PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A systematic literature review and network meta-analysis in highly active relapsing remitting multiple sclerosis and rapidly evolving severe multiple sclerosis.
AUTHORS	Huisman, Eline; Papadimitropoulou, Katerina; Jarrett, James; Bending, Matthew; Firth, Zoe; Allen, Felicity; Adlard, Nick

VERSION 1 - REVIEW

REVIEWER	Eva Havrdová Charles University, First Medical Faculty, Dp. of Neurology and Center for Clinical Neuroscience, MS Center, Czech Republic
REVIEW RETURNED	25-Jul-2016

GENERAL COMMENTS	Authors did a huge and sofisticated statistical exercise to conclude that comparisons across treatments in highly active or rapidly evolving severe RR MS patients are associated with high levels of uncertainty until new data is collected for these subgroups. As clinical trials were aimed to prove the efficacy in a wide variety of patients, these subgroups were not predefined. In future these patients will be mostly treated, not enrolled in clinical trials. The comparison with alemtuzumab is lacking, and though authors explain why, the whole work seems invalid because of it as it provides no guidance for clinicians. In most recommendations or treatment guidelines alemtuzumab and natalizumab are considered to be more effective than fingolimod or dimethyl fumarate, therefore the result of no statistical significant difference between natalizumab and dimethyl fumarate seems not very believable and is not supported by real world data from MSBase. Though well written and a lot of work done this review bring no new knowledge and nothing for clinical practice.

REVIEWER	Elizabeth Sweeney Rice University Unites States
REVIEW RETURNED	12-Aug-2016

GENERAL COMMENTS	Summary: This paper is a meta-analysis of randomized clinical trials for HA and RES RRMS in the UK. The authors use a Bayesian network meta- analysis to compare fingolimod to dimethyl fumarate in HA RRMS and natilizamaub in RES RRMS. The authors follow appropriate guidelines and procedures for the meta-analysis. Due to a small number of trials, the authors cannot make any conclusions for the indirect comparisons in the HA and RES groups. This seems appropriate for the journal BMJ Open, which is open to publishing negative studies.

Major Issues

1) Is there a reason to restrict this study to just the UK? It seems like you could say more and the paper would be more interesting if you used other geographical areas as well.

2) The statistical analysis is unclear (please see comments below). Please provide a more detailed description of the models. Also the citation of the statistical literature is a bit thin.

Minor Issues:

1) You never introduce the abbreviation ARR in the text. It is contained in Table 3, but this should be introduced in text (issue in abstract too).

Incorrect grammar:

1) 'The records title and abstract was'

Statistical Analysis

1) "In this analysis, a linear model with normal likelihood distribution was used to model 3-month and 6-month confirmed disability progression at 24 months (as this is a continuous outcome), and a Poisson likelihood with log link for the annualized relapse rate at 24 months (as this is a Poisson/rate outcome). For both types of outcomes, a value of hazard ratio or risk ratio for the intervention versus the control of lower than 1 indicated greater efficacy."

Comments: Here you say for both types of outcomes a value of hazard ratio / risk ratio lower than 1 indicated grater efficacy – this doesn't make sense for the linear model (for which any statistically significant fixed effect > 0 would indicate greater efficacy)? Please rephrase this paragraph so that it is clearer / easier to understand.

2) You say that you fit a model with fixed and random effects – what were the fixed effects and the random effects? It's not clear at all what models you are using here. Please state explicitly the models you are using along with what your fixed and random effects are (preferably with some notation and equations).

Tables and Figures:

Table 3:

I'm confused here about the median ranking. Does this mean that you have more ARR at 24 months in the fingoloid group. Please explain in the caption what this ranking is (also what is "P (best)"?).

