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)	Comparative efficacy and safety of approved treatments for macular oedema secondary to branch retinal vein occlusion: a network meta-analysis
AUTHORS	Regnier, Stephane; Larsen, Michael; Bezlyak, Vladimir; Allen, Felicity

VERSION 1 - REVIEW

REVIEWER	John Ford University of East Anglia, UK
REVIEW RETURNED	21-Jan-2015

GENERAL COMMENTS	Regnier and colleagues present a network meta-analysis of treatments for macular oedema secondary to branch retinal vein occlusion. This is an important topic area. Several drugs are currently used for this indication and network meta-analyses which combine different drugs are needed. The authors present a technically high-quality analysis of data. It includes a patient level analysis to improve the heterogeneity of the studies. This inclusion of unpublished data is welcome. However I have some significant concerns. Primary bevacizumab has not been included. Whilst bevacizumab is not currently licensed, it is used by clinicians. Research and evidence should drive licensing decisions, not the other way around. Most readers would expect a NMA of “treatments for macular oedema secondary to BRVO” to include bevacizumab. Either the analysis needs revised to include the bevacizumab RCTs, or it needs to be made explicit in the title, abstract and throughout the manuscript which treatments are being assessed. Secondly results are not reported in a balanced manner. • page 2 line 35, the authors report that ranibizumab was numerically superior to aflibercept. This is misleading. The authors should be explicit that there was no statistically significant difference between ranibizumab and aflibercept. • Page 2 line 30-34, the authors only report the probability of letters gained and not proportion of patients gaining > 15 letters. • Abstract conclusion should state that there was no statistically significant difference between aflibercept and ranibizumab. If they authors wish to report likelihood in the abstract, they should also refer to the likelihood of the proportion of patients increasing by >15 letters.

	• Results section needs revised to be more balanced and should explicitly state that there was no statistically significant difference between aflibercept and ranibizumab. • The discussion should be revised to be more balanced. Page 20 line 32, refers to ranibizmab being “numerically, but not statistically, superior to aflibercept”. This should be replaced by words to the effect of “there was no statistically significant difference in either letters gained or proportion of patients gaining more than 15 letters.” • Conclusions should be revised to be more balanced. It is not clear if table 2 includes published data or the data from patient-level data analysis for the three ranibizumab trials. The patient-level summary baseline characteristics and results should be included, if they have not already been. Did the authors look at proportion of patients losing more than 15 letters? It would be helpful is diagrams were presented for each of the analysis, including sensitivity to make it clearer to the reader. The competing interests statement should make it clear that Novartis Pharma AG and Genentech are manufacturers and distributors of ranibizumab.

REVIEWER	Jesse Berlin Johnson & Johnson, USA
REVIEW RETURNED	16-Feb-2015

GENERAL COMMENTS	GENERAL COMMENTS: Overall, the paper is well written and the methods appear to be valid. However, some of the conclusions may need to be a bit more tempered. 1. The number of studies is VERY limited. This has to have made it hard to assess consistency. For some analyses, you have fairly wide confidence intervals. In general, I would strongly suggest adding these points when you discuss the limitations. 2. I’ll have more on this below, but would suggest that it would be very helpful to provide confidence intervals (if it’s possible to do so) for the probabilities of being the best. SPECIFIC COMMENTS 3. I don’t understand a key point at this point in the paper. You mention that triamcinolone is used off-label, but don’t say if you included it or not. You go on to say that bevacizumab is not licensed for the indication (despite being an anti-VEGF agent) and was excluded. Is this NOT used off label? Otherwise, why would you exclude one off-label use and not the other? Or was triamcinolone also excluded but you didn’t mention that here? 4. Page 6: It is admittedly a picky semantic point, but you say the included studies had to have at least two “comparators” of interest. I would suggest saying “interventions” or “therapies,” instead. I tend to think of “comparator” as a contrast to a specific experimental treatment of interest. I may be the only one sensitive to this subtlety, so I leave this to the authors and editors to work out.

