SYSTEMATIC REVIEW

CrossMark

Disease-Modifying Therapies for Relapsing–Remitting Multiple Sclerosis: A Network Meta-Analysis

Rosa C. Lucchetta[1] · Fernanda S. Tonin[1] · Helena H. L. Borba[1] · Letícia P. Leonart[1] · Vinicius L. Ferreira[1] · Aline F. Bonetti[1] · Bruno S. Riveros[1] · Jefferson Becker[2] · Roberto Pontarolo[1] · Fernando Fernandez-Llimós[3] · Astrid Wiens[1]

Abstract

Background A broad range of disease-modifying therapies (DMTs) for relapsing–remitting multiple sclerosis (RRMS) is available. However, the efficacy and safety of traditional DMTs compared with the recently developed DMTs remain unclear.

Objective Therefore, we have synthesised available evidence of clinical outcomes for DMTs in adults with RRMS.

Methods PubMed, Scopus and a manual search were performed. Bayesian network meta-analyses of randomised clinical trials assessing DMTs as monotherapies were conducted. SUCRA and GRADE were used to rank therapies and to assess quality of general evidence, respectively.

Results Thirty-three studies were included in the meta-analyses. The most effective therapies for the outcome of annualised relapse rate were alemtuzumab (96% probability), natalizumab (96%) and ocrelizumab (85%), compared with all other therapies (hazard ratio versus placebo, 0.31, 0.31 and 0.37, respectively; $p < 0.05$ for all comparisons) (high-quality evidence). However, no significant differences among these three therapies were found. Discontinuation due to adverse events revealed similarity across all therapies, except for alemtuzumab, which showed less discontinuation when compared with interferon-1a intramuscular (relative risk 0.37; $p < 0.05$).

Conclusion High-quality evidence shows that alemtuzumab, natalizumab and ocrelizumab present the highest efficacy among DMTs, and other meta-analyses are required regarding adverse events frequency, to better understand the safety of therapies. Based on efficacy profile, guidelines should consider a three-category classification (i.e. high, intermediate and low efficacy).

✉ Astrid Wiens
astrid@ufpr.br

1 Pharmacy, Federal University of Paraná, Curitiba, Brazil

2 Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil

3 Department of Social Pharmacy, Faculty of Pharmacy, Research Institute for Medicines (iMed.ULisboa), University of Lisbon, Lisbon, Portugal

Key Points

Alemtuzumab, natalizumab and ocrelizumab are the most effective therapies for the annualised relapse rate in the treatment of relapsing–remitting multiple sclerosis (RRMS), among disease-modifying therapies.

RRMS guidelines should consider a three-category classification: high efficacy (i.e. alemtuzumab, natalizumab, ocrelizumab), intermediate efficacy (i.e. cladribine, fingolimod, dimethyl fumarate) and low efficacy (i.e. peginterferon, glatiramer acetate, interferons and teriflunamide).

It seems that high- and intermediate-efficacy therapies could be the first-line treatment for patients with more severe RMSS conditions.

Adis

1 Introduction

According to the Multiple Sclerosis International Federation, 2.3–2.5 million people had multiple sclerosis (MS) in 2013 (2.1–140/100,000 residents) [1, 2]. MS is classified into four major phenotypes: relapsing–remitting multiple sclerosis (RRMS), primary progressive MS (PPMS), secondary progressive MS (SPMS) and clinically isolated syndrome (CIS) [3]. RRMS is the most frequent, representing 80–85% of new cases of MS [4, 5].

RRMS is characterised by symptomatic relapse at irregular intervals, interspersed with periods of remission in which there is total or partial recovery of the patient [6]. To reduce the frequency and severity of the relapse, and to delay disease progression, decrease the number of lesions in the central nervous system and maintain or improve patients' quality of life, RRMS treatment should comprise disease-modifying therapies (DMTs) such as alemtuzumab, cladribine, dimethyl fumarate, fingolimod, glatiramer acetate, interferons, mitoxantrone, natalizumab, ocrelizumab, peginterferon or teriflunomide. The Association of British Neurologists (ABN) [7], in guidelines updated in 2015, recommends starting the treatment with a DMT of moderate efficacy (category 1) such as interferons, glatiramer, teriflunomide, dimethyl fumarate and fingolimod. Only for patients in high disease activity, or who do not respond or tolerate a category 1 DMT, are drugs of high efficacy (category 2) such as alemtuzumab and natalizumab recommended. The ABN guideline does not include new therapies such as cladribine and ocrelizumab [7]. These therapies were also not included in previous published meta-analyses with low risk of bias [8]. Network meta-analyses are recommended by the International Society for Pharmacoeconomics and Outcome Research to compare different treatments simultaneously [9, 10]. Thus, we aimed to conduct a network meta-analysis of randomised clinical trials (RCTs) to provide evidence-based hierarchies of the efficacy and safety of all available DMTs for patients with RRMS.

2 Methods

This systematic review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Network Meta-Analyses (PRISMA NMA) [11] and Cochrane Collaboration recommendations [12] and is registered with the International Prospective Register of Systematic Reviews (PROSPERO), number CRD42017059120 [13].

2.1 Search Strategy and Selection Criteria

Electronic searches were conducted in the PubMed and Scopus databases without any time limit or language restriction (updated in May 2017). Trial registration databases (ClinicalTrials) and the reference lists of reviews and included studies were also searched. Complete search strategies are provided in the electronic supplementary material (ESM), p 2.

We included studies that fulfilled the following inclusion criteria: randomised, phase II or later controlled trials (including post hoc analyses) that assessed the efficacy, safety or quality of life (QoL) of a DMT as monotherapy (head-to-head or against placebo) in adults diagnosed with RRMS. The searched DMT therapies were alemtuzumab, 12 mg per day for 5 days (first course) and for 3 days (second course), with a 1-year interval between each course (ALE12 and ALE24) intravenous (IV); azathioprine (AZA) orally (PO); cladribine cumulative dose 3.5 mg, 5.25 mg per kg (CLA3.5 and CLA5.25) PO or (CLA2.45) subcutaneous (SC); daclizumab 150 mg and 300 mg every 4 weeks (DAC150Q4W and DAC300Q4W) IV; dimethyl fumarate 240 mg twice a day or three times a day (BG240BID and BG240TIW) PO; fingolimod 0.5 mg, 1.25 mg and 5 mg per day (FING0.5QD, FING1.25QD and FING5QD) PO; glatiramer acetate 20 mg, 40 mg per day and 40 mg three times a week (GA20QD, GA40QD and GA40TIW) SC; interferon β-1a 30 μg or 60 μg each week (IFNA30QW and IFNA60QW) intramuscular (IM); interferon β-1a 44 or 22 μg three times a week (IFNA44TIW and IFNA22TIW) SC; interferon β-1b 50, 250, 375 or 500 μg, every other day (IFNB50EOD, IFNB250EOD, IFNB375EOD and IFNB500EOD) SC; pegylated interferon 125 μg every 2 or 4 weeks (PIFN125Q2W and PIFN125Q4W) SC; natalizumab 300 mg every 4 weeks (NAT300Q4W) IV; ocrelizumab 600 mg and 2000 mg every 6 months (OCRE600Q6M and OCRE2000Q6M), IV; rituximab (RTX), IV and teriflunomide 7 mg and 14 mg per day (TERI7QD and TERI14QD), PO. The considered outcomes included the annualised relapse rate (ARR), disability progression confirmed at 12 weeks (DPC12), disability progression confirmed at 24 weeks (DPC24), disability improvement confirmed at 12 weeks (DIC12), disability improvement confirmed at 24 weeks (DIC24), discontinuations due to adverse events (DAE) and change in QoL evaluated through Short Form-36 items or 12 items (SF-36 or SF-12). Studies with a follow-up of < 12 weeks or evaluating RRMS with other forms of MS were excluded.

Two researchers independently screened the titles and abstracts of retrieved studies to identify irrelevant records. In a second stage, full-text articles were also independently

evaluated by two researchers according to defined inclusion and exclusion criteria. Discrepancies were reconciled in consensus meetings, using a third researcher as a referee.

2.2 Data Analysis

The following data were independently extracted by two researchers by using Microsoft© Office Excel©: (1) study baseline characteristics (authors' names, year of publication, country, sample size, patients' sex and age, disease duration, onset of symptoms and follow-up, evaluated DMT therapies), (2) methodological aspects (e.g. trial design); (3) clinical outcome results (efficacy, safety or QoL).

The critical evaluation of risk of bias of the included studies was conducted by two independent reviewers, using the Cochrane Collaboration revised Risk of Bias (RoB 2.0) assessment tool [14]. In the absence of consensus, points of disagreement were resolved by the opinion of a third researcher.

Statistical analyses were performed using software R v. 3.4.1/R studio 1.0.153 [15], packages READR [16], META [17], METAFOR [18], GeMTC [19], RJAGS [20] and CODA [21]. Transitivity analyses were performed by comparing population, interventions and control and outcome definitions among the included studies in the meta-analyses. Transitivity was assumed for minor differences in follow-up times, and 48 or 52 weeks were considered to correspond to 1-year follow-up; 96–108 weeks were considered to correspond to 2-year follow-up. Both pairwise and network meta-analyses were performed. Effect size measures were defined for each outcome as follows: hazard ratio for the ARR outcome; relative risk for dichotomous outcomes (DPC12 and 24 weeks, DIC12, DAE); mean difference for the changes in the QoL scores (SF-36 or SF-12).

