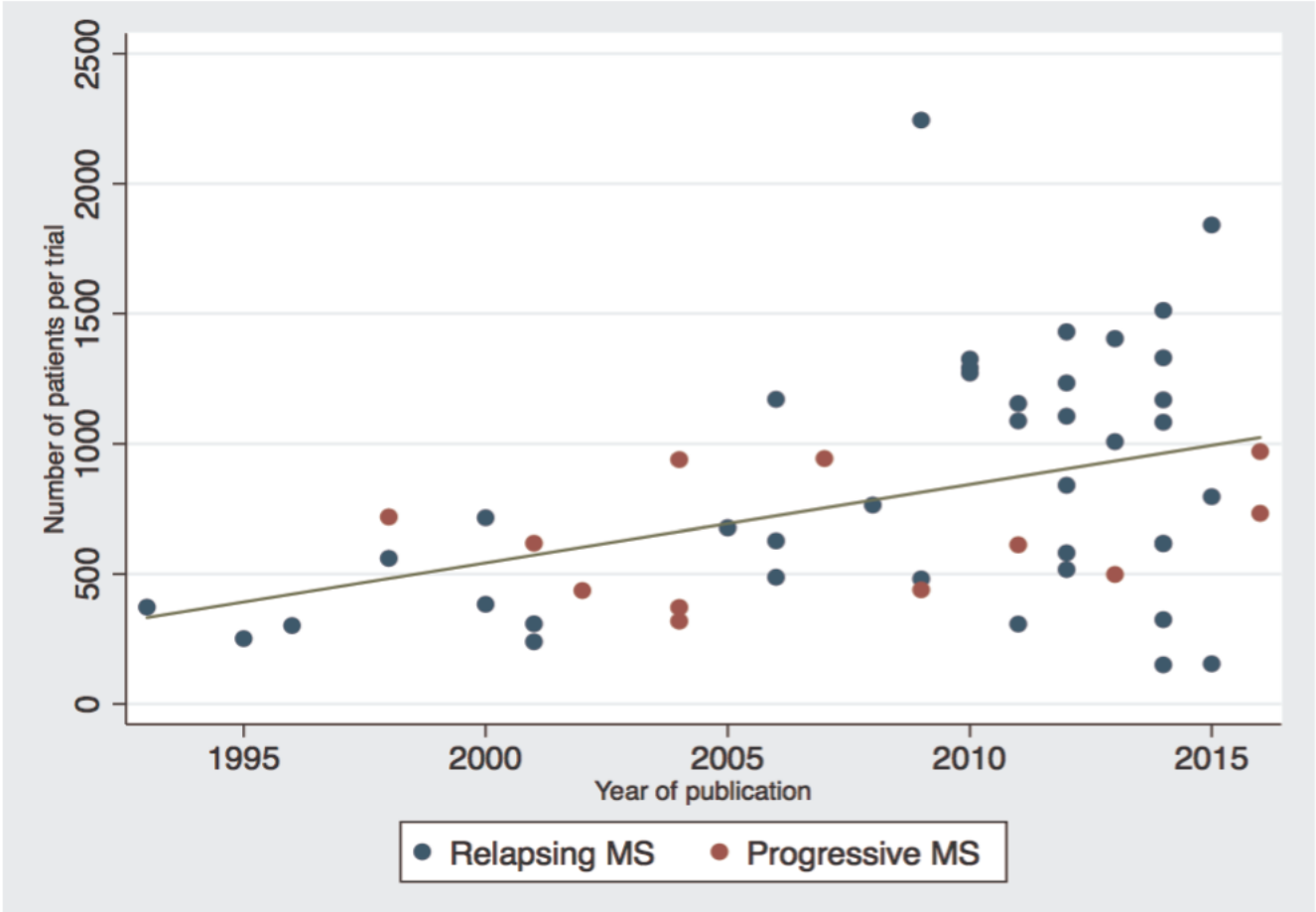
Original OCT outcome	Trial	Condition (number of patients randomised)	Drug, effect	Duration of the trial
Retinal nerve fibre layer thickness	Dorr et al. Trials 2012, phase II	RRMS and CIS (n=80)	Vitamin D (20400 IU every other day) vs. Vitamin D (400 IU every other day), ongoing	24 months
	Horton et al. Neurology 2013, phase II	>6 months after ON in RRMS (n=22)	4-aminopyridine vs. Placebo (crossover), percent change -1.89% vs. 1.45% after 5 weeks for RNFL 60-80µm (p=0.01)	10 weeks
	Cambron et al. Trials 2014, phase II	PPMS and SPMS (n=120, expected)	Fluoxetine (40mg per day) vs. Placebo, ongoing	108 weeks
	Llufriu et al. PloS ONE 2014, phase II	RRMS (n=9)	Autologous Mesenchymal Stem Cells vs. Placebo, change in thickness OD -0.2µm vs. 0.0µm (ns) and OS -0.33µm vs. -0.22µm (ns) after 6, and OD -0.02µm vs. -0.02µm (ns) and OS -0.4µm vs. 0.0µm (ns) after 12 months	12 months
	Diem et al. BMJ Open 2015, phase II	Acute ON in CIS (n=100, expected)	Erythropoietin (33000 IU per day for 3 consecutive days) vs. Placebo, ongoing	6 months
	McKee et al. BMJ Open 2015, phase II	Acute ON in CIS or in RRMS (n=46, expected)	Amiloride vs. Placebo, ongoing	12 months
	Rice et al. Trials 2015, phase II	PPMS (n=20), SPMS (n=20) (expected)	Autologous bone marrow infusion, ongoing	12 months
	Salari et al. J Res Med Sci 2015, phase II	Acute ON in CIS (n=52)	Vitamin D (50000 IU per week) vs. Placebo, change in thickness -19.9µm vs. -17.6µm (ns)	6 months
	Sergott et al. J Neurol Sci 2015, phase II	Acute ON in CIS (n=34)	Atacicept vs. Placebo, change in thickness -8.6µm vs. -17.3µm (p=0.07)	36 weeks
	Raftopoulos et al. Lancet Neurol 2016, phase II	Acute ON in CIS and RRMS (n=86)	Phenytoin vs. Placebo, 30% reduction in thickness in the extent of layer loss with Phenytoin (p=0.021)	6 months

