



REVIEW

Overview of Differences and Similarities of Published Mixed Treatment Comparisons on Pharmaceutical Interventions for Multiple Sclerosis

Maria Pia Sormani · Robert Wolff · Shona Lang · Steven Duffy · Robert Hyde · Elizabeth Kinter · Craig Wakeford · Gavin Giovannoni · Jos Kleijnen

Received: June 5, 2020 / Accepted: September 4, 2020 / Published online: September 25, 2020
© The Author(s) 2020

ABSTRACT

Introduction: Mixed treatment comparisons (MTCs) are increasingly important in the assessment of the benefit–risk profile of pharmaceutical treatments for relapsing–remitting multiple sclerosis (RRMS). Interpretation of MTCs requires a clear understanding of the methods of analysis and population studied. The objectives of this work were to compare

Digital Features To view digital features for this article go to <https://doi.org/10.6084/m9.figshare.12912209>.

Electronic Supplementary Material The online version of this article (<https://doi.org/10.1007/s40120-020-00213-4>) contains supplementary material, which is available to authorized users.

M. Pia Sormani
Department of Health Sciences, University of Genoa, Genoa, Italy

R. Wolff (✉) · S. Lang · S. Duffy
Kleijnen Systematic Reviews Ltd, York, UK
e-mail: robert@systematic-reviews.com

R. Hyde · E. Kinter · C. Wakeford
Biogen, Cambridge, USA

G. Giovannoni
Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK

J. Kleijnen
School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands

MTCs of pharmaceutical treatments for RRMS, including a detailed description of differences in populations, treatments assessed, methods used and findings; and to discuss key considerations when conducting an MTC.

Methods: Fourteen databases were searched until July 2019 to identify MTCs (published during or after 2010) in adults (at least 18 years of age) with RRMS or rapidly evolving severe RRMS treated with any form of pharmaceutical treatment. No language restriction was imposed.

Results: Twenty-seven MTCs assessing 21 treatments were identified. Comparison highlighted many differences in conduct and reporting between MTCs relating to the patient populations or treatments included, duration of follow-up and outcomes of interest measured. The lack of similarity between the MTCs leads to questions about variability in the robustness of analyses and makes comparisons between studies challenging.

Conclusion: Given the importance of MTCs for healthcare decision-making, it is imperative that reporting of methods, results and assumptions is clear and transparent to allow accurate interpretation of findings. For MTCs to be relevant, the choice of outcome measures should reflect clinical practice. Combination of treatments or of outcomes measured at different points of time should be avoided, as should imputation without justification. Furthermore, all approved treatment options should be

included and updates of MTCs should be conducted when data for new treatments are published.

Keywords: Disease-modifying therapies; Healthcare decision-making; Mixed treatment comparison; Multiple sclerosis; Relapsing–remitting multiple sclerosis

Key Summary Points

Interpretation of mixed treatment comparisons (MTCs) requires a clear understanding of the methods of analysis and population studied.

MTCs of pharmaceutical treatments for relapsing–remitting multiple sclerosis were detailed and compared.

A comparison of all identified MTCs showed that the findings of each individual MTC were not directly comparable with the others because of differences in the disease-modifying therapies (DMTs) compared, included studies, effect measures and analysis methods.

Given the importance of MTCs for healthcare decision-making, it is imperative that reporting of methods, results and assumptions is clear and transparent to allow accurate interpretation of findings.

For MTCs to be relevant, the choice of outcome measures should reflect clinical practice, combination of treatments or outcomes measured at different points of time should be avoided, as should imputation without justification, all approved treatment options should be included, and updates of MTCs should be conducted when data for new treatments are published.

DIGITAL FEATURES

This article is published with digital features to facilitate understanding of the article. To view digital features for this article go to <https://doi.org/10.6084/m9.figshare.12912209>.

INTRODUCTION

In the last 8 years, the number of approved disease-modifying therapies (DMTs) for multiple sclerosis (MS), a chronic autoimmune-mediated inflammatory disease of the central nervous system, has more than doubled. In contrast with the requirements of regulatory bodies for randomised controlled trials (RCTs) directly comparing new treatments with placebo and/or one active comparator [1], a broader evidence base is needed to inform how new medicines should fit into existing treatment algorithms [2]. Mixed treatment comparisons (MTCs) utilise a totality of data by combining both direct and indirect evidence [3]. Consequently, they have become increasingly important to the scientific and clinical community as well as reimbursement agencies by enabling objective comparison of the efficacy and safety of MS therapies that have never been compared head-to-head [4–6].

As discussed by Giovannoni et al. [7], the field of MTC continues to evolve and mature. Interpreting MTC findings is complex, particularly in MS where diagnostic criteria [8–11], definition of clinical outcomes and patient populations studied have changed over time [12]. These temporal changes in clinical study design make meeting the necessary assumptions of RCT homogeneity, similarity and transitivity in MTC difficult to achieve in MS [13, 14], resulting in variable conclusions depending on how systematic review and MTC are performed. The combination of RCTs conducted in populations with different subtypes of MS, different sets of treatments, or doses licensed under different regulatory authorities, while pragmatic, can provide limited or invalid information when applied to a different clinical context.

However, progress has been made in enhancing understanding of the strengths and limitations of MTC findings through improved transparency of study conduct and reporting. The application of, for example, standards defined by the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) [15] and the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) guidelines [16] aim to improve completeness of meta-analyses and MTC reporting, providing important insights to decision-makers on the applicability of MTC findings to their patient populations.

We undertook a comparison of existing MTCs of DMTs in MS to assess their relevance to European decision-makers. Here, we present a detailed discussion of the differences identified between MTCs in the populations studied, treatments assessed, methods employed and implications for interpretation of findings.

METHODS

As part of a systematic review to inform an up-to-date MTC of DMTs at European Commission-approved doses for the treatment of relapsing–remitting MS (RRMS), already completed MTCs were also identified and extracted for comparison. This article is based on previously conducted studies and does not contain any studies with human participants or animals performed by any of the authors.

Search Methods

Searches were conducted to identify MTCs of pharmaceutical treatments for MS. The following databases and resources were searched from inception to July 2019: Cochrane Database of Systematic Reviews (CDSR); Database of Abstracts of Reviews of Effects (DARE); Health Technology Assessment (HTA) Database; National Institute for Health Research (NIHR) Journals Library; International Prospective Register of Systematic Reviews (PROSPERO); National Institute for Health and Care Excellence (NICE) Guidance; Kleijnen Systematic Reviews (KSR) Evidence; Canadian Agency for

Drugs and Technologies in Health (CADTH); MEDLINE; MEDLINE Epub Ahead of Print; MEDLINE In-Process; MEDLINE Daily Update; PubMed; and Embase. Reference lists of included studies and systematic reviews were screened to identify any additional relevant studies. The Cochrane Library search strategy and results of all searches are provided in Appendix 1 of the supplementary material.

Inclusion Criteria

Published and unpublished MTCs were included when they reported on adults (at least 18 years of age) with a confirmed diagnosis of RRMS or rapidly evolving severe (RES) RRMS, treated with any form of pharmaceutical treatment. Inclusion was limited to studies published during or after 2010, but no language restriction was imposed.

Methods of Study Selection, Quality Assessment and Data Extraction

Two reviewers independently assessed studies for inclusion and extracted data on the inclusion criteria, systematic review methods, analysis methods and results (relapse, disability progression and safety) for each MTC; any discrepancies were resolved by discussion or by reference to a third reviewer.

We present a narrative synthesis comparing the scope, systematic review methods, analysis methods and findings of MTCs in MS.

RESULTS

Literature searches yielded 1627 references. After removing duplicates ($n = 367$) and references published before 2010 ($n = 261$), a total of 999 references were available for screening. Titles and abstracts were screened and 56 potentially relevant papers were ordered as full texts. One MTC was identified checking the reference lists of full-text papers. Of these, 45 references (relating to 27 individual MTCs) were included (Fig. 1).