Pairwise meta-analyses for the outcome ARR were assessed using the Poisson method. Dichotomous outcomes (DPC12 and 24 weeks, DIC12, DAE) were analysed using the Mantel–Haenszel method. Changes in the QoL scores (SF-36 or SF-12) were evaluated through inverse variance. DerSimonian-Laird estimator of τ^2 was employed in all analyses. All effect size measures were calculated considering 95% confidence intervals (CI). Data entry was performed with contrast-based data. Heterogeneity was evaluated using Higgis inconsistency analyses (I^2) [12]. Sensitivity analyses, using adjustment of the random effects model by Hartung–Knapp and Sidik–Jonjman estimator for τ^2, were carried out.

Network meta-analyses, using a Bayesian framework for each outcome based on the Markov Chain Monte Carlo simulation method, were performed. Arm-level entry data was used. For the inclusion of multiple-arm studies, correlations for the likelihood between arms were considered. A common

heterogeneity parameter was assumed for all comparisons [22]. We opted for a conservative analysis of non-informative priors [23]. Effect size measures were expressed with a 95% credibility interval (CrI). Both fixed- and random-effect models were tested, and the one with the lowest deviance information criteria (DIC) was selected. Convergence was attained based on visual inspection of Brooks-Gelman-Rubin plots and potential scale reduction factor (PSRF) ($1 < PSRF \leq 1.05$). To increase the estimate precision of the relative effect sizes of comparisons and to properly account for correlations between multi-arm trials, ranking probabilities for each outcome were calculated via surface under the cumulative ranking analysis (SUCRA). To estimate the robustness of the network, inconsistency, defined as the difference between the pooled direct and indirect evidence for a comparison, was assessed using node-splitting analysis [24]. Sensitivity analyses with the hypothetical removal or inclusion of the studies were conducted when discrepancies were identified in the network meta-analyses: (1) first scenario: original analyses; (2) second scenario: removal of studies with high risk of bias; (3) third scenario: inclusion of studies with non-approved therapies; (4) fourth scenario: removal of studies with suspicion of impairing transitivity due to important differences in patients' characteristics (e.g. age, Expanded Disability Status Score [EDSS], disease duration, onset of symptoms and previous DMT experience). When possible, subgroup (e.g. population) analyses also were performed.

The quality of the evidence was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group [25] for ARR and DAE96, which were classified as high, moderate, low and very low. IFNA44TIW was assumed as the common comparator.

2.3 Role of Funding Source

3 Results

Our systematic review identified 2797 records after removal of duplicates; 2596 were considered irrelevant during the screening, and 152 were excluded in the full-text appraisal (ESM pp 3–8). The remaining 49 records represent 40 RCTs included in systematic review and 37 in meta-analysis. The included articles were published between 1995 and 2017,

with a median in 2011. Most of the studies in systematic review were multicentric, conducted in more than one country (*n* = 38) and included 29,150 participants (median of 196; interquartile range: 111–417), 66% of whom were women. Eight studies included only treatment-naive participants, and one study assessed only treatment-experienced patients; 16 studies included both treatment-naive and treatment-experienced patients, and 15 articles did not report this information. Altogether, 15 approved dosages of DMT were identified, with 16 clinical trials comparing active therapies (head-to-head trials), 14 comparing different doses of DMT and 10 evaluating the active treatment against placebo. No study evaluating azathioprine or rituximab fulfilled the inclusion criteria and could be included in the systematic review. Most of the studies had a follow-up of 96 weeks (median 96; interquartile range: 48–96). The main characteristics of the included studies are presented in Table 1 (supplemental characteristics are presented in the ESM, p 9).

The methodological quality assessed by RoB 2.0 is presented in the ESM, pp 10–14. The outcomes more frequently associated with 'low risk of bias' were disability improvement and disability progression confirmed at 12 weeks; 'some concerns' appear more frequently in disability progression confirmed at 24 weeks, whereas 'high risk' was associated with QoL and ARR outcomes. The two domains more frequently scored as 'high risk of bias' were measurement of the outcome (due to the lack of masking of the assessors) and domain referring to missing outcome data.

Network diagrams of the possible comparisons for evaluated outcomes in the first scenario (original analysis) are presented in Fig. 1. (Studies included in each network meta-analysis are presented in the ESM, p 15). It was not possible to build a single network for the DIC12 or QoL outcomes (SF-36 or SF-12). Baseline EDSS, disease duration or onset of symptoms were assumed to be homogeneous, because sensitivity analyses that excluded studies suspected of impairing transitivity exhibited similar results. A fixed-effects model was selected for the ARR and DCP24 network meta-analyses, whereas a random-effects model was selected for other DCP12 and DAE analyses.

The network meta-analysis of ARR (Fig. 1a) included 32 studies (*n* = 38,298 patient-years). All therapies were statistically superior to placebo. By SUCRA analysis, ALE12 presented the highest probability of being the best alternative for this outcome (96% probability), followed by NAT300Q4W with 96%, and OCRE600Q6M with 85%. On the other hand, TERI7QD and IFN30QW appeared to be the worst therapies (23% and 7%, respectively, of probability of ranking first) (see Table 2, Fig. 2).

The DPC12 network meta-analysis (Fig. 1b) (*n* = 16 studies; 13,510 patients) revealed that ALE12 and OCRE600Q6M were significantly more efficacious than other therapies (94 and 88% of probabilities, respectively).

The DMTs IFNB250EOD and GA20QD were the worst treatments for this outcome (20 and 24% probability, respectively) (see ESM, p 16).

The DPC24 analyses (Fig. 1c) (*n* = 16 trials; 13,410 patients) presented IFNB250EOD (93% probability) and ALE12 (76% probability) as more efficacious treatments, whereas GA40TIW (5%) and IFNA30QW (22%) were the worst options (see ESM, p. 16). Considering the possibility that the INCOMIN (Independent Comparison of Interferon) [26] study was an outlier in this analysis, after the removal of this study, NAT300Q4W became the most effective (87% probability), followed by ALE12 (82% probability) and OCRE600Q6M (77% probability) (see ESM, p. 17).

The network of DAE96 (Fig. 1d) included 17 trials with 12,221 patients. ALE12 and OCRE600Q6M were considered the best therapies for this outcome (85 and 67% probability, respectively), whereas IFN44TIW and CLA3.5 (22 and 38%, respectively) were ranked last in the SUCRA analyses (see Table 2). It is noteworthy that a statistical difference was found for only one comparison in this outcome (ALE12 and IFNA44TIW, with relative risk [95% CrI] 0.37 [0.17–0.81]).

Pairwise meta-analyses results confirm the results obtained in the network meta-analyses (see ESM, p. 18). In the pairwise meta-analyses, the change from the original statistical method to the Hartung-Knapp method caused the enlargement of the confidence intervals. Consequently, the statistically significant difference between most of the comparisons was lost, except for ARR (ESM, p 19).

The node-splitting technique revealed that no substantial differences in the magnitude or direction between the results of the direct and indirect effects were identified in the network meta-analyses (ESM, p. 20). The sensitivity analyses, using different scenarios of network meta-analyses (scenarios II–IV), found no significant differences compared with the original scenario (ESM, p. 21).

Several studies reported results for patients with more aggressive forms of RRMS. However, due to the inconsistent reports of outcome results, it was not possible to perform subgroup meta-analyses (both pairwise and network) to synthesise evidence.

Table 3 presents the results of the quality of evidence assessment (GRADE) for the ARR and DAE outcomes, with therapies listed by rank order. For ARR, the most efficacious therapies compared with IFNA44TIW presented high-confidence evidence, whereas in the comparison with the least efficacious therapies, the confidence varied from low to moderate due to imprecision and presence of methodological bias. For the outcome of DAE, most comparisons presented high or moderate confidence, the latter downgraded by the presence of methodological bias. None of the comparisons in either outcome was affected by intransitivity.

