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

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>A Comparative Assessment of OnabotulinumtoxinA and Mirabegron for Overactive Bladder: An Indirect Treatment Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Freemantle, Nicholas; Ginsberg, David; McCool, Rachael; Fleetwood, Kelly; Arber, Mick; Khalaf, Kristin; Loveman, Clara; Ni, Quanhong; Glanville, Julie</td>
</tr>
</tbody>
</table>

VERSION 1 - REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Rufus Cartwright</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperial College London, UK</td>
<td></td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>08-Jul-2015</td>
</tr>
</tbody>
</table>

GENERAL COMMENTS

In general I am opposed to systematic reviews with industry authors. It is clearly not possible that this review, performed partly in house at Allergan, would be submitted for publication if the results favoured mirabegron. Having said that, I think this review has been performed rigorously, and usefully quantifies the large difference in treatments, that although obvious from the primary studies, or from clinical practice, is not otherwise available in the literature.

Specific concerns:

1. Was the review protocol registered with Prospero or another review database? Can the review protocol be provided as supplementary material? It is important for the peer review to understand the range of outcomes that were planned to be included (efficacy and safety), and to check for deviation from inclusion criteria, or analytic strategies.

2. The review will be likely more than two years out of date by the time of publication. There are new relevant trials that could be included. Is a update possible?

3. The problem of baseline differences in severity seems intractable here. We clearly know that both drugs are likely to show larger absolute and relative improvements for patients with worse severity (see for example Eur Urol. 2015 Jan;67(1):11-4. doi: 10.1016/j.eururo.2014.06.052). My understanding of the supplementary materials is that the meta-regression could not effectively adjust for those baseline differences in UIE or urgency. This problem / limitation is not addressed at all in the main text.

4. Even after close re-reading I do not understand how the analysis accounts for variable duration of action and variable persistence. The paper concludes strongly that Botox offers efficacy benefits, but does not make it clear at what timepoint(s) this conclusion applies. Clearly one treatment is only effect if patients continue to take it, which most won't, while the other has a defined duration of action up to around 12 months. How does the analysis deal with different
The authors reported the superiority of onabotulinumtoxinA to mirabegron using new statistical methods “network meta-analysis”. The number of publication of indirect comparisons increasing. The reviewer is interested in the results that one of the new treatment for OAB may be superior to the other one in the absence of head-to-head comparisons.

Minor
I think that the term “micturition” should be changed into “urinary frequency”.

GENERAL COMMENTS

The authors present the comparison of two treatments (BTX and MBG) for overactive bladder. For this purpose, network meta-analyses (NMAs) were conducted for seven different outcomes. Results are based on data from 10 studies. Only indirect evidence is available as no direct comparisons of the two treatments have been conducted.

*** Major comments ***

(1) Potential industry bias

Research and manuscript generation was funded by Allergan, the producer of one treatment which appears to be the superior treatment. Allergan did not only provide funding but was actively involved in the review (three co-authors from the company; involvement of Allergan in exclusion of studies - see Table B.1). Accordingly, review results might be biased due to industry involvement.

(2) Comparability of studies

As stated by the authors, a crucial point in network meta-analysis is whether the included studies are sufficiently similar. In my view, this assumption is not fulfilled for the data used in the paper for the following reasons.
a) Baseline differences in disease severity / network meta-regression

This point was already mentioned and discussed by the authors. Network meta-regression based on IPD can be an adequate approach to adjust for imbalances in treatment groups. However, network meta-regressions conducted by the authors are only based on IPD from two studies comparing BTX with placebo whereas no IPD are available for MBG studies. In this situation, results from network meta-regression are of very limited value.

b) Results in patients receiving placebo

On page 12, the following percentages are reported for patients receiving placebo:
- complete continence (BTX: 6.5% - 10.7%, MBG: 36.8% - 40.5%)
- 50% reduction in daily UI episodes (BTX: 28.9% - 33.2%, MBG: 59.2% - 60.1%)

These numbers clearly demonstrate that results for untreated patients are different in BTX and MBG studies (which is an indication that BTX and MBG populations are in general different which contradicts the comparability assumption). Nevertheless, the authors state on page 15: “For the purposes of the analyses it was assumed that placebo treatment would be equally effective.”

c) Differences in percentage of women (see Table 2)

With a single exception (Sahai 2007, only 34 patients, 55.9% women), percentage of women is larger in BTX studies ranging from 84.5% to 100% as compared to MBG studies (68.7% - 89.3%). I cannot judge whether this may have any influence on NMA results.