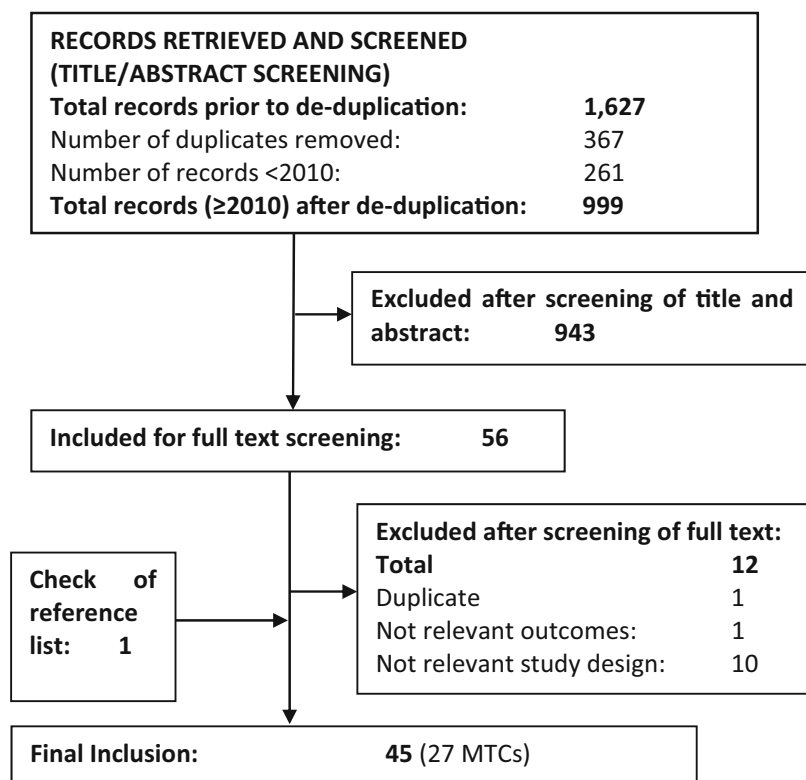


Fig. 1 Flow chart of study searches and inclusion. MTC mixed treatment comparison

Comparison of Inclusion Criteria and Methodology

The majority of MTCs included patients with RRMS (Table 1). Two [17, 18] included studies in patients with all subtypes of MS, and four [19–22] included studies which enrolled patients with relapsing MS, encompassing RRMS, secondary progressive multiple sclerosis with relapses, progressive-relapsing MS or combinations thereof. These six studies did not provide subgroup analyses of patients with RRMS only. Two MTCs [23, 24] focused on RES and highly active RRMS, while two [25, 26] focused on RRMS, with the stipulation that the population in the included trials must be at least 80% patients with RRMS. The CADTH 2013 [27] study also focused on trials in patients with RRMS but specified that the population should be more than 50% RRMS.

Many of the MTCs did not report full details of their methodologies. In 12 of the studies [17, 19, 21, 24, 28–35] omissions were identified

in reporting of inclusion screening, data extraction, quality assessment or discrepancy resolution. Table 1 details the inclusion criteria and methods used in the systematic review underlying the MTCs.

Treatments Assessed by Included MTCs

Each MTC included different DMTs, in part due to the timing of new treatments becoming available after previous MTCs had been conducted (Table 2). Studies also focussed on different aspects of DMTs formulation or mechanism of action. For example, Tolley [25] only included treatments that were delivered by subcutaneous injection, resulting in the exclusion of alemtuzumab, dimethyl fumarate, fingolimod, natalizumab and teriflunomide; while Filippini [18] focused on immunomodulators and immunosuppressants for all subtypes of MS. The majority of MTCs focussed on monotherapies; however, three [18, 27, 36] also

Table 1 Inclusion criteria for and systematic review methods of included MTCs

Study	Population	Last search	Included studies	Inclusion screening	Data extraction	Quality assessment	Discrepancy resolution
Li (2019) [37]	RRMS	10/2018	23	Double	Double	Cochrane ROB tool, 2 reviewers	Discussion (all review authors)
McCool (2019) [19]	Relapsing MS	07/2017	33	Double	NR	Cochrane ROB tool, NR	Third reviewer
Hamidi (2018) [20]	RRMS	11/2015	48	Double	Single + checking	Cochrane ROB tool, number of reviewers NR	Discussion + third reviewer
Lucchetta (2018) [38, 39]	RRMS	05/2017	37	Double	Double	Cochrane ROB 2.0 tool, 2 reviewers	Discussion + third reviewer
Martina (2018) [28]	RRMS	NR	NR	NR	NR	NR	NR
Salas (2018) [29]	RRMS	01/2018	13	Double	NR	Cochrane ROB tool, 2 reviewers	Discussion
Siddiqui (2018) [40]	RRMS	01/2017	44	Double	Double	NICE, IQWiG, HAS checklists and Jadad score. Number of reviewers NR	Third reviewer
Sultanov (2018) [30]	RRMS ^f	NR	5	NR	NR	NR	NR
Xu (2018) [36]	RRMS	09/2017	14	Double	Double	Cochrane ROB tool, NR	Consultation
Huisman (2017) [23]	HA and RES RRMS	11/2014	5	Double	Single + checking	NICE checklist, number of reviewers NR	Third reviewer
ICER (2017) [41]	RRMS	09/2016	33 ^e	Single	Single	US Preventive Services Task Force criteria, NR	Second reviewer
Melendez-Torres (2017) [42]	Relapsing MS, CIS	10/2015	32	Double	Double	Cochrane ROB tool, 2 reviewers	Discussion + third reviewer
Couto (2016) [43]	RRMS	11/2015 ^c	37	Double	Single + checking	NKCHS ROB tool, single + checking	Discussion
Fogarty (2016) [31]	RRMS	09/2014	28	NR	NR	Cochrane ROB tool, number of reviewers NR	NR
Guo (2016) [24]	RRMS with RES disease	09/2015	NR	NR	NR	NICE checklist, number of reviewers NR	NR
Aggarwal (2015) [17]	MS ^a	NR	10	NR	NR	NR	NR
Tolley (2015) [25]	RRMS ^d	03/2014	39	Double	Double	Jadad score, 2 reviewers	Third reviewer
Tramacere (2015) [44]	RRMS	09/2014	39	Double	Double	Cochrane ROB tool, 3 reviewers	Third reviewer
Zagnutt (2015) [32]	RRMS	NR	3	NR	NR	NR	NR

Table 1 continued

Study	Population	Last search	Included studies	Inclusion screening	Data extraction	Quality assessment	Discrepancy resolution
Hutchinson (2014) [26]	RRMS ^d	11/2012	34	Double	Double	Jadad score, 2 reviewers	Third reviewer
CADTH (2013) [27]	RRMS ^b	10/2013	30	Double	Single + checking	SIGN-50 major items, 2 reviewers	Discussion
Cutter (2013) [33]	RRMS	11/2012	30	NR	NR	NR	NR
Filippini (2013) [18]	RRMS; SPMS, PRMS and PPMS combined	02/2012	44	Double	Double	Cochrane ROB tool, 2 reviewers	Discussion
Hadjigeorgiou (2013) [21]	Relapsing MS (RRMS, SPMS with relapses, PRMS or combinations)	11/2012	48	Double	Double	NR	NR
Del Santo (2012) [34]	RRMS	NR	10	NR	NR	NR	NR
Tyas (2012) [35]	RRMS	NR	NR	NR	NR	CRD criteria, number of reviewers NR	NR
Zintzaras (2012) [22]	Relapsing MS (RRMS, SPMS with relapses, PRMS or combinations)	01/2011	109	Double	Double	Assessed blinding and randomisation, number of reviewers NR	Third reviewer

CADTH Canadian Agency for Drugs and Technologies in Health, *GIS* clinically isolated syndrome, *CRD* Centre for Reviews and Dissemination, *HA* highly active, *HAS* Haute Autorité de Santé, *ICER* Institute for Clinical and Economic Review, *IQWiG* Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, *MS* multiple sclerosis, *MTC* mixed treatment comparison, *NICE* National Institute for Health and Care Excellence, *NKCHS* Norwegian Knowledge Centre for the Health Services, *NR* not reported, *PPMS* primary progressive multiple sclerosis, *PRMS* progressive-relapsing multiple sclerosis, *RES* rapidly evolving severe, *ROB* risk of bias, *RRMS* relapsing–remitting multiple sclerosis, *SIGN* Scottish Intercollegiate Guidelines Network, *SPMS* secondary progressive multiple sclerosis, *US* United States