Table 1 Characteristics of the included studies in the systematic review

Study acronym/ first author and year NCT	Year	Follow-up, weeks	Evaluated alternatives	# Participants (# Women)	Age, mean in years (SD)	Baseline EDSS, mean (SD)	Disease duration, mean in years (SD)	Symptom onset, mean in years (SD)	Previous DMT
ADVANCE NCT00906399	2014	48	PIFN125Q2W	512 (361)	36.9 (9.8)	2.5 (1.3)	4.0 (5.1)	6.9 (6.6)	17%
			PIFN125Q4W	500 (352)	36.4 (9.9)	2.5 (1.2)	3.4 (4.4)	6.5 (6.1)	
			PLA	500 (358)	36.3 (9.7)	2.4 (1.2)	3.5 (4.6)	6.3 (6.3)	
AFFIRM NCT00027300	2006	96	NAT300Q4W	627 (449)	35.6 (8.5)	2.3 (1.2)	5.0 (0–34)[a]	NR	NR
			PLA	315 (211)	36.7 (7.8)	2.3 (1.2)	6.0 (0–33)[a]		
BEYOND NCT00099502	2009	96	IFNB250EOD	897 (627)	35.8 (28–43)[b]	2.35 (1.5–3.0)[b]	5.4 (3)	NR	Naive
			IFNB500EOD	899 (629)	35.9 (28–43)[b]	2.33 (1.5–3.0)[b]	5.3 (3)		
			GA20QD	448 (306)	35.2 (27–43)[b]	2.28 (1.5–3.0)[b]	5.1 (3)		
BRAVO NCT00605215	2014	96	INFA30QW	447 (307)	38.5 (30.3–45.9)[b]	2.5 (1.5–3.5)[b]	1.4 (0.3–4.7)[b]	5.3 (2.4; 10.3)[b]	9%
			PLA	450 (321)	37.5 (30.3–45.4)[b]	2.5 (1.5–3.5)[b]	1.2 (0.3–4.0)[b]	4.7 (2.0; 9.7)[b]	6%
Calabrese, 2012	2012	96	IFNA44TIW	46 (32)	35.9 (9.1)	1.9 (1.0)	5.7 (4.9)	NR	Naive
			IFNA30QW	47 (32)	34.8 (9.6)	1.9 (0.8)	5.3 (5.1)		
			GA20QD	48 (35)	38.9 (10.2)	2.1 (1.1)	5.5 (6.1)		
CAMMS223 NCT00050778	2008	144	IFNA44TIW	111 (71)	32.8 (8.8)	1.9 (0.8)	NR	1.4 (0.2; 6.3)[a]	Naive
			ALE12	112 (72)	31.9 (8.0)	1.9 (0.7)	NR	1.3 (0.1; 3.5)[a]	
			ALE24	110 (71)	32.2 (8.8)	2.0 (0.7)	NR	1.2 (0.3; 3.2)[a]	
CARE-MS I NCT00530348	2012	96	IFNA44TIW	187 (122)	33.2 (8.5)	2.0 (0.8)	NR	2.0 (1.3)	Naive
			ALE12	376 (243)	33.0 (8.0)	2.0 (0.8)	NR	2.1 (1.4)	
CARE-MS II NCT00548405	2012	96	IFNA44TIW	202 (131)	35.8 (8.8)	2.7 (1.2)	NR	4.7 (2.9)	100%
			ALE12	426 (281)	34.8 (8.4)	2.7 (1.3)	NR	4.5 (2.7)	
			ALE24	170 (120)	35.1 (8.4)	2.7 (1.2)	NR	4.3 (2.8)	
Clanet, 2002[g]	2002	144	IFNA60QW	400 (272)	36.7 (7.9)	3.6 (1.0)	6.5 (5.3)	NR	NR
			IFNA30QW	402 (273)	36.9 (7.9)	3.6 (1.0)	6.6 (5.6)		
CLARITY NCT00213135	2010	96	CLA3.5	433 (298)	37.9 (10.3)	2.8 (1.2)	7.9 (7.2)	NR	26%
			CLA5.25	456 (312)	39.1 (9.9)	3.0 (1.4)	9.3 (7.6)		32%
			PLA	437 (288)	38.7 (9.9)	2.9 (1.3)	8.9 (7.4)		33%
CMSSG NCT00004814	1995	96	GA20QD	125 (88)	34.6 (6.0)	2.8 (1.2)	7.3 (4.9)	NR	NR
			PLA	126 (96)	34.3 (6.5)	2.4 (1.3)	6.6 (5.1)		
Cohen, 2007[f]	2007	36	GA40QD	46 (37)	37.4 (6.5)	2.1 (1.0)	3.8 (4.8)	6.7 (6.4)	NR
			GA20QD	44 (31)	37.1 (7)	2.0 (1.2)	3.2 (3.7)	7.4 (6.2)	
CombiRx NCT00211887	2013	144	IFNA30QW	250 (173)	37.6 (10.2)	2.0 (1.2)	1.4 (4.0)	NR	NR
			GA20QD	259 (185)	39.0 (9.5)	1.9 (1.2)	1.0 (2.9)		
CONFIRM NCT00451451	2012	96	BG240BID	359 (245)	37.8 (9.4)	2.6 (1.2)	4.9 (5.1)	NR	28%
			BG240TID	345 (250)	37.8 (9.4)	2.5 (1.2)	4.6 (5.2)		29%
			GA20QD	350 (247)	36.7 (9.1)	2.6 (1.2)	4.4 (4.7)		29%
			PLA	363 (251)	36.9 (9.2)	2.6 (1.2)	4.8 (5.0)		31%
CORAL[g] NCT00337779	2011	48	GA20QD	586 (421)	36.3 (9.0)	2.2 (1.2)	3.0 (4.0)	NR	NR
			GA40QD	569 (407)	36.3 (9.0)	2.1 (1.1)	3.3 (4.8)		
Tramacere NCT01064401	2015	96–144	IFNA30QW	922 (627)	36.2 (9.3)	2.5 (1.3)	4.1 (4.7)	6.9 (6.3)	41%
			DAC150Q4W	919 (625)	36.4 (9.4)	2.5 (1.2)	4.2 (5.0)	7.0 (6.3)	41%

Table 1 (continued)

Study acronym/ first author and year NCT	Year	Follow-up, weeks	Evaluated alternatives	# Participants (# Women)	Age, mean in years (SD)	Baseline EDSS, mean (SD)	Disease duration, mean in years (SD)	Symptom onset, mean in years (SD)	Previous DMT
DEFINE NCT00420212	2012	96	BG240BID	410 (296)	38.1 (9.1)	2.4 (1.3)	5.6 (5.4)	NR	40%
			BG240TID	416 (306)	38.8 (8.8)	2.4 (1.2)	5.1 (5.3)		40%
			PLA	408 (306)	38.5 (9.1)	2.5 (1.2)	5.8 (5.8)		42%
DMSG	2006	96	IFNA22QW	143 (95)	37 (18–55)[a]	3.0 (0–5.5)[a]	6.0 (0–35)[a]	NR	NR
			IFNB250EOD	158 (99)	38 (19–55)[a]	2.8 (0–5.5)[a]	6.0 (0–31)[a]		
Durelli, 2008[f]	2008	24	IFNB250EOD	40 (27)	33.3 (8.6)	2.0 (0.9)	6.3 (4.8)	NR	Naive
			IFNB375EOD	36 (27)	33.7 (7.3)	2.0 (1.0)	5.7 (3.6)		
ECGA	2001	36	PLA	120 (NR)/114 (83)	34.4 (7.4)	2.4 (1.2)	8.3 (5.6)	NR	NR
			GA20QD	190 (NR)/113 (87)	34.0 (7.6)	2.3 (1.1)	8.0 (5.6)		
EVIDENCE NCT00292266	2007	96–192	IFNA44TIW	339 (254)	38.3 (18–55)[a]	2.0 (NR)[a]	4.0 (NR)[a]	NR	Naive
			IFNA30QW	338 (252)	37.4 (18–55)[a]	2.0 (NR)[a]	4.1 (NR)[a]		
FREEDOMS NCT00289978	2010	96	PLA	418 (298)	37.2 (8.6)	2.5 (1.3)	NR	8.1 (6.4)	59%
			FING0.5QD	425 (296)	36.6 (8.8)	2.3 (1.3)		8.0 (6.6)	43%
			FING1.25QD	429 (295)	37.4 (8.9)	2.4 (1.4)		8.4 (6.9)	40%
FREEDOMS II NCT00355134	2014	96	PLA	355 (288)	40.1 (8.4)	2.4 (1.3)	NR	10.6 (7.9)	73%
			FING0.5QD	358 (275)	40.6 (8.4)	2.4 (1.3)		10.4 (8.0)	74%
			FING1.25QD	370 (281)	40.9 (8.9)	2.5 (1.3)		10.8 (8.2)	78%
GALA	2013	48	PLA	461 (313)	38.1 (9.2)	2.7 (1.2)	NR	7.6 (6.4)	14%
			GA40TIW	943 (641)	37.4 (9.4)	2.8 (1.2)		7.7 (6.7)	14%
GATE NCT01489254	2015	36	GA20QD[c]	357 (238)	33.8 (9.0)	2.7 (1.2)	NR	6.4 (6.0)	NR
			PLA	84 (57)	32.6 (8.7)	2.7 (1.2)		5.7 (6.0)	
GLACIER[f] NCT01874145	2015	16	GA20QD	101 (83)	50.4 (9.3)	2.4 (1.4)	12.1 (10.0)	16.2 (11.0)	NR
			GA40TIW	108 (89)	50.9 (11.0)	2.5 (1.4)	10.8 (8.6)	15.7 (11.1)	
IMPROVE NCT00441103	2012	16	IFNA44TIW	120 (NR)	NR	NR	NR	NR	NR
			PLA	60 (NR)					
INCOMIN	2002	96	IFNA30QW	92 (57)	34.9 (7.9)	1.96 (0.7)	6.7 (5.4)	NR	NR
			IFNB250EOD	96 (66)	38.8 (7.1)	1.97 (0.7)	5.9 (4.2)		
Kappos, 2011 NCT00676715	2011	48	PLA	54 (36)	38.0 (8.8)	NR	2.7 (0.1–19.2)[a]	4.8 (0.6; 26.2)[a]	30%
			IFNA44TIW	54 (38)	38.1 (9.3)		3.3 (0.1–20.2)[a]	5.3 (0.8; 35.2)[a]	31%
			OCRE600Q6M	55 (35)	35.6 (8.5)		3.6 (0.1–16.5)[a]	6.5 (0.5; 20.5)[a]	53%
			OCRE2000Q6M	55 (32)	38.5 (8.7)		4.4 (0.1–19.2)[a]	7.7 (0.3; 28.0)[a]	51%
MSCRG	1996	96	PLA	143 (103)	36.9 (0.6)	2.3 (0.1)	6.4 (0.5)	NR	Naive
			IFNA30QW	158 (118)	36.7 (0.6)	2.4 (0.1)	6.6 (0.5)		
OPERA I NCT01247324	2017	96	IFNA44TIW	411 (272)	36.9 (9.3)	2.9 (1.2)	3.71 (4.63)	6.25 (5.98)	29%
			OCRE600Q6M	410 (270)	37.1 (9.3)	2.8 (1.3)	3.82 (4.8)	6.74 (6.37)	26%
OPERA II NCT01412333	2017	96	IFNA44TIW	418 (280)	37.4 (9.0)	2.8 (1.3)	4.13 (5.07)	6.68 (6.13)	25%
			OCRE600Q6M	417 (271)	37.2 (9.1)	2.8 (1.4)	4.15 (4.95)	6.72 (6.10)	27%