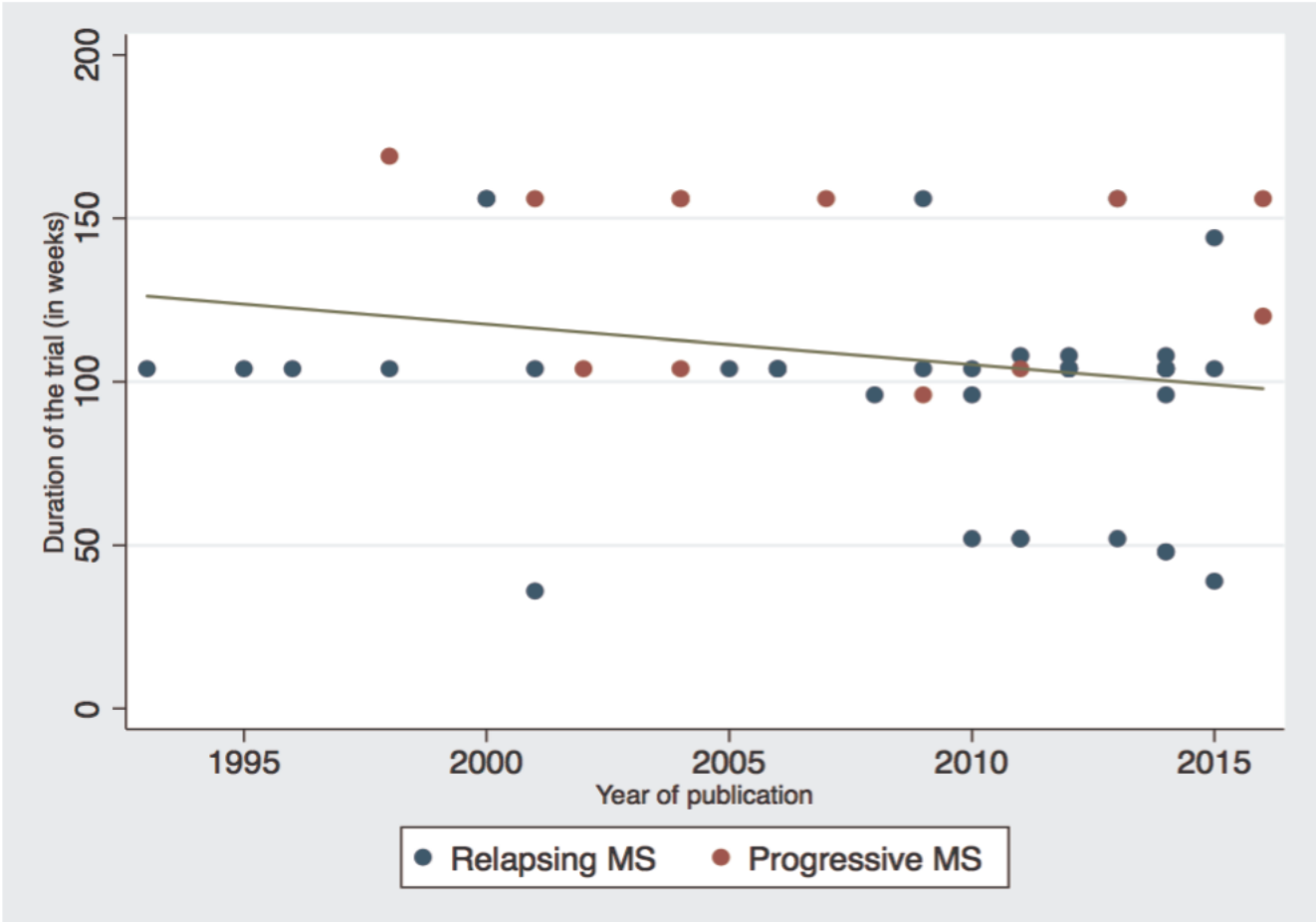
Ganglion cell layer thickness	McKee et al. BMJ Open 2015, phase II	ON in CIS or in RRMS (n=46)	Amiloride vs. Placebo, ongoing	12 months
Macular volume	Dorr et al. Trials 2012, phase II	RRMS and CIS (n=80)	Vitamin D (20400 IU every other day) vs. Vitamin D (400 IU every other day), ongoing	24 months
	Zarbin et al Ophthalmlogy 2013, phase II and phase III (pooled data analysis)	RRMS (n=2615)	Fingolimod, macular oedema detection	5 years
	Cambron et al. Trials 2014, phase II	PPMS and SPMS (n=120, expected)	Fluoxetine (40mg per day) vs. Placebo, ongoing	108 weeks
	Llufriu et al. PloS ONE 2014, phase II	RRMS (n=9)	Autologous Mesenchymal Stem Cells vs. Placebo, volume change OD -0.02mm^3 vs. 0.0mm^3 (ns) and OS -0.02mm^3 vs. -0.02mm^3 (ns) after 6, and OD -0.02mm^3 vs. 0.0mm^3 (ns) and OS -0.01mm^3 vs. 0.01mm^3 (ns) after 12 months	12 months
	Diem et al. BMJ Open 2015, phase II	Acute ON in CIS (n=100, expected)	Erythropoietin (33000 IU per day for 3 consecutive days) vs. Placebo, ongoing	6 months
	McKee et al. BMJ Open 2015, phase II	ON in CIS or in RRMS (n=46, expected)	Amiloride vs. Placebo, ongoing	12 months
	Rice et al. Trials 2015, phase II	PPMS (n=20), SPMS (n=20) (expected)	Autologous bone marrow infusion, ongoing	12 months