	5. Why limit yourselves to proprietary data from Novartis? Which product is sold by Novartis? Evidently a similar search of sponsor files was not performed for other products, true? Specifics would be helpful here. (Genentech got to review a draft of the paper – and they make ranibizumab. Why not go to their files?) 6. Page 7: Can this small number of studies in a network meta-analysis adequately support a random-effects model? Do you have enough data? You end up showing that FE models give the same results, despite the heterogeneity. 7. Page 9: Here, you seem to say that RABAMES met the inclusion criteria. Earlier, in the Methods section, you imply that RABAMES was only included in sensitivity analyses because the treatment regimen (3 monthly injections followed by 3 months of observation) didn't quite meet your inclusion criterion of using monthly injections. Again on page 11, I get more confused about inclusion criteria. Here you say that monthly injections were ONLY used in one trial (BRAVO), although the other studies, despite being considered PRN, pretty much had monthly injections for most subjects, based on the means you report. Then Table 2 seems to imply that several studies used 6 doses given monthly. All of this makes me wonder why you mention RABAMES as part of a sensitivity analysis. Then on page 14 you do make it clear that RABAMES had substantially fewer injections than the other studies. In any case, please clarify your strategy in the Methods section. 8. Page 11, top: Presumably you mean 1743 subjects. The current wording makes it sound a bit like you had 1743 studies. 9. Page 14: When you talk about creating imbalance on gender, are you talking about the overall groups, collapsed across studies (ignoring stratification by study)? That's what appears to be happening in Table 2. I would think that the more relevant worry would be the within-study balance, since the analysis ultimately preserves the within-study comparisons. I can't imagine that deleting 3 subjects from a trial could matter much, but it's possible, I suppose. 10. Supplementary Table 3 makes it clear that there are some studies with zero events in one or more treatment arms, but you don't mention specifically how you handle these zero cells in the analysis. Does the Bayesian approach just magically handle the problem? Please give a brief explanation. The sparseness of the data for looking at the binary outcome certainly reveals itself in the width of the confidence intervals for the results. 11. Given the sparseness of the data for the odds ratios, do you really feel comfortable reporting probabilities of being the best therapy? (I don't really have much confidence in those probabilities, based solely on the small numbers of events.) 12. I'm not seeing a table of pairwise comparisons of mean BCVA between therapies. Table 4 has the comparisons with laser therapy, but the text reports on the pairwise comparison between means for the two anti-VEGF treatments (page 16, lines 14-18). Shouldn't those pairwise comparisons of means be in a table? 13. KEY POINT: It seems inconsistent to combine the two anti-

	VEGF treatments for the safety analysis but not the efficacy analysis. In fact, the whole question of the pairwise comparison between the two anti-VEGF agents needs some careful reconsideration. You feel OK combining the agents for the safety analysis. There is a numerical difference in the means (+1.4 letters [−5.2 to 8.5), which is not statistically significant. You go on to conclude (Discussion) that "The efficacy of ranibizumab 0.5 mg PRN was numerically, but not statistically, superior to aflibercept 2q4." You make a point of mentioning the higher probability of ranibizumab being the most effective, but you don't give confidence intervals for those probabilities. Apparently, the conclusion is based on the mean difference of 1.4 letters, which is not statistically significant. I'm not familiar enough with the scale to know for sure, but have a hard time believing that 1.4 letters represents a clinically important difference. I think, at a minimum, you need to be much more cautious about your probability statements, and at least make sure to highlight the small difference and the lack of statistical significance.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1's comments

1. Primary bevacizumab has not been included. Whilst bevacizumab is not currently licensed, it is used by clinicians. Research and evidence should drive licensing decisions, not the other way around. Most readers would expect a NMA of "treatments for macular oedema secondary to BRVO" to include bevacizumab. Either the analysis needs revised to include the bevacizumab RCTs, or it needs to be made explicit in the title, abstract and throughout the manuscript which treatments are being assessed.

Response: Please see response to Editorial comment #1.

2. Page 2 line 35, the authors report that ranibizumab was numerically superior to aflibercept. This is misleading. The authors should be explicit that there was no statistically significant difference between ranibizumab and aflibercept.

Response: The abstract has been updated and is now explicit that there was no statistically significant difference between ranibizumab and aflibercept (page 2, lines 24–25 and 32–33).

3. Page 2 line 30-34, the authors only report the probability of letters gained and not proportion of patients gaining > 15 letters.

Response: The probability of being the best treatment for the subset of patients gaining ≥ 15 letters has been added to the abstract (page 2, lines 21–24).

4. Abstract conclusion should state that there was no statistically significant difference between aflibercept and ranibizumab. If they authors wish to report likelihood in the abstract, they should also refer to the likelihood of the proportion of patients increasing by ≥ 15 letters.