Table 1 (continued)

Study acronym/ first author and year NCT	Year	Follow-up, weeks	Evaluated alternatives	# Participants (# Women)	Age, mean in years (SD)	Baseline EDSS, mean (SD)	Disease duration, mean in years (SD)	Symptom onset, mean in years (SD)	Previous DMT
PRISMS	1998	96	IFNA22TIW	189 (67)	34.8 (29.3–39.8)[b]	2.5 (1.2)	5.4 (3.0–11.2)[b]	NR	NR
			IFNA44TIW	184 (66)	35.6 (28.4–41.0)[b]	2.5 (1.3)	6.4 (2.9–10.3)[b]		
			PLA	187 (75)	34.6 (28.8–40.4)[b]	2.4 (1.2)	4.3 (2.4–8.4)[b]		
REGARD NCT00078338	2008	96	IFNA44TIW	386 (267)	36.7 (9.8)	2.4 (1.3)	NR	5.93 (6.25)	NR
			GA20QD	378 (272)	36.8 (9.5)	3.2 (1.3)		6.55 (7.10)	
Saida, 2017 NCT01440101	2017	24	PLA	47 (32)	35.1 (8.2)	2.1 (1.5)	5.1 (4.9)	6.8 (5.5)	85%
			NAT300Q4W	47 (34)	37.7 (8.6)	2.5 (1.6)	5.9 (5.0)	8.7 (5.7)	91%
SELECT NCT00390221	2013	52	PLA	204 (128)	36.6 (9.0)	2.7 (1.2)	2.0 (1–6)[b]	NR	24%
			DAC150Q4W	208 (140)	35.5 (8.9)	2.8 (1.2)	3.0 (1–7)[b]		25%
			DAC300Q4W	209 (134)	35.2 (8.7)	2.7 (1.2)	3.0 (1–6)[b]		22%
Stelmasiak, 2009[g]	2009	48	PLA	44 (27)	NR	4.0 (3.4–4.5)[d]	NR	NR	Naive
			CLA5	40 (22)		3.8 (3.3–4.4)[d]			
TEMSO (Cohort RRMS) NCT00134563	2011	96	PLA TERI7QD TERI14QD	995 (NR)	NR	NR	NR	NR	NR
TENERE NCT00883337	2014	48	IFNA44TIW[e]	104 (71)	37.0 (10.6)	2.0 (1.2)	NR	7.7 (7.6)	24%
			TERI7QD[e]	109 (70)	35.2 (9.2)	2.0 (1.2)		7.0 (6.9)	21%
TRANS-FORMS NCT00340834	2010	48	IFNA30QW	435 (295)	36.0 (8.3)	2.2 (1.3)	NR	7.4 (6.3)	56%
			FING0.5QD	431 (282)	36.7 (8.8)	2.2 (1.3)		7.5 (6.2)	55%
			FING1.25QD	426 (293)	35.8 (8.4)	2.2 (1.3)		7.3 (6.0)	59%

ALE12/ALE24 alemtuzumab 12 or 24 mg/day for 5 days and, 12 months later, for 3 days, *BG240BID/BG240TD* dimethyl fumarate 240 mg twice or three times daily, *CLA3.5/CLA5.25* cladribine, cumulative dose 3.5 or 5.25 mg/kg, *DAC150Q4W/DAC300Q4W* daclizumab 150 or 300 mg every 4 weeks, *DMT* disease-modifying therapies, *EDSS* Expanded Disability Status Score, *FING0.5QD/FING1.25QD* fingolimod 0.5 or 1.25 mg daily, *GA20QD* glatiramer acetate 20 mg daily, *GA40TIW* glatiramer acetate 40 mg three times weekly, *IFNA22TIW/IFNA44TIW* interferon β-1a 22 or 44 μg three times weekly, *IFNA30QW/IFNA60QW* interferon β-1a 30 or 60 μg weekly, *IFNB50EOD/IFNB250EOD/IFN-B375EOD/IFNB500EOD* interferon β-1b 50, 250, 275 or 500 μg every other day, *SD* standard deviation, *NAT300Q4W* natalizumab 300 mg every 4 weeks, *NR* not reported, *OCRE600Q6M* ocrelizumab 600 mg every 6 months, *PIFN125Q2W/PIFN125Q4W* peginterferon 125 μg every 2 or 4 weeks, *PLA* placebo, *TERI7QD* teriflunomide 7 mg daily, *TERI14QD* teriflunomide 14 mg daily

[a]Median (minimum, maximum)

[b]Median (interquartile range)

[c]Considered glatiramer acetate branded

[d]Mean (95% confidence interval

[e]Considered only arms of therapies that included only patients as RRMS

[f]Studies not included in the meta-analyses

[g]Studies not included in the base case

4 Discussion

In our study, we compared efficacy and safety of DMTs through a systematic review of 40 studies (29,150 participants), with 33 studies (26,133 participants) included in the original analysis. Previous network meta-analyses of DMTs in RRMS, published by Tramacere et al. [8], Fogarty et al. [27], Siddiqui et al. [28] and Hamidi et al. [29] included, respectively, 39, 28, 44 and 49 studies. However, our network presents more strict inclusion criteria, considering only RRMS patients and not SPMS patients [8]. We also included the three new therapies (CLA3.5, DAC150Q4W and OCRE600Q6M) not considered in Tramacere et al. Fogarty et al. or Hamidi et al. [8, 27, 29].

The ARR outcome was the most reported by the studies, followed by DAE; although DPC was reported by most
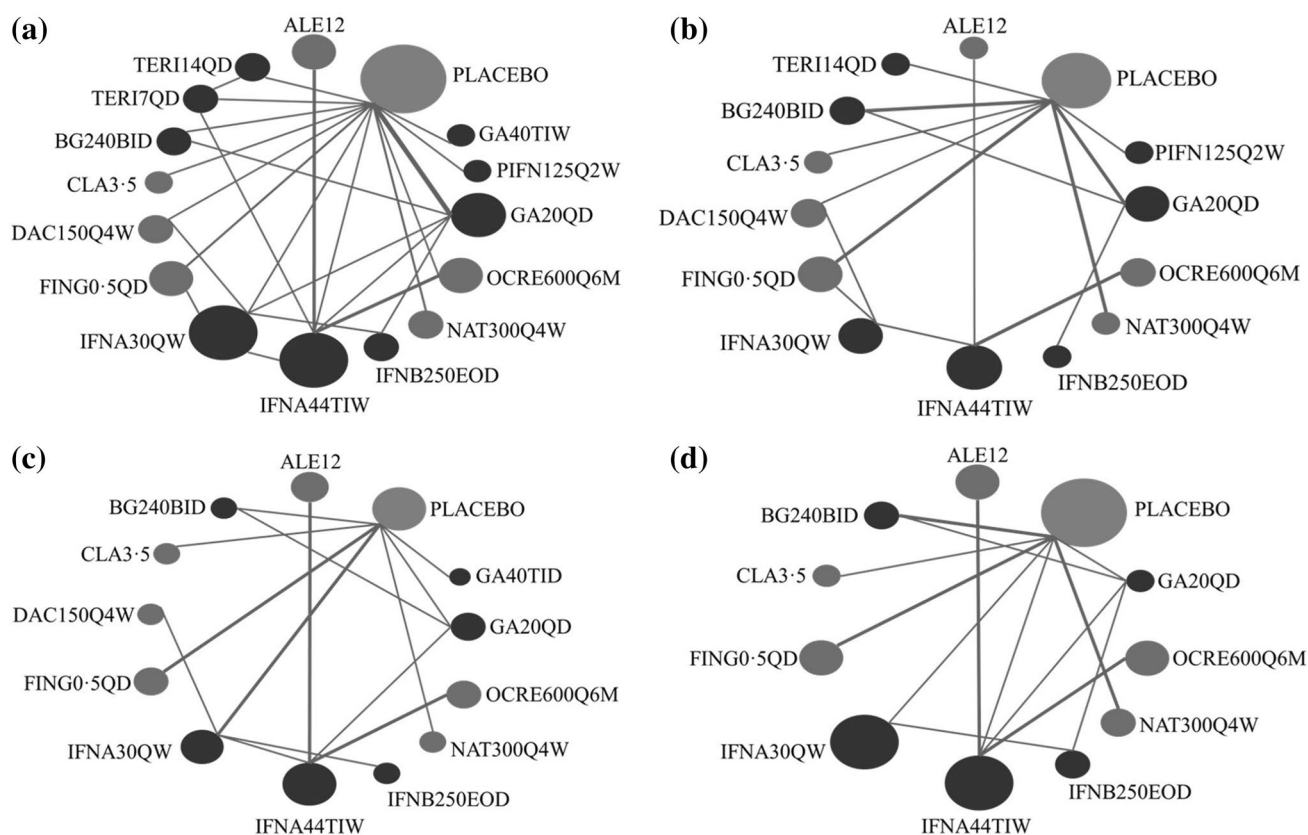
(a)