In summary, the comparability of BTX and MBG studies is questionable and network meta-regressions to adjust for imbalances are based on a very limited amount of data. The abstract does not mention any of these limitation and is thus too optimistic.

*** Further comments ***

The literature search until end of August 2013 is outdated and should be updated.

Excluding non-english studies is certainly a source of bias as correctly stated by the authors in the discussion. I would suggest to give some information on studies published in other languages in the
discussion.

Please report number of studies and patients for each outcome in the results section and/or in Figures 3 and 4.

Adverse events have not been considered in the paper which is OK in my view. However, I would expect some information on common and severe adverse events in the discussion.

Page 5, last sentence: Please explain what is meant by “Studies that were considered similar ... were included in the network meta-analyses”.

Page 9, Assessment of heterogeneity:

The authors state that “... no suitable statistic is available to assess heterogeneity directly between the two bodies of evidence ...”. I am not sure whether I understand this statement correctly, however, statistical methods to evaluate inconsistency are available, e.g. Higgins et al. (2012), Krahn et al. (2013).

Discrepancy between Table 3 and Table E.1: Numbers of studies do not match in these tables for the first outcome: 6 vs 7

References


strategies. summarizes the eligibility criteria for the network meta-analysis and we have provided all of the details to these methods in appendices A-E, above and beyond what is usually included for such analyses. The current appendices provide extensive detail of the searches conducted, studies excluded following full-text review, the comparability assessment, statistical methodology, and summary of studies included in the networks. We have also provided the PRISMA checklist as requested by the editor.

2. The review will be likely more than two years out of date by the time of publication. There are new relevant trials that could be included. Is a update possible? Please see first comment above in response to the editorial comments.

3. The problem of baseline differences in severity seems intractable here. We clearly know that both drugs are likely to show larger absolute and relative improvements for patients with worse severity (see for example Eur Urol. 2015 Jan;67(1):11-4. doi: 10.1016/j.eururo.2014.06.052). My understanding of the supplementary materials is that the meta-regression could not effectively adjust for those baseline differences in UIE or urgency. This problem / limitation is not addressed at all in the main text. A meta-regression approach was taken to specifically address the issue of baseline differences between populations. Like all statistical models meta regression requires assumptions: in this case, we assumed that there is a linear relationship between baseline severity and change from baseline. Unfortunately individual patient data (IPD) were not available for the mirabegron trials and we only had IPD from onabotulinumtoxinA trials to work with. Thus we took a pragmatic approach rather than one of the alternatives, i.e. to completely ignore the differences between the studies, or not to conduct an NMA at all. We have therefore expanded on the Discussion (pp 15-16) as follows:

“Patients in the onabotulinumtoxinA trials had more severe urinary incontinence and urgency symptoms at baseline than those in the mirabegron trials and all patients in the onabotulinumtoxinA trials had failed antimuscarinic therapy. Our logistic regression model assessed the impact of continuous baseline UI severity on OR (onabotulinumtoxinA versus placebo) and suggested that continuous baseline UI severity did not significantly impact the OR for either 50% or 100% reduction in UI episodes/day. However, a recent analysis of the effect of categorical baseline UI severity on 100% reduction in UI episodes/day did show a trend for increasing OR (onabotulinumtoxinA versus placebo) with increasing categorical baseline UI severity (Drake et al., EAU Congress, 2015). The NMR model that we used adjusts for the baseline severity by assuming a linear relationship between baseline severity and change from baseline. This assumption was tested using individual patient data from two onabotulinumtoxinA trials. The assumption held up in this test but should still be considered a limitation of the analysis given that such patient level data were not available for mirabegron trials. Based on the data available, this approach was the most pragmatic approach possible.”

4. Even after close re-reading I do not understand how the analysis accounts for variable duration of action and variable persistence. The paper concludes strongly that Botox offers efficacy benefits, but does not make it clear at what time point(s) this We agree with your comments about the variations in duration and persistence with the treatments. Within the Results section (Included studies), Table 2 (pp7-8) provides details of duration of treatment/follow-up and the text states: “The majority of studies reported a treatment period of 12 weeks, with follow-up ranging
conclusion applies. Clearly one treatment is only effect if patients continue to take it, which most won't, while the other has a defined duration of action up to around 12 months. How does the analysis deal with different lengths of follow up, different strategies for loss to follow (LOCF vs multiple imputation)?

Reviewer #2:
1. I think that the term “micturition” should be changed into “urinary frequency”.