Response: Please see our responses to comments #2 and #3 above.

5. The Results section needs to be revised to be more balanced and should explicitly state that there was no statistically significant difference between aflibercept and ranibizumab. Response: Page 22 (lines 35–36) states: "Mean (95% CrI) BCVA letters gained from baseline for ranibizumab monotherapy over aflibercept remained non-statistically significant".

Page 23 (lines 7–9) states: "patients receiving ranibizumab monotherapy showed a non-significant mean (95% CrI) gain of 1.4 letters (−6.3 to 9.5) over aflibercept".

Page 23 (lines 9–11) states: "patients receiving ranibizumab monotherapy showed a non-significant mean (95% CrI) gain of 1.3 letters (−6.2 to 8.6) compared with those receiving aflibercept".

6. The discussion should be revised to be more balanced. Page 20 line 32, refers to ranibizumab

being “numerically, but not statistically, superior to aflibercept”. This should be replaced by words to the effect of “there was no statistically significant difference in either letters gained or proportion of patients gaining more than 15 letters.”
Response: The Discussion on page 24 (lines 19–21) states:
“there was no statistically significant difference between ranibizumab 0.5 mg PRN and aflibercept 2q4 in either letters gained or proportion of patients gaining more than 15 letters”.

7. Conclusions should be revised to be more balanced.
Response: The Conclusions (page 26, lines 19–22) now state that there was no statistical differences between ranibizumab monotherapy and aflibercept.
8. It is not clear if Table 2 includes published data or the data from patient-level data analysis for the three ranibizumab trials. The patient-level summary baseline characteristics and results should be included, if they have not already been.
Response: Table 2 shows published data for non-ranibizumab trials and data from patient-level data analysis for ranibizumab trials. A footnote was added to Table 2 to clarify the source of the data.

Supplementary Table 2 shows baseline characteristics before and after patient-level data analysis and outcomes before patient-level data analysis.
9. Did the authors look at the proportion of patients losing more than 15 letters?
Response: We did not because we had already analyzed a large number of endpoints in this study.
10. It would be helpful if diagrams were presented for each of the analyses, including sensitivity to make it clearer to the reader.
Response: Please see Figure 4 and 5.
11. The competing interests statement should make it clear that Novartis Pharma AG and Genentech are manufacturers and distributors of ranibizumab
Response: Please see our response to Editorial comment #2.

Reviewer 2’s comments
1. The number of studies is VERY limited. This has to have made it hard to assess consistency. For some analyses, you have fairly wide confidence intervals. In general, I would strongly suggest adding these points when you discuss the limitations.
Response: The limitations paragraph of the Discussion has been amended and now contains the text (page 26, lines 14–16): “Finally, data from only a limited number of trials were included in this analysis and future analyses will be strengthened once additional clinical trial data becomes available.”

The issue of the broad confidence intervals were addressed in the Discussion of the main results on page 24, lines 19-20.
2. I’ll have more on this below, but would suggest that it would be very helpful to provide confidence intervals (if it’s possible to do so) for the probabilities of being the best treatment.
Response: The probabilities of being the best treatment do not have confidence (or credible) intervals. They are based on Markov Chain Monte Carlo simulations and are simply the percentage of simulations in which a treatment had the highest treatment effect.

3. I don’t understand a key point at this point in the paper. You mention that triamcinolone is used off-label, but don’t say if you included it or not. You go on to say that bevacizumab is not licensed for the indication (despite being an anti-VEGF agent) and was excluded. Is this NOT used off label? Otherwise, why would you exclude one off-label use and not the other? Or was triamcinolone also excluded but you didn’t mention that here?
Response: This is now clarified on page 5 (lines 22–23): “Triamcinolone and bevacizumab were both excluded from this analysis”
4. Page 7: It is admittedly a picky semantic point, but you say the included studies had to have at least two “comparators” of interest. I would suggest saying “interventions” or “therapies,” instead. I

tend to think of “comparator” as a contrast to a specific experimental treatment of interest. I may be the only one sensitive to this subtlety, so I leave this to the authors and editors to work out.
Response: We agree – this has now been amended as follows (page 7, lines 13–14):
“In addition, RCTs had to comprise at least two treatment arms..”