^a Reported as an abstract, appears to include MS and subtypes

^b Included trials involve at least 50% patients with RRMS

^c Updated search of CADTH 2013

^d Included trials involve at least 80% patients with RRMS

^e Real-world evidence studies plus 14 RCTs (only 4 of which explicitly referenced)

^f Only reported as abstract

Table 2 Treatments assessed by included MTCs

Study	Cladribine	Ocrelizumab	Daclizumab	Natalizumab	PEG IFNβ1a	DMF	Alemtuzumab	Teriflunomide	Fingolimod	Azathioprine	Mitoxantrone
Li (2019) [37]	-	✓	✓	✓	✓	✓	✓	✓	✓	-	✓
McCool (2019) [19]	✓✓	✓	✓	✓	✓	✓	✓	✓✓	✓	-	-
Hamidi (2018) [20]	-	-	-	✓	✓	✓	✓	✓	✓	-	-
Lucchetta (2018) [38, 39]	✓✓	✓	✓	✓	✓	✓✓	✓✓	✓✓	✓✓	-	-
Martina (2018) [28]	-	-	-	✓	-	-	-	-	✓✓	-	-
Salas (2018) [29]	-	-	✓	✓	-	-	✓	✓	✓	-	-
Siddiqui (2018) [40]	✓	✓	✓	✓	✓	✓	✓	✓✓	✓	-	-
Sultanov (2018) [30]	-	-	-	✓	-	-	-	-	-	-	-
Xu (2018) [36]	-	✓	✓	✓	-	-	✓	-	-	-	-
Huisman (2017) [23]	-	-	-	✓	-	✓	-	-	✓	-	-
ICER (2017) [41]	-	✓	✓	✓	✓	✓	✓	✓✓	✓	-	-
Melendez-Torres (2017) [42]	-	-	-	-	✓	-	-	-	-	-	-
Couto (2016) [43]	-	-	-	✓	✓✓	✓✓	✓	✓✓	✓✓	-	-
Fogarty (2016) [31]	-	-	-	✓	✓	-	✓	✓	✓	-	-
Guo (2016) [24]	-	-	-	✓	-	-	✓	-	-	-	-
Aggarwal (2015) [17]	-	-	-	✓	-	-	-	-	✓	-	-
Tolley (2015) [25]	-	-	-	-	✓	-	-	-	-	-	-
Tramacere (2015) [44]	-	-	✓✓	✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓	✓
Zagnutt (2015) [32]	-	-	-	-	-	✓✓	-	✓✓	-	-	-
Hutchinson (2014) [26]	-	-	-	✓	-	✓	-	✓✓	✓	-	-
CADTH (2013) [27]	-	-	-	✓	-	✓	✓✓	✓✓	✓	-	-
Cutter (2013) [33]	-	-	-	-	-	-	-	✓✓	-	-	-
Filippini (2013) [18]	-	-	-	✓	-	-	-	✓✓	-	✓✓	✓✓
Hadjigeorgiou (2013) [21]	-	-	-	✓	-	-	✓	✓✓	✓	-	-
Del Santo (2012) [34]	-	-	-	✓	-	-	-	-	✓	-	-
Tyas (2012) [35]	-	-	-	-	-	-	-	-	✓	-	-
Zintzaras (2012) [22]	-	-	-	✓✓	-	✓✓	✓✓	✓	✓✓	✓	✓✓

Table 2 continued

Study	Glatiramer acetate	IFN β 1a SC	IFN β 1a IM	IFN β 1b	Immunoglobulins ^a	Corticosteroids ^a	Cyclophosphamide ^b	Laquinimod ^b	Methotrexate ^b	Placebo
Li (2019) [37]	✓	✓	✓	✓	-	-	-	-	-	✓
McCool (2019) [19]	✓	✓	✓	✓	-	-	-	-	-	✓
Hamidi (2018) [20]	✓	✓	✓	✓	-	-	-	-	-	✓
Lucchetta (2018) [38, 39]	✓	✓	✓	✓	-	-	-	-	-	✓
Martina (2018) [28]	✓	✓	✓	✓	-	-	-	-	-	-
Salas (2018) [29]	✓	✓	✓	✓	-	-	-	-	-	-
Siddiqui (2018) [40]	✓	✓	✓	✓	-	-	-	-	-	✓
Sultanov (2018) [30]	✓	-	-	-	-	-	-	-	-	✓
Xu (2018) [36]	-	-	✓ ^c	-	-	-	-	-	-	✓
Huisman (2017) [23]	-	-	-	-	-	-	-	-	-	✓
ICER (2017) [41]	✓	✓	✓	✓	-	-	-	-	-	-
Melendez-Torres (2017) [42]	✓	✓	✓	✓	-	-	-	-	-	✓
Couto (2016) [43]	✓	✓	✓	✓	-	-	-	-	-	✓
Fogarty (2016) [31]	✓	✓	✓	✓	-	-	-	-	-	✓
Guo (2016) [24]	-	-	-	-	-	-	-	-	-	✓
Aggarwal (2015) [17]	✓	✓ ^d	-	-	-	-	-	-	-	✓
Tolley (2015) [25]	✓	✓	✓	✓	-	-	-	-	-	✓
Tramacere (2015) [44]	✓	✓	✓	✓	✓	-	-	✓	-	✓
Zagnutt (2015) [32]	✓	-	-	-	-	-	-	-	-	✓
Hurchinson (2014) [26]	✓	✓ ^c	-	-	-	-	-	-	-	✓
CADTH (2013) [27]	✓	✓	✓	✓	-	-	-	-	-	✓
Cutter (2013) [33]	✓	✓	✓	✓	-	-	-	-	-	✓
Filippini (2013) [18]	✓	✓	✓	✓	✓	✓	✓	-	✓	✓
Hadjigeorgiou (2013) [21]	✓	✓	✓	✓	-	-	-	-	-	✓
Del Santo (2012) [34]	✓	✓	✓	✓	-	-	-	-	-	✓

Table 2 continued

Study	Cladribine		Ocrelizumab Fingolimod	Daclizumab Azathioprine		Natalizumab Mitoxantrone		PEG IFNβ1a		DMF
	Teriflunomide							Tyas (2012) [35]		
Alentuzumab	✓								✓	
Zinzaras (2012) [22]	✓		✓	✓	✓	✓	✓			✓

Papers have been arranged in descending order and treatments have been rearranged according to date of EMA approval from most recent (left) to the oldest treatments (right)

✓ Included (single dose of the drug), ✓✓ Included (multiple doses of the drug), CADTH Canadian Agency for Health, DMF dimethyl fumarate, EMA European Medicines Agency, IGER Institute for Clinical and Economic Review, IFN interferon, IM intramuscular, MTC mixed treatment comparison, PEG polyethylene glycol, SC subcutaneous

^a Corticosteroids and immunoglobulins include a wide range of treatments therefore no explicit date of EMA approval can be found; no EMA license, off-label use

^b No EMA license, off-label use

^c Alone or in combination with natalizumab

^d Type and dose of interferon not reported

^e Pooled IFN

allowed the inclusion of combination treatments.

Further variation arose because of reviews limiting/restricting the inclusion of studies using treatment doses approved/licensed in particular countries/regions (e.g. Tolley [25] and Hutchinson [26] limited inclusion to treatment doses that were approved in the European Union or the USA). In CADTH 2013 [27], inclusion was limited to doses that were licensed in Canada and approved by Health Canada plus a selection of emerging treatments at that time (alemtuzumab, dimethyl fumarate, teriflunomide) that were not limited by dose.

Of the most recent MTCs published, the study by McCool [19] included daclizumab, which, although approved at the time of the analysis, was subsequently withdrawn by the manufacturer because of concerns about the benefit–risk profile. This MTC also included cladribine 5.25 mg/kg, a dosing regimen which is off-label for the treatment of MS. Removal of these drugs/doses from the networks may potentially lead to different results and conclusions.

Annualised Relapse Rate

Annualised relapse rate (ARR), often included in clinical trials as a surrogate for future disability, was the most widely reported outcome in the MTCs identified and was therefore chosen to compare methods and results related to it. Table 3 presents an overview of methods used in the analyses. A summary of the available results for all treatments compared with placebo is presented in Table 4.