(b)

(c)

(d)

Fig. 1 Network meta-analyses. Network geometry: each node represents a therapy, and the lines represent direct comparisons in the literature; thicker lines represent largest number of studies identified, and larger nodes represent the largest number of studies for a therapy; dark-coloured nodes correspond to first-line therapies and light-coloured nodes to second- or third-line therapies and placebo. **a** Annualised relapse rate; **b** disability progression confirmed at 12 weeks; **c** disability progression confirmed at 24 weeks; **d** discontinuation due to adverse events at 96 weeks. *ALE12* alemtuzumab 12 mg daily for 5 days and, 12 months later, 12 mg daily for 3 days,

BG240BID dimethyl fumarate 240 mg twice daily, *CLA3.5* cladribine, cumulative dose 3.5 mg/kg, *DAC150Q4W* daclizumab 150 mg monthly, *FING0.5QD* fingolimod 0.5 mg daily, *GA20QD* glatiramer acetate 20 mg daily, *GA40TIW* glatiramer acetate 40 mg three times weekly, *IFNA30QW* interferon β-1a 30 μg weekly, *IFNA44TIW* interferon β-1a 44 μg three times weekly, *IFNB250EOD* interferon β-1b 250 μg every other day, *NAT300Q4W* natalizumab 300 mg monthly, *OCRE600Q6M* ocrelizumab 600 mg every 6 months, *PIFN125Q2W* peginterferon 125 μg every 2 weeks, *PLA* placebo, *TERI7QD* teriflunomide 7 mg daily, *TERI14QD* teriflunomide 14 mg daily

of the studies, the statistical synthesis of this outcome in the present review showed limitations due to differences in its assessment: some studies reported DPC or DIC only at 12 weeks, others at 24 weeks. QoL was poorly reported, and in a few cases, it was described incompletely (without distribution statistics) and with different assessment tools, including specific ones for MS, EQ-5D and SF-12/36, which precluded a network development. QoL and other patient-reported outcomes are highly relevant in the context of RRMS, considering that the disease is a progressively incapacitating condition. Hence, as important as avoiding a relapse and reducing disease intensity, the benefits obtained with the treatment should include physical and mental improvement from the perspective of the patient. The heterogeneity in measuring and reporting the addressed outcomes may be explained by the absence of a core outcome set for RRMS in adults, which would be paramount to guide future

studies, contributing, therefore, to the consistency and pertinence of new findings.

ARR was the outcome that presented the best robustness. The greater precision for this outcome is probably because (as recurrence of relapses) it is the primary endpoint in most of the studies, being used in the definition of sample size. ALE12, NAT300Q4W and OCRE600Q6M were the most effective treatments against placebo and active comparators, with over 80% probability of being the best choice (SUCRA). No significant differences in terms of efficacy were identified among them; however, differences in costs and administration schedule exist. NAT300Q4W is administered monthly, OCRE600Q6M is given every 6 months and alemtuzumab in two unique courses (one in 5 days in the first year and a second over 3 days in the second year). Despite the potential benefit of ALE12 administration, little evidence of long-term efficacy exists [30].

Table 2 Network meta-analysis results for ARR, regarding all follow-up durations (lower) and DAE for 96 weeks of follow-up (upper)

ALE12	0.47 (0.11; 2.50)	0.69 (0.23; 2.00)	0.44 (0.09; 2.60)	0.50 (0.15; 2.30)	-	0.52 (0.16; 2.10)	-	-	0.37 (0.17; 0.81)	0.49 (0.15; 1.50)	-	0.48 (0.10; 2.10)	-	0.52 (0.14; 2.40)	0.68 (0.22; 2.60)
0.99 (0.78; 1.30)	**NAT300Q4W**	1.50 (0.28; 5.70)	0.93 (0.22; 3.90)	1.10 (0.36; 3.40)	-	1.10 (0.35; 3.20)	-	-	0.80 (0.18; 2.60)	1.00 (0.26; 2.80)	-	1.00 (0.20; 3.70)	-	1.10 (0.34; 3.50)	1.40 (0.56; 3.60)
0.83 (0.68; 1.00)	0.84 (0.65; 1.10)	**OCRE600Q6M**	0.64 (0.14; 3.70)	0.73 (0.23; 3.40)	-	0.76 (0.24; 3.00)	-	-	0.54 (0.27; 1.10)	0.70 (0.23; 2.10)	-	0.70 (0.15; 2.90)	-	0.76 (0.22; 3.40)	0.98 (0.35; 3.70)
0.73 (0.55; 0.98)	0.74 (0.57; 0.96)	0.89 (0.66; 1.20)	**CLA3.5**	1.10 (0.33; 4.30)	-	1.20 (0.32; 4.10)	-	-	0.85 (0.17; 3.20)	1.10 (0.26; 3.60)	-	1.10 (0.19; 4.60)	-	1.20 (0.31; 4.60)	1.50 (0.51; 4.80)
0.66 (0.53; 0.83)	0.67 (0.55; 0.81)	0.80 (0.63; 1.00)	0.90 (0.70; 1.10)	**FING0.5QD**	-	1.10 (0.40; 2.30)	-	-	0.75 (0.20; 2.00)	0.98 (0.30; 2.00)	-	0.95 (0.21; 2.90)	-	1.00 (0.37; 2.70)	1.30 (0.70; 2.50)
0.65 (0.53; 0.81)	0.66 (0.54; 0.80)	0.79 (0.63; 0.98)	0.89 (0.69; 1.10)	0.99 (0.83; 1.20)	**DAC150Q4W**	-	-	-	-	-	-	-	-	-	-
0.65 (0.51; 0.81)	0.65 (0.53; 0.80)	0.78 (0.61; 0.99)	0.88 (0.68; 1.10)	0.98 (0.81; 1.20)	0.99 (0.82; 1.20)	**BG240BID**	-	-	0.72 (0.22; 1.80)	0.93 (0.37; 1.70)	-	0.91 (0.24; 2.70)	-	0.99 (0.39; 2.60)	1.30 (0.74; 2.60)
0.48 (0.36; 0.65)	0.49 (0.37; 0.63)	0.58 (0.43; 0.79)	0.66 (0.48; 0.89)	0.73 (0.57; 0.94)	0.74 (0.57; 0.95)	0.75 (0.58; 0.97)	**PIFN125Q2W**	-	-	-	-	-	-	-	-
0.47 (0.37; 0.61)	0.48 (0.38; 0.60)	0.57 (0.44; 0.75)	0.65 (0.49; 0.84)	0.72 (0.58; 0.88)	0.73 (0.58; 0.90)	0.73 (0.59; 0.91)	0.98 (0.74; 1.30)	**GA40TIW**	-	-	-	-	-	-	-
0.45 (0.40; 0.52)	0.46 (0.37; 0.56)	0.55 (0.47; 0.63)	0.62 (0.48; 0.79)	0.68 (0.57; 0.82)	0.69 (0.59; 0.81)	0.70 (0.58; 0.85)	0.94 (0.72; 1.20)	1.00 (0.88; 1.10)	**IFNA44TIW**	1.30 (0.56; 3.00)	-	1.30 (0.35; 4.50)	-	1.40 (0.49; 5.30)	1.80 (0.81; 5.40)
0.45 (0.38; 0.55)	0.46 (0.38; 0.54)	0.55 (0.45; 0.66)	0.62 (0.49; 0.78)	0.68 (0.59; 0.80)	0.69 (0.60; 0.80)	0.70 (0.60; 0.82)	0.94 (0.73; 1.20)	0.95 (0.79; 1.20)	1.00 (0.88; 1.10)	**GA20QD**	-	0.99 (0.35; 2.60)	-	1.10 (0.47; 3.30)	1.40 (0.80; 3.30)
0.45 (0.35; 0.58)	0.46 (0.37; 0.57)	0.55 (0.42; 0.70)	0.62 (0.47; 0.80)	0.68 (0.56; 0.84)	0.69 (0.56; 0.85)	0.70 (0.56; 0.87)	0.94 (0.71; 1.20)	0.95 (0.75; 1.20)	1.00 (0.81; 1.20)	1.00 (0.83; 1.20)	**TERI14QD**	-	-	-	-
0.44 (0.36; 0.55)	0.44 (0.36; 0.54)	0.53 (0.42; 0.66)	0.60 (0.46; 0.77)	0.67 (0.55; 0.81)	0.67 (0.57; 0.80)	0.68 (0.56; 0.83)	0.91 (0.70; 1.20)	0.97 (0.86; 1.10)	0.97 (0.82; 1.20)	0.93 (0.74; 1.20)	0.97 (0.78; 1.20)	**IFNB250EOD**	-	1.10 (0.38; 4.00)	1.40 (0.53; 5.10)
0.43 (0.34; 0.54)	0.43 (0.35; 0.53)	0.51 (0.40; 0.66)	0.58 (0.44; 0.75)	0.64 (0.53; 0.79)	0.65 (0.53; 0.79)	0.66 (0.53; 0.81)	0.88 (0.67; 1.20)	0.90 (0.71; 1.10)	0.94 (0.77; 1.10)	0.94 (0.78; 1.10)	0.94 (0.79; 1.10)	0.97 (0.78; 1.20)	**TERI7QD**	-	-
0.37 (0.30; 0.44)	0.37 (0.31; 0.44)	0.44 (0.36; 0.54)	0.50 (0.39; 0.63)	0.56 (0.48; 0.64)	0.56 (0.51; 0.62)	0.57 (0.48; 0.67)	0.76 (0.60; 0.97)	0.78 (0.64; 0.94)	0.81 (0.71; 0.93)	0.81 (0.73; 0.91)	0.81 (0.67; 0.98)	0.83 (0.72; 0.97)	0.86 (0.72; 1.0)	**IFNA30QW**	1.30 (0.62; 2.80)
0.31 (0.26; 0.38)	0.31 (0.27; 0.36)	0.37 (0.31; 0.46)	0.42 (0.34; 0.52)	0.47 (0.41; 0.53)	0.47 (0.42; 0.54)	0.48 (0.42; 0.55)	0.64 (0.51; 0.80)	0.66 (0.55; 0.78)	0.69 (0.60; 0.79)	0.69 (0.62; 0.76)	0.69 (0.58; 0.81)	0.70 (0.61; 0.82)	0.73 (0.63; 0.85)	0.84 (0.77; 0.93)	**PLA**