	Raftopoulos et al. Lancet Neurol 2016, phase II	Acute ON in CIS and RRMS (n=86)	Phenytoin vs. Placebo, 34% volume reduction in the extent of volume loss with Phenytoin (p=0.005)	6 months
Macular thickness	McKee et al. BMJ Open 2015, phase II	ON in CIS or in RRMS (n=46, expected)	Amiloride vs. Placebo, ongoing	12 months
	Salari et al. J Res Med Sci 2015, phase II	Acute ON in CIS (n=52)	Vitamin D (50000 IU per week) vs. Placebo, thickness change -0.8µm vs. -3.1µm (ns)	6 months

Abbreviations: CIS: clinically isolated syndrome; ns: not significant; ON: optic neuritis; PPMS: primary progressive multiple sclerosis; RRMS: relapsing-remitting multiple sclerosis; SPMS: secondary progressive multiple sclerosis.


Phase 3 clinical trials per year
10
9
8
7
6
5
4
3
2
1
0
1993
95
00
05
10
2015
2016
Time (year)
Relapsing MS (including CIS)
Progressive MS (SPMS or PPMS)


Number of endpoints per trial
5
10
15
20
25
1995
2000
2005
2010
2015
Year of publication
Relapsing MS
Progressive MS


Number of patients per trial
0
500
1000
1500
2000
2500
1995
2000
2005
2010
2015
Year of publication
Relapsing MS
Progressive MS


Duration of the trial (in weeks)
0
50
100
150
200
1995
2000
2005
2010
2015
Year of publication
Relapsing MS
Progressive MS