In 19 of the MTCs [19, 20, 23–29, 31, 33, 35–43] relapse was analysed on the basis of the ARR. Appropriate data on MS subtype were more commonly reported in the more recent studies and were often unavailable for older treatments. In patients with high disease activity despite previous treatment, there were insufficient data to form a network, as shown in the study by Huisman [23].

Differences were observed in the statistical approaches applied in measuring ARR between the MTCs. Fourteen [19, 20, 24–27, 31, 35,

Table 3 MS relapse: analysis methods

Study	Outcome	# studies	# treatments	Method	Software	Primary model	Input data	Data type	Follow-up
Li (2019) [37]	ARR ^d	23	13	Bayesian	Stata	Fixed effect, random effects	Total relapses	Poisson	2 years
McCool (2019) [19]	ARR	33	17	Bayesian	R	Random effects	Proportion relapse-free	Poisson	≥ 12 weeks
Hamidi (2018) [20]	ARR	NR	19	Bayesian	WinBUGS	Random effects	Total relapses, total person-years	Poisson	NR
Lucchetta (2018) [38, 39]	ARR	37	15	Bayesian	R	Fixed effect, random effects	NR	Poisson	2 years
Martina (2018) [28]	ARR	NR	9	Bayesian	WinBUGS, Stata, R and Microsoft Projects	Fixed effect, random effects	Number of relapses, treatment exposure time in person-years	NR	18 months
Salas (2018) [29]	NA ^c								
Siddiqui (2018) [40]	ARR	41	NA	Bayesian	WinBUGS	Random effects ^a	NR	NR	NR
Sultanov (2018) [30]	Proportion experiencing at least one relapse	5	2	NR	Stata	NR	NR	NR	NR
Xu (2018) [36]	ARR	14	7	Frequentist	Stata	NR	NR	NR	24–144 weeks

Table 3 continued

Study	Outcome	# studies	# treatments	Method	Software	Primary model	Input data	Data type	Follow-up
Huisman (2017) [23]	ARR	5	4 ^c	Bayesian	OpenBUGS, R	Fixed effect	Rate ratio, SE	Log transformed rate ratio	NR
ICER (2017) [41]	ARR	33	17	Bayesian	WinBUGS	Fixed effect, random effects	Number of relapses, treatment exposure time in person-years	Poisson	≥ 2 years
Melendez-Torres (2017) [42]	ARR	NR	8	Frequentist	Stata	Random effects	Total relapses, total person-years	Poisson	NR
Couto (2016) [43]	ARR	NR	20	Bayesian	WinBUGS	Random effects	Total relapses, total person-years	Poisson	≥ 16 weeks
Fogarty (2016) [31]	ARR	28	13	Bayesian	WinBUGS, R	Fixed effect	Total relapses, total person-years	Poisson	NR
Guo (2016) [24]	ARR	NR	3	Bayesian	NR	Random effects	Total relapses, total person-years	Poisson	NR
Aggarwal (2015) [17]	Proportion experiencing relapse	10	6	Bayesian	WinBUGS	Random effects	NR	NR	NR
Tolley (2015) [25]	ARR	16	7	Bayesian	WinBUGS	Random effects ^a	Total relapses, total person-years	Poisson	≥ 6 months
Tramacere (2015) [44]	Proportion experiencing new relapse	26	14	Frequentist	Stata	Random effects	Number with relapse, total number analysed	Binary	≥ 16 weeks
Zagmutt (2015) [32]	NA ^c								

Table 3 continued

Study	Outcome	# studies	# treatments	Method	Software	Primary model	Input data	Data type	Follow-up
Hutchinson (2014) [26]	ARR	22	8	Frequentist	SAS	Random effects	Total relapses, total person-years	Poisson	≥ 6 months
CADTH (2013) [27]	ARR	27	14	Bayesian	WinBUGS	Random effects ^a	Total relapses, total person-years	Poisson	≥ 16 weeks
Currer (2013) [33]	ARR	30	6	Bayesian	NR	NR	NR	NR	NR
Filippini (2013) [18]	Recurrence of relapse	16	11	Bayesian	WinBUGS	Random effects	Number with relapse, total number analysed	Binary	2 years ^b
Hadjigeorgiou (2013) [21]	Proportion not experiencing relapse	43	10	Bucher indirect comparison	NET-MS	Random effects	Number without relapse, total number analysed	Binary	NR
Del Santo (2012) [34]	Proportion not experiencing relapse	10	7	NR	ITC (CADTH)	NR	Relapse-free rate	Time to event	12 months
Tyas (2012) [35]	ARR	NR	7	NR	NR	NR	Total relapses	Poisson	NR
							Total person-years		

Table 3 continued

Study	Outcome	# studies	# treatments	Method	Software	Primary model	Input data	Data type	Follow-up
Zintzaras (2012) [22]	Proportion not experiencing relapse	82	92	Bucher indirect comparison	NET-MS	Random effects	Number with relapse, total number analysed	Binary	NR

AE adverse event, *ARR* annualised relapse rate, *CADTH* Canadian Agency for Drugs and Technologies in Health, *GLM* generalised linear model, *GLMM* generalised linear mixed model, *HA* highly active, *ICER* Institute for Clinical and Economic Review, *ITC* Indirect Treatment Comparison software, *MTC* mixed treatment comparison, *NA* not applicable, *NR* not reported, *RES* rapidly evolving severe, *RRMS* relapsing–remitting multiple sclerosis, *SE* standard error, *WinBUGS* Bayesian inference Using Gibbs Sampling (for Microsoft Windows)

- ^a Fixed effect model included for comparison
- ^b Results for 12 months were also reported for the subgroup of trials with patients with RRMS only
- ^c Stratified by HA and RES
- ^d Bi-annualised relapse rate
- ^e MTC of AEs

37–39, 41–43] analysed ARR as a Poisson outcome based on the total number of relapses and the total number of person-years of follow-up, while six [17, 18, 21, 22, 34, 44] used a different approach by considering recurrence of relapse as a binary outcome based on the number of patients with relapse and the total number of patients. Four MTCs [[17, 18, 21, 22] reported the results as odds ratios (ORs), two [37, 44] as risk ratios, and three [34, 38, 39] as hazard ratios. This approach meant that the results of these MTCs are not directly comparable to each other, or to other MTCs using a rate ratio as the outcome measure.

In ten out of 16 MTCs the analysis was based on using Bayesian methods to fit generalised linear models [18, 19, 25, 27–29, 33, 38, 39, 41, 43]. Hutchinson [26] also used linear models but applied frequentist methods in SAS software to fit the model. Two MTCs [36, 44] used linear models applying frequentist methods using the Stata statistical software. Six other studies used Bayesian methods but did not report further details [17, 20, 23, 24, 31, 40] while Melendez-Torres 2017 [42] used frequentist methods without providing further details. Two studies [21, 22] used the Bucher method for indirect comparison, and three studies [30, 34, 35] did not report the methods for MTC. Zagmutt [32] assessed adverse events (AEs) rather than relapses so was not included in this comparison. In 14 of 19 MTCs, the primary analysis used a random-effects model. A fixed-effects model was the primary analysis for seven studies [23, 28, 31, 37–39, 41] and was also reported for comparison in three [25, 27, 40]. Three studies [33–35] did not report their statistical model while Zagmutt 2015 [32] included no MTC of ARR.