Interpretation: left therapy compared with right therapy; for example, ALE12 is more effective than PLA with an HR of 0.31 (95% CrI 0.26–0.38), although it is safer than IFNA44TIW with a relative risk (RR) of 0.37 (95% CrI 0.17–0.81). *Cells with light-grey fill* represent measures that presented difference with statistical significance

ALE12 alemtuzumab 12 mg/day for 5 days (1st course) and for 3 days (2nd course) with 1-year interval between each course, *ARR* annualised relapse rate, *BG240BID* dimethyl fumarate 240 mg twice a day, *CLA3.5* cladribine, cumulative dose 3.5 mg/kg, *CrI* credibility interval, *DAC150Q4W* daclizumab 150 mg every 4 weeks, *DAE* discontinuation due to adverse events, *FING0.5QD* fingolimod 0.5 mg/day, *GA20QD* glatiramer acetate 20 mg/day, *GA40TIW* glatiramer acetate 40 mg three times weekly, *IFNA30QW* interferon β-1a 30 μg every week, *IFNA44TIW* interferon β-1a 44 μg three times weekly, *IFNB250EOD* interferon β-1b 250 μg every other day, *NAT300Q4W* natalizumab 300 mg every 4 weeks, *OCRE600Q6M* ocrelizumab 600 mg every 6 months, *PIFN125Q2W* peginterferon 125 μg every 2 weeks, *PLA* placebo, *TERI7QD* teriflunomide 7 mg/day, *TERI14QD* teriflunomide 14 mg/day

Cost–utility studies comparing DMTs identified ALE12 as more effective and less costly than treatment alternatives, followed by NAT300Q4W [29]. No pharmacoeconomic studies were found that evaluated ocrelizumab with ALE12 or NAT300Q4W, but comparison with IFNA44TIW shows OCRE600Q6M as the most efficient therapy [31]. CLA3.5 is administered orally, which may be an additional benefit compared with daclizumab (SC) and dimethyl fumarate (PO), with similar efficacy. Conversely to Siddiqui [28], we also identified that CLA3.5, DAC150Q4W, BG240BID and FING0.5QD present an intermediate efficacy profile in ARR, below ALE12 and NAT300Q4W. Potential reason for this discrepancy between the meta-analyses could be the different patients' inclusion criteria.

Very similar results were obtained for DPC12 and 24. It is important to note that the INCOMIN study [26] was reported as an outlier for this outcome and may have overestimated the efficacy of IFNB250EOD [32, 33]. The influence of this trial in the ARR network was not relevant because of the greater number of studies included. Sensitivity analyses had a greater impact on secondary and tertiary outcomes, but not on ARR. This suggests that outcomes with lower statistical power are more sensitive to the inclusion or exclusion of studies.

We evaluated safety, considering only DAE96, because of two reasons. It was not possible to build a 48-week network, but the main reason was that ALE12 is administered once a year, which could overestimate its safety in periods
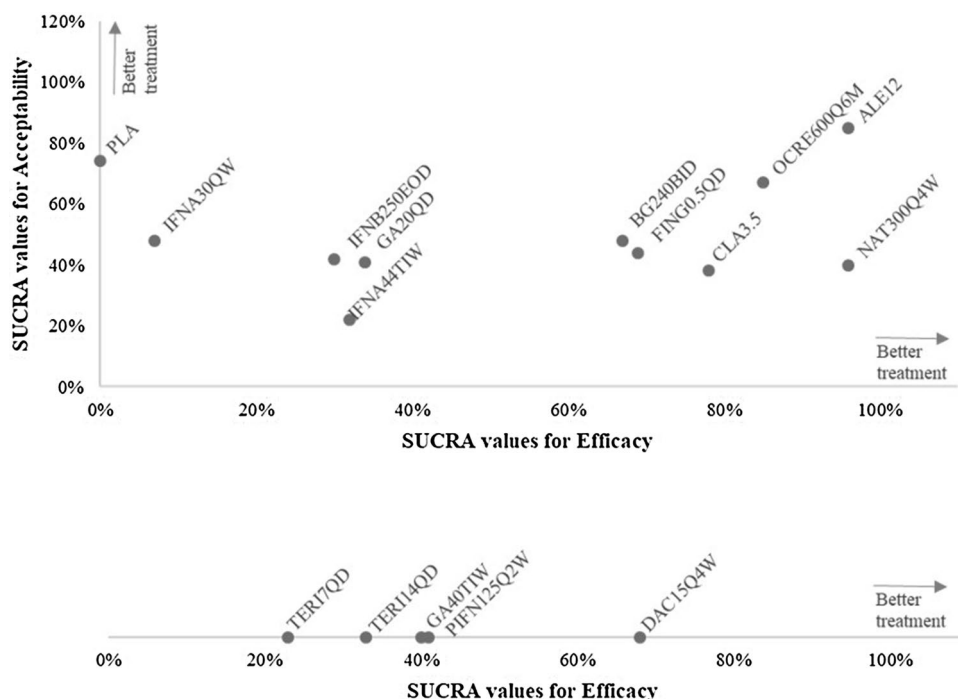
Fig. 2 Ranking plot based on analysis of surface under the cumulative ranking curve (*SUCRA*) values for efficacy (annualised relapse rate) and acceptability (treatment discontinuation due to adverse events) over 96 weeks. Treatments lying in the upper-right corner are more effective and acceptable than the other treatments. TERI7QD, TERI14QD, GA40TIW, PIFN125Q2W and DAC150Q4W could not be included in network meta-analysis because of discontinuation due to adverse events. *ALE12* alemtuzumab 12 mg daily for 5 days and, 12 months later, 12 mg daily for 3 days, *BG240BID* dimethyl fumarate 240 mg twice daily, *CLA3.5* cladribine, cumulative dose of < 2 years. For this outcome, whereas in individual studies 3.5 mg/kg, *DAC150Q4W* daclizumab 150 mg monthly, *FING0.5QD* fingolimod 0.5 mg daily, *GA20QD* glatiramer acetate 20 mg daily, *GA40TIW* glatiramer acetate 40 mg three times weekly, *IFNA30QW* interferon β-1a 30 μg weekly, *IFNA44TIW* interferon β-1a 44 μg three times weekly, *IFNB250EOD* interferon β-1b 250 μg every other day, *NAT300Q4W* natalizumab 300 mg monthly, *OCRE600Q6M* ocrelizumab 600 mg every 6 months, *PIFN125Q2W* peginterferon 125 μg every 2 weeks, *PLA* placebo, *TERI7QD* teriflunomide 7 mg daily, *TERI14QD* teriflunomide 14 mg daily