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name
Eva Havrdová

Institution and Country
Charles University, First Medical Faculty, Dp. of Neurology and Center for Clinical Neuroscience, MS Center, Czech Republic

Please state any competing interests or state 'None declared':
None declared

Please leave your comments for the authors below
Authors did a huge and sophisticated statistical exercise to conclude that comparisons across treatments in highly active or rapidly evolving severe RR MS patients are associated with high levels of uncertainty until new data is collected for these subgroups. As clinical trials were aimed to prove the efficacy in a wide variety of patients, these subgroups were not predefined. In future these patients will be mostly treated, not enrolled in clinical trials.
The comparison with alemtuzumab is lacking, and though authors explain why, the whole work seems invalid because of it as it provides no guidance for clinicians. In most recommendations or treatment guidelines alemtuzumab and natalizumab are considered to be more effective than fingolimod or dimethyl fumarate, therefore the result of no statistical significant difference between natalizumab and dimethyl fumarate seems not very believable and is not supported by real world data from MSBase.
Though well written and a lot of work done this review bring no new knowledge and nothing for clinical practice.

Response:
Thank-you for the comments on the choice of comparison and we have a justification for this selection. The SLR and NMA feasibility revealed the scarcity of available subgroup data in multiple sclerosis trials which is a key message of this study. Although subgroup data have been considered for evaluation of disease-modifying treatments by Health Technology Assessment bodies (such as NICE) for alemtuzumab, these data are not publically available and could not be incorporated in this analysis. Our research utilized all publically available data identified from a rigorously performed SLR which meets the requirements of PRISMA. The view shared by BMJ Open is that negative or "non-inferior" studies should be published and we therefore recommend that it is important these results are shared with respect to the challenges in making these comparisons.

Reviewer: 2

Reviewer Name
Elizabeth Sweeney

Institution and Country
Rice University
Unites States

Please state any competing interests or state ‘None declared’:
None declared

Please leave your comments for the authors below

Summary: This paper is a meta-analysis of randomized clinical trials for HA and RES RRMS in the UK. The authors use a Bayesian network meta- analysis to compare fingolimod to dimethyl fumarate in HA RRMS and natalizumab in RES RRMS. The authors follow appropriate guidelines and procedures for the meta-analysis. Due to a small number of trials, the authors cannot make any conclusions for the indirect comparisons in the HA and RES groups. This seems appropriate for the journal BMJ Open, which is open to publishing negative studies.

Major Issues

1) Is there a reason to restrict this study to just the UK? It seems like you could say more and the paper would be more interesting if you used other geographical areas as well.

Response: Thank-you for this observation on the geographical remit of our study. The data used in the NMA are not from UK studies and agree that although many HTA appraisals using NMA have been performed in the UK setting it is misleading to have this restriction. We did not intend this and have therefore eliminated UK-specific references in the revised version of the manuscript.

2) The statistical analysis is unclear (please see comments below). Please provide a more detailed description of the models. Also the citation of the statistical literature is a bit thin.

Thank you for your comment; we have provided more details below and revised the methods section accordingly.

Minor Issues:

1) You never introduce the abbreviation ARR in the text. It is contained in Table 3, but this should be introduced in text (issue in abstract too).

Thank-you, we have now added the abbreviation.

Incorrect grammar:

1) ‘The records title and abstract was’

Corrected.

Statistical Analysis

1) “In this analysis, a linear model with normal likelihood distribution was used to model 3-month and 6-month confirmed disability progression at 24 months (as this is a continuous outcome), and a Poisson likelihood with log link for the annualized relapse rate at 24 months (as this is a Poisson/rate

outcome). For both types of outcomes, a value of hazard ratio or risk ratio for the intervention versus the control of lower than 1 indicated greater efficacy."

Comments: Here you say for both types of outcomes a value of hazard ratio / risk ratio lower than 1 indicated grater efficacy – this doesn't make sense for the linear model (for which any statistically significant fixed effect > 0 would indicate greater efficacy)? Please rephrase this paragraph so that it is clearer / easier to understand.
Response: Apologies for not making the interpretation clear enough in the text. Please see below our response and we have reformulated the text of the revised version accordingly.