Findings

Overall, a comparison of all identified MTCs showed that the findings of each individual MTC were not directly comparable with the others because of differences in the DMTs compared, included studies, effect measures and analysis methods (Table 4). Nevertheless, the estimated treatment effects relative to placebo

Table 4 MS relapse—results

Treatment vs. placebo	Effect measure	IFN beta-1a, 30 mcg, once weekly	IFN beta-1a, 22 mcg, tiw	IFN beta-1a, 44 mcg, tiw	PEG IFN beta-1a, 125 mcg, q2w	IFN beta-1b, 250 mcg EOD	GA, 20 mg od	GA, 40 mg tiw	Teriflumonide, 7 mg, od	Teriflumonide, 14 mg, od
Li (2019) [37]	Risk ratio	0.87 (0.77–0.99)	0.81 (0.72–0.90) ^b	0.63 (0.52–0.77)	0.63 (0.52–0.77)	0.81 (0.72–0.91)	0.80 (0.72–0.89) ^b		NR	0.78 (0.66 to 0.92)
McCool (2019) [19]	Rate ratio	0.82 (0.74–0.92)	NR	0.63 (0.53–0.74)	0.64 (0.49–0.84)	0.69 (0.61–0.79)	0.67 (0.59–0.74)	0.66 (0.53–0.82)	0.78 (0.69–0.90)	0.66 (0.58 to 0.76)
Hamidi (2018) [20]	Odds ratio	NR ^b								
Luchetta (2018) [38, 39]	Hazard ratio	0.84 (0.77–0.93)	NR	0.69 (0.60–0.79)	0.64 (0.51–0.80)	0.70 (0.61–0.82)	0.69 (0.62–0.76)	0.66 (0.55–0.78)	0.73 (0.63–0.85)	0.69 (0.58 to 0.81)
Marrina (2018) [28]	Rate ratio (SE)	0.783 (0.07)	0.766 (0.08)	0.748 (0.08)	NR	0.700 (0.07)	NR	0.601 (0.05)	NR	NR
Salas (2018) [29]	Odds ratio	NR ^c								
Siddiqui (2018) [40]	Rate ratio	NR ^d								
Sultanov (2018) [30]	NR	NR ^b								
Xu (2018) [36]	Relative risk	NR	NR	NR	NR	NR	NR	NR	NR	NR
Huisman (2017) [23]	Odds ratio	NR	NR	NR	NR	NR	NR	NR	NR	NR
ICER (2017) [41]	Rate ratio	0.83 (0.74–0.94)	0.70 (0.55–0.85)	0.64 (0.54–0.73)	0.63 (0.47–0.86)	0.65 (0.55–0.77)	0.63 (0.55–0.71)	0.67 (0.52–0.86)	0.77 (0.67–0.93)	0.67 (0.56 to 0.79)
Melendez-Torres (2017) [42]	Rate ratio	0.80 (0.72–0.88)	0.72 (0.61–0.85)	0.68 (0.60–0.76)	0.64 (0.50–0.83)	0.69 (0.62–0.76)	0.66 (0.54–0.80)	NR	NR	NR
Couto (2016) [43]	Rate ratio	0.82 (0.73–0.91)	0.69 (0.57–0.83)	0.64 (0.56–0.72)	0.65 (0.49–0.85)	0.66 (0.57–0.76)	0.65 (0.59–0.73)	0.66 (0.52–0.82)	0.77 (0.68–0.90)	0.67 (0.58 to 0.77)
Fogarty (2016) [31]	Risk ratio	0.85 (0.78–0.93)	0.72 (0.63–0.82)	0.67 (0.60–0.74)	0.64 (0.51–0.80)	0.67 (0.60–0.75)	0.65 (0.59–0.71)	0.65 (0.55–0.78)	NR	0.67 (0.59 to 0.75)
Guo (2016) [24]	Rate ratio	NR ^e								
Aggarwal (2015) [17]	Odds ratio	NR ^f					NR ^g	NR	NR	NR
Tolley (2015) [25]	Rate ratio	0.74 (0.65–0.84)	0.71 (0.59–0.86)	0.66 (0.57–0.75)	0.65 (0.49–0.87)	0.68 (0.58–0.80)	0.65 (0.57–0.73)	NR	NR	NR

Table 4 continued

Treatment vs. placebo Study	Effect measure	IFN beta-1a, 30 mcg, once weekly	IFN beta-1a, 22 mcg, tiw	IFN beta-1a, 44 mcg, tiw	IFN beta-1a, 125 mcg, q2w	IFN beta-1b, 250 mcg EOD	GA, 20 mg od	GA, 40 mg tiw	Teriflunomide, 7 mg, od	Teriflunomide, 14 mg, od
Tramacet (2015) [44]	Risk ratio	0.91 (0.82–1.02) ^b	NR	0.86 (0.77–0.95) ^h	0.89 (0.70–1.13) ^{h,i}	0.85 (0.77–0.94) ^h	0.83 (0.75–0.91) ^h	NR	0.88 (0.75–1.03) ^b	0.89 (0.81 to 0.98) ^h
Zagmutt (2015) [32]	Odds ratio	NA ^c								
Hutchinson (2014) [26]	Rate ratio	NR ⁱ	NR	NR ⁱ	NR	NR	NR	NR	0.53 (0.45–0.62)	NR
CADTH (2013) [27]	Rate ratio	0.87 (0.76–0.98)	0.71 (0.60–0.83)	0.67 (0.59–0.76)	NR	0.67 (0.59–0.77)	0.67 (0.60–0.74)	NR	0.69 (0.56–0.83)	0.68 (0.56 to 0.83)
Cutter (2013) [33]	Risk ratio	0.78 (0.67–0.91)	NR	0.62 (0.51–0.76)	NR	0.68 (0.52–0.87)	0.64 (0.53–0.76)	NR	0.76 (0.66–0.89)	0.67 (0.57 to 0.77)
Filippini (2013) [18]	Odds ratio	1.10 (0.69–1.82)	NR	0.46 (0.25–0.71)	NR	0.50 (0.31–0.82)	0.50 (0.29–0.71)	NR	1.06 (0.59–1.89) ^r	
Hadjigeorgiou (2013) [21]	Rate ratio	0.82 (0.73–0.91)	0.69 (0.57–0.83)	0.64 (0.56–0.72)	0.65 (0.49–0.85)	0.66 (0.57–0.76)	0.65 (0.59–0.73)	0.66 (0.52–0.82)	0.77 (0.68–0.90)	0.67 (0.58 to 0.77)
DeI Santo (2012) [34]	Hazard ratio	NR ^k					1.28 (1.02–1.62) ^s		NR	NR
Tyas (2012) [35]	Rate ratio	NR	NR	NR	NR	NR	NR	NR	NR	NR
Zintzaras (2012) [22]	Odds ratio	NR ⁱ								
Treatment vs. placebo Study		Dimethyl Fumarate, 240 mg bid	Fingolimod, 0.5 mg od	Dactizumab, 150 mg, q4w	Natalizumab, 300 mg, q4w	Alentuzumab, 12 mg od	Cladribine	Ocrelizumab		
Li (2019) [37]		0.65 (0.56–0.74)	0.57 (0.50–0.65)	NR	0.51 (0.43–0.61)	0.49 (0.40–0.59)	NR	0.49 (0.40–0.61)		
McCool (2019) [19]		0.50 (0.42–0.59)	0.46 (0.40–0.54)	0.45 (0.38–0.54)	0.32 (0.26–0.39)	0.28 (0.22–0.35)	0.42 (0.33–0.55)	0.34 (0.26–0.43)		
Hamidi (2018) [20]										
Lucchetti (2018) [38, 39]		0.48 (0.42–0.55)	0.47 (0.41–0.53)	0.47 (0.42–0.54)	0.31 (0.27–0.36)	0.31 (0.26–0.38)	0.42 (0.34–0.52)	0.37 (0.31–0.46)		
Martina (2018) [28]		NR	0.413 (0.05)	NR	0.407 (0.07)	NR	NR	NR		
Salas (2018) [29]										