of < 2 years. For this outcome, whereas in individual studies and direct meta-analyses some therapies demonstrate superiority over IFNA44TIW and placebo, in our network we identified no differences among most of the comparisons. The only statistical difference found was between ALE12 and IFNA44TIW, favouring ALE12, probably because of the difference administration schedules make (once a year vs three times weekly, respectively). Despite the wide credibility intervals, SUCRA analysis suggests that ALE12 and OCRE600Q6M are the safest therapies. It is important to note that DAE is usually associated with serious adverse events, and the ranking we obtained may not correspond with the frequency of other adverse events [33]. This analysis demonstrates how considering DAE as the only relevant safety outcome could cause health professionals to make wrong decisions, especially regarding new therapies. Unlike the DAE analysis, clinical practice demonstrates that most of these therapies are associated with high rates of adverse events, and many long-term adverse events are identified after the end of treatment (e.g. acute acalculous cholecystitis, thyroid disorders, immune thrombocytopenia with

alemtuzumab) [34, 35] or even after the end of the studies (e.g. cancer, severe infections and deaths with ocrelizumab or progressive multifocal leukoencephalopathy, severe liver failures and lymphoma cases with natalizumab) [36, 37]. Moreover, the case of daclizumab illustrates this concern: despite the absence of DAE at 96 weeks or differences with interferon on report of severe adverse events in the DECIDE (Efficacy and Safety of Daclizumab High Yield Process Versus Interferon β 1a in Patients With Relapsing–Remitting Multiple Sclerosis) and SELECT (Daclizumab High-Yield Process in Relapsing–Remitting Multiple Sclerosis) trials, recently the manufacturer announced voluntary worldwide withdrawal of marketing authorisations for Zinbryta® for RRMS due to the identification of cases of inflammatory encephalitis and meningoencephalitis [38]. Therefore, besides the need to consider frequency of adverse events, it is mandatory to consider results of trial extension and real-world evidence data with adequate confidence in evidence, in order to conclude about the safety profile of DMTs.

We selected INFA44TIW as the comparator for our analyses of confidence in the evidence, instead of placebo

Table 3 Evidence quality assessment for annualised relapse rate and discontinuation due to adverse events (GRADE)

Patient: patients with relapsing–remitting multiple sclerosis (RRMS)
Setting: secondary healthcare centres
Intervention: immunomodulators or immunosuppressants used for RRMS
Comparison: interferon β-1a, 44 μg three times weekly (IFNA44TIW)

Intervention	Illustrative comparative risks		Relative effect (95% CrI)	Confidence in the evidence (GRADE)
	Assumed risk with IFNA44TIW	Corresponding risk with intervention (95% CrI)		
Annualised relapse rate				
ALE12	35 per 100	16 per 100 (14–18)	0.45 (0.40–0.52)	⊕⊕⊕⊕ High
NAT300Q4W	35 per 100	16 per 100 (13–21)	0.46 (0.38–0.60)	⊕⊕⊕⊕ High
OCRE600Q6M	35 per 100	19 per 100 (17–22)	0.55 (0.47–0.63)	⊕⊕⊕⊕ High
CLA3.5	35 per 100	22 per 100 (17–28)	0.62 (0.49–0.79)	⊕⊕⊕⊕ High
FING0.5QD	35 per 100	24 per 100 (20–29)	0.68 (0.57–0.82)	⊕⊕⊕⊕ High
DAC150Q4W	35 per 100	24 per 100 (21–28)	0.69 (0.59–0.81)	⊕⊕⊕⊕ High
BG240BID	35 per 100	24 per 100 (20–30)	0.70 (0.58–0.85)	⊕⊕⊕⊕ High
PIFN125Q2W	35 per 100	33 per 100 (25–42)	0.94 (0.72–1.20)	⊕⊕⊕○[1] Moderate
GA40TIW	35 per 100	35 per 100 (31–39)	1.00 (0.88–1.10)	⊕⊕⊕○[1] Moderate
GA20QD	35 per 100	35 per 100 (32–40)	1.00 (0.91–1.14)	⊕⊕○○[1,2] Low
TERI14QD	35 per 100	36 per 100 (29–43)	1.03 (0.83–1.23)	⊕⊕⊕○[1] Moderate
IFNB250EOD	35 per 100	36 per 100 (29–43)	1.03 (0.83–1.22)	⊕⊕○○[1,2] Low
TERI7QD	35 per 100	37 per 100 (32–46)	1.06 (0.91–1.30)	⊕⊕⊕○[1] Moderate
IFNA30QW	35 per 100	43 per 100 (38–49)	1.23 (1.08–1.41)	⊕⊕○○[1,2,3] Low
Discontinuation due to adverse events				
ALE12	11 per 100	4 per 100 (2–9)	0.37 (0.17–0.81)	⊕⊕⊕⊕ High
OCRE600Q6M	11 per 100	6 per 100 (3–12)	0.54 (0.27–1.10)	⊕⊕⊕⊕ High
IFNA30QW	11 per 100	8 per 100 (2–22)	0.71 (0.19–2.04)	⊕⊕⊕○[2] Moderate
BG240BID	11 per 100	8 per 100 (2–20)	0.72 (0.22–1.80)	⊕⊕⊕⊕ High
FING0.5QD	11 per 100	8 per 100 (2–22)	0.75 (0.20–2.00)	⊕⊕⊕⊕ High
GA20QD	11 per 100	9 per 100 (4–20)	0.77 (0.33–1.79)	⊕⊕⊕○[2] Moderate
IFNB250EOD	11 per 100	9 per 100 (2–32)	0.77 (0.22–2.86)	⊕⊕⊕○[2] Moderate

Table 3 (continued)

Patient: patients with relapsing–remitting multiple sclerosis (RRMS)
Setting: secondary healthcare centres
Intervention: immunomodulators or immunosuppressants used for RRMS
Comparison: interferon β-1a, 44 μg three times weekly (IFNA44TIW)

Intervention	Illustrative comparative risks		Relative effect (95% CrI)	Confidence in the evidence (GRADE)
	Assumed risk with IFNA44TIW	Corresponding risk with intervention (95% CrI)		
NAT300Q4W	11 per 100	9 per 100 (2–29)	0.80 (0.18–2.60)	⊕⊕⊕⊕ High
CLA3.5	11 per 100	9 per 100 (2–35)	0.85 (0.17–3.20)	⊕⊕⊕⊕ High

1: imprecision; 2: methodological bias—measurement of the outcome; 3: methodological bias—selection of the reported result

ALE12 alemtuzumab 12 mg daily for 5 days and, 12 months later, 12 mg daily for 3 days, *BG240BID* dimethyl fumarate 240 mg twice a day, *CLA3.5* cladribine, cumulative dose 3.5 mg/kg, *CrI* credibility interval, *DAC150Q4W* daclizumab 150 mg every 4 weeks, *FING0.5QD* fingolimod 0.5 mg/day, *GA20QD* glatiramer acetate 20 mg/day, *GA40TIW* glatiramer acetate 40 mg three times weekly, *GRADE* Grading of Recommendations Assessment, Development and Evaluation, *IFNA30QW* interferon β-1a 30 μg every week, *IFNA44TIW* interferon β-1a 44 μg three times weekly, *IFNB250EOD* interferon β-1b 250 μg every other day, *NAT300Q4W* natalizumab 300 mg every 4 weeks, *OCRE600Q6M* ocrelizumab 600 mg every 6 months, *PIFN125Q2W* peginterferon 125 μg every 2 weeks, *TERI7QD* teriflunomide 7 mg/day, *TERI14QD* teriflunomide 14 mg/day

like Tramacere et al. [8]. Due to the effects of imprecision, selecting a high-efficacy comparator, such as NAT300Q4W, would downgrade the quality of evidence for each comparison, whereas selecting a low-efficacy comparator, such as IFNA30QW or placebo, would excessively upgrade the quality of evidence. High-efficacy comparators present the highest quality of evidence for efficacy, whereas intermediate- and low-efficacy comparators present moderate and low quality. Evidence obtained for safety was mainly high quality. None of the comparisons were downgraded due to heterogeneity, publication bias or intransitivity; however, the low heterogeneity may have been caused by the small number of studies in each pairwise meta-analysis. Our sensitivity analyses demonstrated no problems associated with transitivity.

As a result of our analyses, especially of efficacy evaluation, we suggest a three-category classification for RRMS instead of the two categories recommend by ABN [7]. We identified three efficacy clusters: high efficacy (i.e. ALE12, NAT300Q4W and OCRE600Q6M), intermediate efficacy (i.e. CLA3.5, FING0.5QD and BG240BID) and low efficacy (i.e. PIFN125Q2W, GA40TIW, IFNA44TIW, GA20QD, TERI14QD and IFNB250EOD). This reclassification would have an impact on selecting first-line therapies for patients with more aggressive conditions (highly active or rapidly evolving severe); although there is not sufficient evidence, it seems that high- and intermediate-efficacy therapies could be the first-line choice for patients with more aggressive conditions, which makes a difference, especially for FING0.5QD and BG240BID, currently in category 1 in the ABN guidelines [7], but also for CLA3.5, not included in the guideline. Although ABN guidelines draw no difference

between interferons and teriflunomide, our network identified TERI7QD and IFNA30QW as the worst options for ARR, which probably justifies removing them from the guideline. Considering its efficacy profile, DAC150Q4W could be proposed as intermediate efficacy; however, recently the manufacturer announced voluntary worldwide withdrawal of marketing authorisations for Zinbryta® for RRMS due to identification of cases of inflammatory encephalitis and meningoencephalitis [38].