5. Why limit yourselves to proprietary data from Novartis? Which product is sold by Novartis? Evidently a similar search of sponsor files was not performed for other products, true? Specifics would be helpful here. (Genentech got to review a draft of the paper – and they make ranibizumab. Why not go to their files?)
Response: There were no additional patient-level data owned by Genentech.
6. Page 7: Can this small number of studies in a network meta-analysis adequately support a random-effects model? Do you have enough data? You end up showing that FE models give the same results, despite the heterogeneity.
Response: This is a valid point: please see our response to comment #1. There is always a balance to be struck between the number of studies and expected heterogeneity when choosing the right model. However, since we suspected heterogeneity between the dexamethasone study and the anti-VEGF ones, we chose a random-effects model. This choice does not appear to have significantly impacted the results.
7. Page 9: Here, you seem to say that RABAMES met the inclusion criteria. Earlier, in the Methods section, you imply that RABAMES was only included in sensitivity analyses because the treatment regimen (3 monthly injections followed by 3 months of observation) didn’t quite meet your inclusion criterion of using monthly injections.
Response: RABAMES had a much lower injection frequency than the other studies. That is why the study was only included in sensitivity analysis (page 10, lines 16–18).
8. Again on page 11, I get more confused about inclusion criteria. Here you say that monthly injections were ONLY used in one trial (BRAVO), although the other studies, despite being considered PRN, pretty much had monthly injections for most subjects, based on the means you report. Then Table 2 seems to imply that several studies used 6 doses given monthly. All of this makes me wonder why you mention RABAMES as part of a sensitivity analysis. Then on page 14 you do make it clear that RABAMES had substantially fewer injections than the other studies. In any case, please clarify your strategy in the Methods section.
Response: The inclusion/exclusion criteria are clearly described in the Methods section on page 7. Inclusion criterion 4 states that patients must have received anti-VEGF therapy via a PRN or monthly regimen. We were explicit in stating that RABAMES was excluded because in this study patients received substantially fewer injections than the other studies.
9. Page 11, top: Presumably you mean 1743 subjects. The current wording makes it sound a bit like you had 1743 studies.
Response: This has now been clarified.
10. Page 14: When you talk about creating imbalance on gender, are you talking about the overall groups, collapsed across studies (ignoring stratification by study)? That’s what appears to be happening in Table 2. I would think that the more relevant worry would be the within-study balance, since the analysis ultimately preserves the within-study comparisons. I can’t imagine that deleting 3 subjects from a trial could matter much, but it’s possible, I suppose.
Response: We have deleted this section from the manuscript.
11. Supplementary Table 3 makes it clear that there are some studies with zero events in one or more treatment arms, but you don’t mention specifically how you handle these zero cells in the analysis. Does the Bayesian approach just magically handle the problem? Please give a brief explanation. The sparseness of the data for looking at the binary outcome certainly reveals itself in the width of the confidence intervals for the results.
Response: In Bayesian network meta-analyses, there is indeed no need to adjust for cells with zero value.
12. Given the sparseness of the data for the odds ratios, do you really feel comfortable reporting

probabilities of being the best therapy? (I don't really have much confidence in those probabilities, based solely on the small numbers of events.)
Response: As long as the model converges, the probabilities of being the best therapy can be reported. The fact that there a small number of events is reflected in the distribution of the events and, therefore, impacts the probability of being the best treatment.

13. I'm not seeing a Table of pairwise comparisons of mean BCVA between therapies. Table 4 has the comparisons with laser therapy, but the text reports on the pairwise comparison between means for the two anti-VEGF treatments (page 16, lines 14-18). Shouldn't those pairwise comparisons of means be in a Table?
Response: Table 5a (Pairwise odds ratios for gaining ≥ 15 letters from baseline) and 5b (Pairwise difference (95% CrI) for letters gained from baseline) have now been added to the manuscript.
14. KEY POINT: It seems inconsistent to combine the two anti-VEGF treatments for the safety analysis but not the efficacy analysis. In fact, the whole question of the pairwise comparison between the two anti-VEGF agents needs some careful reconsideration. Response: Ideally, we would have like to report the results independently but the model did not converge. This is the reason why, for IOP, we chose to analyze the class effect (namely, anti-VEGF vs corticosteroid implants vs laser vs sham).
15. You feel OK combining the agents for the safety analysis. There is a numerical difference in the means (+1.4 letters [−5.2 to 8.5), which is not statistically significant. You go on to conclude (Discussion) that "The efficacy of ranibizumab 0.5 mg PRN was numerically, but not statistically, superior to aflibercept 2q4." You make a point of mentioning the higher probability of ranibizumab being the most effective, but you don't give confidence intervals for those probabilities. Apparently, the conclusion is based on the mean difference of 1.4 letters, which is not statistically significant. I'm not familiar enough with the scale to know for sure, but have a hard time believing that 1.4 letters represents a clinically important difference. I think, at a minimum, you need to be much more cautious about your probability statements, and at least make sure to highlight the small difference and the lack of statistical significance.
Response: The conclusion and discussion has been updated to address these concerns. Please see our response to Reviewer 1 comments #5 and #6.