Table 4 continued

Treatment vs. placeboStudy	Dimethyl Fumarate, 240 mg bid	Fingolimod, 0.5 mg od	Daclizumab, 150 mg, q4w	Natalizumab, 300 mg, q4w	Alemtuzumab, 12 mg od	Cladribine	Ocrelizumab
Siddiqui (2018) [40]							
Sultanov (2018) [30]							
Xu (2018) [36]	NR	NR	0.13 (0.10–0.19)	0.12 (0.09–0.16)	0.09 (0.06–0.14)	NR	0.13 (0.09–0.20)
Huisman (2017) [23]	HA: 0.57 (0.39–0.84)	HA: 0.52 (0.40–0.68)	NR	RES: 0.25 (0.16–0.39)	NR	NR	NR
		RES: 0.43 (0.24–0.76)					
ICER (2017) [41]	0.53 (0.43–0.63)	0.46 (0.39–0.55)	0.46 (0.38–0.58)	0.31 (0.25–0.40)	0.28 (0.22–0.35)	NR	0.35 (0.27–0.44)
Melendez-Torres (2017) [42]	NR	NR	NR	NR	NR	NR	NR
Couto (2016) [43]	0.50 (0.42–0.60)	0.46 (0.39–0.54)	NR	0.30 (0.24–0.36)	0.29 (0.23–0.35)	NR	NR
Fogarty (2016) [31]	0.50 (0.43–0.59)	0.47 (0.41–0.53)	NR	0.31 (0.27–0.36)	0.31 (0.26–0.36)	NR	NR
Guo (2016) [24]							
Aggarwal (2015) [17]	NR	NR ^a	NR	NR ^a	NR	NR	NR
Tolley (2015) [25]	NR	NR	NR	NR	NR	NR	NR
Tramacere (2015) [44]	0.72 (0.64–0.81) ^b	NR	0.56 (0.47–0.66) ^b	0.46 (0.38–0.55) ^b	NR	NR	NR
Zagmutt (2015) [32]							
Hutchinson (2014) [26]	NR	NR	NR	NR	NR	NR	NR
CADTH (2013) [27]	0.50 (0.42–0.59)	0.44 (0.37–0.53)	NR	0.32 (0.26–0.38)	0.30 (0.25–0.37)	NR	NR
Cutter (2013) [33]	NR	NR	NR	NR	NR	NR	NR
Filippini (2013) [18]	NR	NR	NR	0.31 (0.19–0.55)	NR	NR	NR
Hadjigeorgiou (2013) [21]	0.50 (0.42–0.60)	0.46 (0.39–0.54)	NR	0.30 (0.24–0.36)	0.29 (0.23–0.35)	NR	NR
Del Santos (2012) [34]	NR	1.29 (1.18–1.40) ^a	NR	1.38 (1.24–1.53) ^a	NR	NR	NR
Tyas (2012) [35]	NR	0.43 (NR)	NR	NR	NR	NR	NR

Table 4 continued

Treatment vs. placeboStudy	Dimethyl Fumarate, 240 mg bid	Fingolimod, 0.5 mg od	Dacizumab, 150 mg, q4w	Natalizumab, 300 mg, q4w	Alemtuzumab, 12 mg od	Cladribine	Ocrelizumab
----------------------------	-------------------------------	-----------------------	------------------------	--------------------------	-----------------------	------------	-------------

Zintzaras (2012) [22]

AE adverse event, *bid* twice daily, *CADTH* Canadian Agency for Drugs and Technologies in Health, *CI* confidence interval, *CrI* credible interval, *EOD* every other day, *GA* glatiramer acetate, *HA* highly active, *HR* hazard ratio, *ICER* Institute for clinical and economic review, *IFN* interferon, *mg* milligram, *MS* multiple sclerosis, *MTC* mixed treatment comparison, *NA* not applicable, *NR* not reported, *od* once daily, *PEG* polyethylene glycol, *q2w* once every 2 weeks, *q4w* once every 4 weeks, *RES* rapidly evolving severe, *SE* standard error, *tiw* three times a week

Bolded cells denote results in favour of the intervention, representing a statistically significant difference/difference (in Bayesian terms). **Italicized** cells denote results in favour of the intervention where the 95% *CI*/*CrI* did include 1, representing no statistically significant difference/difference (in Bayesian terms)

Bold *Italicized* cells denote results in favour of placebo where the 95% *CI*/*CrI* did include 1, representing no statistically significant difference/difference (in Bayesian terms)

^a Both doses combined

^b No results for comparison versus placebo reported

^c *MTC* of *AEs*

^d Only results for comparison to cladribine reported

^e Results for comparison of natalizumab and alemtuzumab

^f Type and dose of interferon not reported

^g Dose not reported

^h In multi-arm studies which compared the same treatment at different doses, results were merged into a single-arm by summing the number of events and sample size, i.e. these results may not correspond directly to the doses indicated in the row labels

ⁱ Results for 12 months follow-up

^j Results of four *IFNs* were pooled and expressed as a single comparator

^k *HR* available for combination of *IFNs* beta 1a and 1b

^l *IFN* beta-1b 250 mcg *EOD* was used as the common comparator, no effect estimates for comparisons against placebo reported

were generally in the same direction for those treatment effects that were reported by multiple studies. This trend was consistent across outcomes and across populations, but the magnitude of the effects and the associated uncertainty varied as a result of the differences between MTCs (e.g. risk ratio of 0.32 for natalizumab in CADTH [27], compared with 0.46 in Tramacere [44]). In particular, Filippini [18] was an outlier, with the ORs for IFN β 1a and teriflunomide favouring placebo, in contrast to other MTCs of the same treatments.

The results available from nine MTCs [19, 21, 22, 24, 26, 29, 30, 36, 40] were limited in applicability to decision-making as they reported the relative effects of one drug compared with other treatments in line with the objectives of the studies, and therefore did not present results for all treatments compared with placebo. Three studies [17, 26, 34] combined all of the IFN β 1a treatments into a single treatment class, while two [18, 44] merged treatments in studies comparing the same agent at different doses but summing the number of events and the sample size.

DISCUSSION

Twenty-seven MTCs were identified. As detailed above, variation in the clinical question to be answered, as defined by the population, intervention, comparator, outcomes (PICO) framework, accounted for the differences in included populations, treatments assessed and methods of analysis used. Many of these differences have a potential impact on the findings of the review or possibly limit the relevance of a particular review for decision-makers. However, the estimated treatment effects in comparison with placebo were generally in the same direction across MTCs of the same treatments (Table 4). A number of examples of the differences between MTCs, and the potential implications, are discussed below.

Population

Consistent more generally with other neurological diseases, diagnosis of MS poses

challenges, which, together with changing clinical criteria, have resulted in an evidence base for treatment which has a high level of heterogeneity in study populations (the ‘Will Rogers phenomenon’) [45]. In addition, outcome measures and definitions have changed over time such that two studies, conducted 20 years apart and both reporting relapse outcomes, may not measure the same clinical state [8, 11]. The increase in treatment options has also influenced the types of patients included in more recent clinical trials. This creates considerable limitations in comparing the effectiveness of new treatments to existing treatment options, where the previous evidence base may largely comprise heterogeneous studies. Therefore, authors of MTCs should carefully examine the inclusion criteria and characteristics of study participants in the identified studies, to ensure the subsequent analyses uphold similarity assumptions. For example, whether the definition of “serious adverse event” included or excluded relapse of MS differed between RCTs. It would be advisable to investigate the effect of this issue in sensitivity analyses, and it may be useful to involve clinical input to ensure that all assumptions are valid, and that all-important disease modifiers, characteristics or biases are investigated.

Five of the identified MTCs included a trial in which 19% of patients had a diagnosis of clinically isolated syndrome [46]. While inclusion of studies with a limited proportion of participants with other subtypes of MS is acceptable, that proportion should be relatively small (e.g. less than 15% [47]), to reduce the potential impact of including results of patients with higher or lower disease stage which might have a different treatment response compared with the population of interest.

Intervention and Comparators

Traditional meta-analyses compare only a few, often two, treatments while MTCs enable the comparison of all relevant treatments in a certain disease. In many cases, the selection of treatments assessed by an MTC is related to the date of publication; as treatments and trials for

MS have evolved over time, we have witnessed a corresponding natural evolution of MTCs from simple networks involving IFN and glatiramer acetate to more complex networks involving newer products such as natalizumab and fingolimod. As many treatments in MS have never been directly compared, MTCs should ideally include all relevant treatments, thus presenting the reader with a comprehensive overview of all available treatment options while increasing the connectivity of the network. If treatments are excluded, the reasons for choosing particular treatments should be transparent and the rationale for exclusion documented. Regular updates are also needed to include newly available treatments and data.