As in any systematic search, there is a chance that studies were missed. However, the grey literature and manual searches found no additional studies, reinforcing the quality of our search. Although we applied strict selection criteria for the type of MS, we found poor reporting of raw data in subgroup analyses of primary studies, which precludes us from performing subgroup meta-analyses (e.g. age, EDSS, disease duration, disease activity). The rapid evolution of the diagnostic criteria may produce differences in efficacy assessment, and we could not make sensitivity analyses, because most DMTs were evaluated by a single criteria. Moreover, analyses for DAE may not represent accurate safety concerns as seen in real-world settings.

5 Conclusion

High-quality evidence shows that alemtuzumab, natalizumab and ocrelizumab present the highest efficacy among DMTs, and other meta-analyses are required to evaluate the frequency of adverse events to better understand the safety profile of these therapies. Based on efficacy profile,

guidelines should considerer a three-category classification (i.e. high, intermediate and low efficacy). Specific studies should be conducted for a more precise selection of therapies for more aggressive RRMS conditions.

Compliance with Ethical Standards

References

1. Multiple Sclerosis International Federation. Atlas of MS 2013. Mapping multiple sclerosis around the world. 2014. http://www.msif.org/wp-content/uploads/2014/09/Atlas-of-MS.pdf. Accessed 11 July 2018.

2. Montalban X, Gold R, Thompson AJ, Otero-Romero S, Amato MP, Chandraratna D, et al. ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. Mult Scler J. 2018;24:96–120. https://doi.org/10.1177/1352458517751049.

3. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology. 2014;83:278–86.

4. Pickin M, Cooper CL, Chater T, O'Hagan A, Abrams KR, Cooper NJ, et al. The multiple sclerosis risk sharing scheme monitoring study—early results and lessons for the future. BMC Neurol BioMed Cent. 2009;9:1.

5. Kasper DL, Fauci AS, Hauser S, Longo D, Jameson JL, Loscalzo J. Harrisons manual of medicine. 19th ed. New York City: McGraw-Hill Education; 2016.

6. Leary SM, Porter B, Thompson AJ. Multiple sclerosis: diagnosis and the management of acute relapses. The Fellowship of Postgraduate Medicine. Postgrad Med J. 2005;81:302–8.

7. Scolding N, Barnes D, Cader S, Chataway J, Chaudhuri A, Coles A, et al. Association of British Neurologists: revised (2015) guidelines for prescribing disease-modifying treatments in multiple sclerosis. Pract Neurol. 2015;15:273–9. https://doi.org/10.1136/practneurol-2015-001139.

8. Tramacere I, Del Giovane C, Salanti G, D'Amico R, Filippini G. Immunomodulators and immunosuppressants for relapsing-remitting multiple sclerosis: a network meta-analysis. Cochrane database Syst Rev. https://doi.org/10.1002/14651858.CD011381.pub2.

9. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR task force on indirect treatment comparisons good research practices: part 2. Value Heal. 2011;14:429–37. https://doi.org/10.1016/j.jval.2011.01.011.

10. Tonin FS, Rotta I, Mendes AM, Pontarolo R. Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. Pharm Pract (Granada). 2017;15:943.

11. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med. 2015;162:777–84.

12. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011.

13. Lucchetta RC, Tonin FS, Borba HHL, Leonart LP, Ferreira VL, Bonetti AF, et al. Systematic review and network meta-analysis of disease-modifying therapies in relapsing-remitting multiple sclerosis—CRD42017059120. PROSPERO. 2017. http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD420170559120. Accessed 11 July 2018.

14. Higgins J, Sterne J, Savović J, Page M, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials. In: Chandler J, McKenzie J, Boutron I, Welch V (eds) Cochrane Methods Cochrane Database Syst Rev. 2016;Suppl.1.

15. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/. Accessed 11 July 2018.

16. Wickham H, Hester J, François R. Readr: Read rectangular tet data. R package version 1.1.1. 2017. https://cran.r-project.org/package=readr. Accessed 11 July 2018.

17. Schwarzer G. Meta: An R package for meta-analysis. R News. 2007;7:40–5.

18. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36:1–48.

19. Valkenhoef G van, Kuiper J. GeMTC: Network Meta-Analysis Using Bayesian Methods. R Packag version 08-2. 2016. https://cran.r-project.org/package=gemtc. Accessed 11 July 2018.

20. Plummer M. RJags: Bayesian Graphical Models using MCMC. R Packag version 4-6. 2016. https://cran.r-project.org/package=rjags. Accessed 11 July 2018.

21. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R News. 2006;6:7–11.

22. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004;23:3105–24. https://doi.org/10.1002/sim.1875.

23. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol. 2012;41:818–27. https://doi.org/10.1093/ije/dys041.

24. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making. 2013;33:641–56. https://doi.org/10.1177/0272989X12455847.

25. GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490. https://doi.org/10.1136/bmj.328.7454.1490.

26. Durelli L, Verdun E, Barbero P, Bergui M, Versino E, Ghezzi A, et al. Every-other-day interferon beta-1b versus once-weekly interferon beta-1a for multiple sclerosis: results of a 2-year prospective randomised multicentre study (INCOMIN). Lancet (London, England). 2002;359:1453–60.

27. Fogarty E, Schmitz S, Tubridy N, Walsh C. Comparative efficacy of disease-modifying therapies for patients with relapsing remitting multiple sclerosis: systematic review and network meta-analysis. Mult Scler Relat Disord 2016;9:23–30. https://doi.org/10.1016/j.msard.2016.06.001.

28. Siddiqui MK, Khurana IS, Budhia S, Hettle R, Harty G, Wong SL. Systematic literature review and network meta-analysis of cladribine tablets versus alternative disease-modifying treatments for relapsing–remitting multiple sclerosis. Curr Med Res Opin. 2017;34:1361–71. https://doi.org/10.1080/03007995.2017.1407303.

29. Hamidi V, Couto E, Ringerike T, Klemp M. A multiple treatment comparison of eleven disease-modifying drugs used for multiple sclerosis. J Clin Med Res. 2018;10:88–105.

30. Ziemssen T, Thomas K. Alemtuzumab in the long-term treatment of relapsing-remitting multiple sclerosis: an update on the clinical trial evidence and data from the real world. Ther Adv Neurol Disord 2017;10:343–59. https://doi.org/10.1177/1756285617722706.

31. Yang H, Duchesneau E, Foster R, Guerin A, Ma E, Thomas NP. Cost-effectiveness analysis of ocrelizumab versus subcutaneous interferon beta-1a for the treatment of relapsing multiple sclerosis. J Med Econ. 2017;20:1056–65. https://doi.org/10.1080/13696998.2017.1355310.

32. Vartanian T. An examination of the results of the EVIDENCE, INCOMIN, and phase III studies of interferon beta products in the treatment of multiple sclerosis. Clin Ther. 2003; 25:105–18.

33. Tolley K, Hutchinson M, You X, Wang P, Sperling B, Taneja A, et al. A network meta-analysis of efficacy and evaluation of safety of subcutaneous pegylated interferon beta-1a versus other injectable therapies for the treatment of relapsing-remitting multiple sclerosis. PLoS One. 2015;10:e0127960.

34. Croteau D, Flowers C, Kulick CG, Brinker A, Kortepeter CM, Croteau C, Kulick CG, Brinker A, Kortepeter CMDF, et al. Acute acalculous cholecystitis: a new safety risk for patients with MS treated with alemtuzumab. Neurology. 2018;90:e1548–52. https://doi.org/10.1212/WNL.0000000000005422.

35. Coles JA, Fox EJ, Giovannoni G, Hartung HP, Havrdova E, Schippling S, Selmaj KW, Traboulsee A, Compston DAS, Margolin DH, Thangavelu K, Chirieac MC, Jody D, Xenopoulos P, Hogan RJ, Panzara MA, Arnold DAJC. Alemtuzumab CARE-MS II 5-year follow-up: efficacy and safety findings. Neurology. 2017;89:1117–26.

36. Clerico CA, Di Liberto A, Rolla S, Bardina V, Barbero P, De Mercanti SF, Durelli LMA. Long-term safety evaluation of natalizumab for the treatment of multiple sclerosis. Expert Opin Drug Saf. 2017;16:963–72.

37. McGinley BP, Cohen JAMPM. Safety of monoclonal antibodies for the treatment of multiple sclerosis. Expert Opin Drug Saf. 2017;16:89–100.

38. Biogen and AbbVie Announce the Voluntary Worldwide Withdrawal of Marketing Authorizations for ZINBRYTA® (daclizumab) for Relapsing Multiple Sclerosis. 2018. https://www.businesswire.com/news/home/20180302005168/en/Biogenand-AbbVie-Announcethe-VoluntaryWorldwide-Withdrawal-Marketing-Authorizations. Accessed 11 July 2018.