The models use the log hazard ratio which is assumed to follow a Normal distribution. The modeling takes place in the log scale and not the natural scale and the results are interpreted in hazard ratio and risk (rate) ratio scale where a positive outcome is translated to hazard ratio and a risk ratio lower than one. Suggest to rephrase as follows:

"Trial results were reported as trial-based summary measures, i.e., 3-month and 6-month confirmed disability progression at 24 months were reported as hazard ratios and annualized relapse rate at 24 months were reported as risk ratios. In these cases, we assumed a Normal distribution for the continuous measure of the treatment effect. The modeling is performed in the log scale. The output of the analyses are summary measures i.e, hazard ratios and risk ratios of the treatment of interest vs the comparator. A value equal to 1 translates to no difference between the competing treatments and a value lower than 1 translated to greater efficacy (lower hazard and/or lower risk of relapse)".

2) You say that you fit a model with fixed and random effects – what were the fixed effects and the random effects? It's not clear at all what models you are using here. Please state explicitly the models you are using along with what your fixed and random effects are (preferably with some notation and equations).

Response: Network meta-analyses can be performed with fixed-or random effects models. The former assumes that there is no variation in the relative treatment effects across studies for a particular pairwise comparison. The observed differences for a particular comparison among study results are solely due to chance. The general fixed-effects model for network meta-analysis can be specified as follows:

η_{jk}= μ_{jb} if b=A,B,C for k=b or η_{jk}=μ_{jb} + d_{bk}=μ_{jb}+d_{Ak}-d_{Ab} for k=B,C,D if k is "after" than b
where μ_{jb} is the outcome for treatment b in study j, and d_{bk} is the fixed effect of treatment k relative to treatment b.
The random effects model assumes that the true relative effects are exchangeable across studies and can be described as a sample from a Normal/Gaussian distribution whose mean is the pooled relative effect and SD reflects the heterogeneity. The model notation of the random effects model is as follows:
η_{jk}= μ_{jb} if b=A,B,C for k=b or η_{jk}=μ_{jb} + δ_{jbk} for k=B,C,D if k is "after" than b

where δ_{jbk} are the trial-specific effect of treatment k relative to treatment b. These trial-specific effects are drawn from a random-effects distribution with the following properties:
δ_{jbk} ~ N (d_bk, σ^2).

Tables and Figures:

Table 3:

I'm confused here about the median ranking. Does this mean that you have more ARR at 24 months in the fingolimod group? Please explain in the caption what this ranking is (also what is "P (best)"?).

Response: We have reformulated the text as follows to explain how the ranking was calculated:

"The Bayesian NMA provided joint posterior distributions of the relative treatment effects across interventions accompanied with pairwise probabilities of one treatment being better than another for each of the outcomes. These probabilities were calculated based on the proportion of MCMC cycles in which a specific treatment estimate was better than the comparator and can be interpreted as there is x% probability that treatment A is better than treatment B. The ranking probabilities are summarized by a median and an associated 95% CrI. Additional ranking outcomes monitored are the probability of being best (Pbest) and SUCRA. The former is calculated as the proportion of MCMC cycles which a given treatment ranks first out of all competing interventions. The SUCRA measure was calculated as Surface Under the Cumulative Ranking Curve (SUCRA); SUCRA is 1 when a treatment is certain to be the best and 0 when a treatment is certain to be the worst."

Regarding your specific question on Table 3 please see our response:

The median rank in Table 3 can be interpreted as follows: Fingolimod can be ranked as first or second best treatment among its competitors but in the majority of the MCMC cycles it ranked first. These ranking probabilities are not associated with greater efficacy in the specific outcome. They provide an overview of ranking of the best treatment. P best is calculated as the proportion of cycles in which a given treatment ranks first out of the all competing treatments. P best should sum to 100% when looking at the competing interventions. For ARR at 24 months, Fingolimod ranked as the first treatment with probability of being the best treatment equal to 64% followed by the second best treatment, DMF which showed 36% probability of being the best.