Five of the identified MTCs combined individual DMTs into classes or combined different doses of a single treatment before conducting the MTC. Drugs grouped into the same class may still differ by structure, pharmacokinetics and mode of action and these differences translate into clinical practice in terms of both efficacy and tolerability. Therefore, no two drugs are really the same and grouping them into a similar class requires many, often limiting, assumptions and risks a potential loss of important information about the comparative benefits and harms [7]. It should be noted that the results obtained by this approach would be applicable only to a hypothetical average treatment rather than any of the individual treatment doses. It is likely that this decision will also have an impact on the relative effect of other treatments in the same network. If treatments are combined, e.g. to increase sample size, this should be clearly documented, and the effect explored in sensitivity analyses.

Outcome

Careful consideration of the outcome definition is required in terms of its description but also with regards to the type of data (discrete or continuous) and whether intention-to-treat or per-protocol analysis was employed, ensuring similarity assumptions. This is especially important if comparing results between different reviews, and clinical opinion is useful for

establishing the best outcome to measure effectiveness or safety.

When analysing endpoints, such as AEs, as a binary outcome, studies can only be combined in the analysis if they measure the outcome at the same point in time. Two studies [27, 43] appear to have combined results for the endpoints of confirmed disease progression after 3 and 6 months (CDP3M, CDP6M), respectively (data not shown). Combining endpoints makes it much harder, or even impossible, to compare the outcome with other studies and to interpret the results. It could be assumed that the results would be mainly driven by CDP3M as this is more widely reported. As different disability criteria and lengths of confirmation can cause variation in observed results, the definition of disease progression should be clearly reported [48].

As shown in Table 3, 13 MTCs analysed ARR as a Poisson outcome based on the total number of relapses and the total number of person-years of follow-up. These values were frequently not reported in the primary studies and therefore would have required imputation. The total number of relapses was typically imputed from the reported ARR values and the exposure time in person-years was typically imputed from the study duration multiplied by the number of patients who completed the trial. Two MTCs [27, 43] assumed that all patients completed the trial if the percentage of completers was not reported. In the trials which did report the percentage of completers, patients who discontinued may have contributed to the total number of relapses, but would be excluded from contributing to the exposure time, therefore increasing the imputed relapse rate. In those trials where the number of completers was not reported and 100% completion was assumed, those patients who discontinued could contribute to the exposure time but could not contribute any further relapses, therefore decreasing the imputed relapse rate. Two studies [26, 31] imputed the exposure time as study duration multiplied by the number of completers, plus half of the study duration multiplied by the number who withdrew. This has a similar impact to the assumptions of the CADTH 2013 [27] study in that the patients

who discontinued did not contribute equally to the number of relapses and the exposure time. The other studies reporting ARR did not report the assumptions made to impute missing data. The net effect of these assumptions is difficult to assess as this depends on the number of studies that did or did not report specific values which cannot be determined from the published articles. Therefore, imputation should either be avoided or assumptions should be clearly reported while following standard recommendations [4, 49, 50].

Systematic Review Methods

Studies analysed in MTC should be identified by a systematic literature review in order to minimise the risk of publication bias and other biases [4, 51]. Nine MTCs did not report how many people were involved in the screening for relevant studies which could indicate a risk that relevant studies may have been missed. Twelve MTCs did not report double data extraction of studies, which increases the risk that relevant data were missed or extracted incorrectly [4].

All methods should be clearly reported, following established guidelines, to enable the reader to assess the risk of bias [15, 16, 52]. Assessment of methodological rigour is hampered when reporting is limited (e.g. five of the identified MTCs were published only as conference abstracts).

Analysis Methods

When conducting MTCs, authors can choose between a frequentist and a Bayesian framework. In general, both methods are accepted by decision-makers. However, more guidance is available for the Bayesian approach as it can incorporate previous evidence in the prior distributions [6]. While effect estimates of the two methods are comparable, different measures of uncertainties (Bayesian: credible intervals, frequentist: confidence interval) are obtained.

Similar to classical meta-analyses, authors can decide to run a fixed-effects or a random-effects analysis. While both approaches are acceptable [6], some authors advise using a

random-effects analysis to account for potential 'additional heterogeneity in an indirect comparison compared to a direct comparison' [53]. However, Bayesian random effects analyses are less applicable when there are only a few studies to compare [54].

Two reviews [18, 44] assumed 'that treated and control group participants who contributed to missing outcome data both had an unfavourable outcome'. This assumption is likely to have contributed to differences in the effect estimates compared with the other MTCs (Table 4). Ideally, such assumptions should be explored in sensitivity analyses and discussed in detail, in order to ensure transparency.

CONCLUSIONS

This paper aimed to raise awareness of issues related to the conduct and reporting of MTCs in MS to support an accurate interpretation of results by decision-makers so that limitations in the robustness of analyses can be factored into an understanding of the applicability and relevance of findings.

Clear and transparent reporting of methods, results and assumptions made in the analyses should be presented [4, 15, 51, 52, 55]. The choice of outcome measure should be explicit and reflect standard clinical practice. Both the pooling of treatments (assumption of class effect) and the combination of outcomes measured at different points in time should ideally be avoided, or the effect explored in sensitivity analyses. Further, imputation of results without a clear justification should also be discouraged. Basic assumptions of homogeneity, similarity and consistency should be explored when conducting MTCs [14]. MTCs should include all relevant treatment options.

A second paper by Giovannoni et al. [7] presents the results of a recent MTC of DMTs for RRMS and discusses how to avoid some of the issues identified in this paper.

ACKNOWLEDGEMENTS

Richard Birnie produced tables and figures, screened search results, selected full-text studies for inclusion and performed data extraction.

Funding. Work on this systematic review was sponsored by Biogen. Biogen was given the opportunity to comment on the draft paper, but the authors had full editorial freedom. Biogen funded the journal's Rapid Service Fee.

Medical Writing and Editorial Assistance. Editorial assistance in the preparation of this article was provided by Zoe Whittington and James Davies of OPEN VIE (Marlow, UK). Funding for this assistance was provided by Biogen.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Disclosures. Maria Pia Sormani, Robert Wolff, Shona Lang, Steven Duffy, Robert Hyde, Elizabeth Kinter, Craig Wakeford, Gavin Giovannoni and Jos Kleijnen have nothing to disclose.

Compliance with Ethics Guidelines. This article is based on previously conducted studies and does not contain any studies with human participants or animals performed by any of the authors.

Data Availability. All data generated or analyzed during this study are included in this published article/as supplementary information files.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide

a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. European Medicines Agency. Guideline on clinical investigation of medicinal products for the treatment of Multiple Sclerosis. EMA/CHMP/771815/2011, Rev. 2. March 2015. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500185161.pdf. Accessed 3 Feb 2020.
2. Claflin SB, Broadley S, Taylor BV. The effect of disease modifying therapies on disability progression in multiple sclerosis: a systematic overview of meta-analyses. *Front Neurol*. 2019;9:1–25.
3. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105–24.
4. Higgins J, Thomas J, Chandler J, et al. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. (updated March 2011) Cochrane. www.training.cochrane.org/handbook. Accessed 7 Aug 2014.
5. Laws A, Kendall R, Hawkins N. A comparison of national guidelines for network meta-analysis. *Value Health*. 2014;17(5):642–54.
6. Coleman CI, Phung OJ, Cappelleri JC, et al. Use of mixed treatment comparisons in systematic reviews. Rockville: Agency for Healthcare Research and Quality; 2012.
7. Giovannoni G, Lang S, Wolff R, et al. A systematic review and mixed treatment comparison of pharmaceutical interventions for multiple sclerosis. *Neurol Ther*. 2020. <https://doi.org/10.1007/s40120-020-00212-5>.