VERSION 2 – REVIEW

REVIEWER	John Ford UEA, UK
REVIEW RETURNED	07-Apr-2015

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript again. The authors have made it more balanced, but there are still some outstanding issues. Limitations summary on page 2 should state that bevacizumab and triamcinolone were not included Abstract should state that bevacizumab and triamcinolone were excluded Abstract – text relating to "probability of being most efficacious treatment ..." aflibercept should appear before ranibizumab in the text The methods state that bevacizmuab was included in the searches. It should be stated that studies assessing bevacizumab were not included in the systematic review. Page 21 – example of 67% probability is given for ranibizumab.

	Please repeat this for patients gaining at least 15 letters for ranibizumab versus aflibercept

REVIEWER	Jesse Berlin Johnson & Johnson, USA
REVIEW RETURNED	05-Apr-2015

GENERAL COMMENTS	The authors have done a very thorough job of responding to earlier comments. Please accept my apologies for missing some of the points that had already been in the paper. I have just a couple of additional suggestions for your consideration. 1. In your response, you say that there were no additional patient-level data owned by Genentech. Why not add that explicit comment to the text of the paper, just to be as transparent as possible? 2. In your response you say: "Ideally, we would have like to report the results independently but the model did not converge. This is the reason why, for IOP, we chose to analyze the class effect (namely, anti-VEGF vs corticosteroid implants vs laser vs sham)." In the current version, you mention the low rates of increase of IOP and the comparability of rates between the drugs. Again, why not add something explicit, along the lines of what you say in your response? You could say that the rates were low and, consequently, the models didn't converge. Because the low rates were also similar between drugs, we combined…

VERSION 2 – AUTHOR RESPONSE

Reviewer 1's comments
1. In your response, you say that there were no additional patient-level data owned by Genentech. Why not add that explicit comment to the text of the paper, just to be as transparent as possible?
Response: We added the following statement: "To our knowledge, there were no additional relevant proprietary data on file at Genentech."

2.In your response you say: "Ideally, we would have like to report the results independently but the model did not converge. This is the reason why, for IOP, we chose to analyze the class effect (namely, anti-VEGF vs corticosteroid implants vs laser vs sham)."
In the current version, you mention the low rates of increase of IOP and the comparability of rates between the drugs. Again, why not add something explicit, along the lines of what you say in your response? You could say that the rates were low and, consequently, the models didn't converge. Because the low rates were also similar between drugs, we combined…
Response: The text now reads: "The model did not converge when we consider IOP/OH rates separately for ranibizumab and aflibercept. Because the rates of increased IOP/OH were small and comparable for ranibizumab (0–5%) and aflibercept (2%), rates of increased IOP/OH for ranibizumab and aflibercept were pooled to give an anti-VEGF rate of increased IOP/OH."

Reviewer 2's comments
1.Limitations summary on page 2 should state that bevacizumab and triamcinolone were not included
Response: The limitations section now reads: "Bevacizumab and triamcinolone were not included in the study."

2. Abstract should state that bevacizumab and triamcinolone were excluded

Response: This is now stated in the abstract.

3. Abstract – text relating to “probability of being most efficacious treatment ...” aflibercept should appear before ranibizumab in the text
Response: The order is now switched.

4.The methods state that bevacizmuab was included in the searches. It should be stated that studies assessing bevacizumab were not included in the systematic review.
Response: The methods section now reads: “bevacizumab was not included in our analysis.”

5. Page 21 – example of 67% probability is given for ranibizumab. Please repeat this for patients gaining at least 15 letters for ranibizumab versus aflibercept
Response: The following sentence was added: “Based on the percentage of patients gaining at least 15 letters, the probability that ranibizumab monotherapy is a better treatment than aflibercept is 53%.”