8. Poser CM, Paty DW, Scheinberg L, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol*. 1983;13(3):227–31.
9. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. *Ann Neurol*. 2001;50(1):121–7.
10. Polman CH, Reingold SC, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald criteria”. *Ann Neurol*. 2005;58(6):840–6.
11. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol*. 2011;69(2):292–302.
12. Zhang Y, Salter A, Wallström E, Cutter G, Stüve O. Evolution of clinical trials in multiple sclerosis. *Ther Adv Neurol Disord*. 2019;12:1756286419826547.
13. Tonin FS, Rotta I, Mendes AM, Pontarolo R. Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. *Pharm Pract*. 2017;15(1):943.
14. Song F, Loke YK, Walsh T, Glenny A-M, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating health-care interventions: survey of published systematic reviews. *BMJ*. 2009;338:b1147.
15. Hutton B, Salanti G, Caldwell DM, 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(11):777–84.
16. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health*. 2011;14(4):417–28.
17. Aggarwal S, Kumar S, Topalogu H. Comparison of network meta-analysis and traditional meta-analysis for prevention of relapses in multiple sclerosis. *Value Health*. 2015;18(7):A660.
18. Filippini G, Del Giovane C, Vacchi L, et al. Immunomodulators and immunosuppressants for multiple sclerosis: a network meta-analysis. *Cochrane Database Syst Rev*. 2013;6:CD008933.
19. McCool R, Wilson K, Arber M, et al. Systematic review and network meta-analysis comparing ocrelizumab with other treatments for relapsing multiple sclerosis. *Mult Scler Relat Disord*. 2019;29:55–61.
20. 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(2):88–105.
21. Hadjigeorgiou GM, Doxani C, Miligkos M, et al. A network meta-analysis of randomized controlled trials for comparing the effectiveness and safety profile of treatments with marketing authorization for relapsing multiple sclerosis. *J Clin Pharm Ther*. 2013;38(6):433–9.
22. Zintzaras E, Doxani C, Mprotsis T, Schmid CH, Hadjigeorgiou GM. Network analysis of randomized controlled trials in multiple sclerosis. *Clin Ther*. 2012;34(4):857–69.
23. Huisman E, Papadimitropoulou K, Jarrett J, et al. Systematic literature review and network meta-analysis in highly active relapsing-remitting multiple sclerosis and rapidly evolving severe multiple sclerosis. *BMJ Open*. 2017;7(3):e013430.
24. Guo J, Das Gupta R, Fahrback K, Wissinger E, Cox F. A network meta-analysis comparing alemtuzumab to natalizumab in patients with relapsing-remitting multiple sclerosis with rapidly evolving severe disease. *Mult Scl*. 2016;22(3 Suppl):772.
25. Tolley K, Hutchinson M, You X, 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(6):e0127960.
26. Hutchinson M, Fox RJ, Havrdova E, et al. Efficacy and safety of BG-12 (dimethyl fumarate) and other disease-modifying therapies for the treatment of relapsing-remitting multiple sclerosis: a systematic review and mixed treatment comparison. *Curr Med Res Opin*. 2014;30(4):613–27.
27. Canadian Agency for Drugs and Technologies in Health. Comparative clinical and cost-effectiveness of drug therapies for relapsing-remitting multiple sclerosis. CADTH. 2013. <https://www.ncbi.nlm.nih.gov/books/NBK169748/>. Accessed 3 Feb 2020.
28. Martina R, Jenkins D, Bujkiewicz S, Dequen P, Abrams K, GetReal Workpackage 1. The inclusion of real world evidence in clinical development planning. *Trials*. 2018;19(1):468.
29. Salas PAO, Parra CO, Florez CEP, Goetz LM, Velez-van-Meerbeke A, Rodriguez JH. Safety liver profile of teriflunomide versus interferon β in multiple sclerosis: systematic review and indirect

- comparison meta-analysis. *Mult Scler Relat Disord*. 2018;26:192–200.
30. Sultanov M, Almadiyeva A, Absattarova K, Semenova Y. Natalizumab versus glatiramer acetate in relapsing-remitting multiple sclerosis: a network meta-analysis. *Value Health*. 2018;21(Suppl 3):S331.
 31. Fogarty E, Schmitz S, Tubridy N, Walsh C, Barry M. 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.
 32. Zagmutt FJ, Carroll CA. Meta-analysis of adverse events in recent randomized clinical trials for dimethyl fumarate, glatiramer acetate and teriflunomide for the treatment of relapsing forms of multiple sclerosis. *Int J Neurosci*. 2015;125(11):798–807.
 33. Cutter G, Fahrbach K, Huelin R, Wissinger E, Kitio-Dschassi B, Dive-Pouletty C. Relative efficacy of teriflunomide 14 mg in relapsing MS: a mixed-treatment comparison. *Mult Scler*. 2013;19(10):1413.
 34. Del Santo F, Maratea D, Fadda V, Trippoli S, Messori A. Treatments for relapsing-remitting multiple sclerosis: summarising current information by network meta-analysis. *Eur J Clin Pharmacol*. 2012;68(4):441–8.
 35. Tyas D, Roskell N, Zimovetz E, Rycroft C, Eckert B. Annualized relapse rate of first-line treatments for multiple sclerosis: a meta-analysis, including indirect comparisons versus fingolimod. *Neurology*. 2012;78(1 Suppl).
 36. Xu X, Chi S, Wang Q, et al. Efficacy and safety of monoclonal antibody therapies for relapsing remitting multiple sclerosis: a network meta-analysis. *Mult Scler Relat Disord*. 2018;25:322–8.
 37. Li H, Hu F, Zhang Y, Li K. Comparative efficacy and acceptability of disease-modifying therapies in patients with relapsing-remitting multiple sclerosis: a systematic review and network meta-analysis. *J Neurol*. 2019. <https://doi.org/10.1007/s00415-019-09395-w>.
 38. Lucchetta RC, Tonin FS, Borba HHL, et al. Disease-modifying therapies for relapsing-remitting multiple sclerosis: a network meta-analysis. *CNS Drugs*. 2018;32(9):813–26.
 39. Lucchetta RC, Tonin FS, Borba HHL, et al. Immunomodulators and immunosuppressants for patients with relapsing-remitting multiple sclerosis: a network meta-analysis. *Value Health*. 2018;21(Suppl 1):S203–4.
 40. 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*. 2018;34(8):1361–71.
 41. Institute for Clinical Economic Review. Disease-modifying therapies for relapsing remitting and primary-progressive multiple sclerosis: effectiveness and value final evidence report. ICER. 2017. https://icer-review.org/wp-content/uploads/2016/08/CTAF_MS_Final_Report_030617.pdf. Accessed 3 Feb 2020.
 42. Melendez-Torres GJ, Auguste P, Armoiry X, et al. Clinical effectiveness and cost-effectiveness of beta-interferon and glatiramer acetate for treating multiple sclerosis: systematic review and economic evaluation. *Health Technol Assess*. 2017;21(52):1–352.
 43. Couto E, Hamidi V, Ringerike T, Odgaard-Jensen J, Harboe I, Klemp M. Medicines used for multiple sclerosis—a health technology assessment. Oslo: Knowledge Centre for the Health Services at The Norwegian Institute of Public Health (NIPH); 2016.
 44. 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*. 2015;9:CD011381.
 45. Sormani MP. The Will Rogers phenomenon: the effect of different diagnostic criteria. *J Neurol Sci*. 2009;287(1 Suppl):S46–9.
 46. O’Connor P, Filippi M, Arnason B, 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*. 2009;8(10):889–97.
 47. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). General methods: version 5.0. IQWiG. 2016. <https://www.iqwig.de/en/methods/methods-paper.3020.html>. Accessed 3 Feb 2020.
 48. Kalincik T, Cutter G, Spelman T, et al. Defining reliable disability outcomes in multiple sclerosis. *Brain*. 2015;138(Pt 11):3287–98.
 49. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials*. 2007;8:16.
 50. Suthar A, Lawn S, del Amo J, et al. Antiretroviral therapy for prevention of tuberculosis in adults

- with HIV: a systematic review and meta-analysis. *PLoS Med.* 2012;9(7):e1001270.
51. Centre for Reviews and Dissemination B. Systematic reviews. CRDs guidance for undertaking reviews in health care. CRD. 2009. https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf. Accessed 3 Feb 2020.
 52. Hoaglin DC, Hawkins N, Jansen JP, 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 Health.* 2011;14(4):429–37.
 53. Haute Autorite de Sante. Indirect comparisons: methods and validity. HAS. 2009. https://www.has-sante.fr/portail/upload/docs/application/pdf/2011-02/summary_report__indirect_comparisons_methods_and_validity_january_2011_2.pdf. Accessed 3 Feb 2020.
 54. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A (Stat Soc).* 2009;172(1):137–59.
 55. Jansen JP, Trikalinos T, Cappelleri JC, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health.* 2014;17(2):157–73.