Thank you for this suggestion. We have made this change to the text throughout the manuscript where appropriate; in some instances “urination” is more suited to the context of the sentence and thus this is used in those situations.

Reviewer #3:
(1) Potential industry bias
Research and manuscript generation was funded by Allergan, the producer of one treatment which appears to be the superior treatment. Allergan did not only provide funding but was actively involved in the review (three co-authors from the company; involvement of Allergan in exclusion of studies - see Table B.1). Accordingly, review results might be biased due to industry involvement.

(2) Comparability of studies
As stated by the authors, a crucial point in network meta-analysis is whether the included studies are sufficiently similar. In my view, this assumption is not fulfilled for from 2 to 24 weeks or 6 months”. We also noted within the Discussion (p15), “Studies included in the analyses typically assessed short-term outcomes (12 to 13 weeks) as dictated by regulatory agencies; longer term efficacy of onabotulinumtoxinA versus mirabegron cannot be extrapolated from the data reported.”

However, you are quite correct in that we have not explicitly reported the time point at which the conclusions apply; all outcomes were analysed at 12 weeks. We have clarified this in the text as follows:

Methods (Network meta-analysis) (p5): “Studies that were considered similar and reported sufficient data at 12 weeks were included in the network meta-analyses; those that neither compared key treatments nor provided indirect information did not contribute to the networks.”

Results (Network meta-analysis) (p9): “Point estimates for fixed-effect models were consistent with those for random-effects models (see online supplementary appendix F); results of random-effects models are presented. Summary plots for onabotulinumtoxinA versus mirabegron 50 mg and 25 mg are included in Figures 3 and 4 (detailed forest plots are shown in online supplementary appendix H), with results of the analyses summarised in Table 3. All outcomes were analysed at 12 weeks.

Conclusions (pp16-17): “This indirect treatment comparison suggests that, compared with mirabegron, onabotulinumtoxinA is associated with additional reductions in the number of UI episodes and urgency episodes per day and frequency of daily micturition in patients with idiopathic OAB at 12 weeks.

We have been transparent about the authorship – the review team includes independent researchers as well as Allergan. We have adopted standard systematic review methods to conduct this review and have provided extensive supplementary materials to ensure complete transparency of all methods and decisions regarding the studies included and the methods used for modeling. If the reviewer has noticed that studies have been excluded that s/he deems eligible according to the eligibility criteria then we can discuss the reasons for exclusion. Excluded studies are described in the supplementary files.

There are inherent limitations to NMAs that are well understood, and as the reviewer mentions, well highlighted by us in the manuscript.

Please see our response to reviewer 1, comment 3.
the data used in the paper for the following reasons.

a) Baseline differences in disease severity / network meta-regression
This point was already mentioned and discussed by the authors. Network meta-regression based on IPD can be an adequate approach to adjust for imbalances in treatment groups. However, network meta-regressions conducted by the authors are only based on IPD from two studies comparing BTX with placebo whereas no IPD are available for MBG studies. In this situation, results from network meta-regression are of very limited value.

b) Results in patients receiving placebo
On page 12, the following percentages are reported for patients receiving placebo:
- complete continence (BTX: 6.5% - 10.7%, MBG: 36.8% - 40.5%)
- 50% reduction in daily UI episodes (BTX: 28.9% - 33.2%, MBG: 59.2% - 60.1%)
These numbers clearly demonstrate that results for untreated patients are different in BTX and MBG studies (which is an indication that BTX and MBG populations are in general different which contradicts the comparability assumption). Nevertheless, the authors state on page 15: “For the purposes of the analyses it was assumed that placebo treatment would be equally effective.”

We recognize that the two placebo treatments are different - one is a placebo injection and one is a placebo tablet – but in order to make a connected network we need to assume that these two types of placebo are equivalent, i.e. we are assuming that if the two different placebos were given in the same study they would have the same effect. However, we don't necessarily expect the two placebos to have the same effect in different studies. The differences in the placebo response are likely to reflect differences in the patient populations rather than differences in how the placebo was administered. In NMA differences in the placebo response between studies are not problematic. NMA is based on comparisons between treatments within studies rather than the absolute responses. With any mixed treatment comparisons there will be some difference between the trials but in the absence of a head-to-head comparison this is the highest level of evidence that can be generated based on the available data. To further clarify in the text we have made the following changes in the Discussion (p16):

The treatments themselves also differ in their mode of administration: onabotulinumtoxinA and the placebo in the onabotulinumtoxinA trials were administered as a single procedure, while both mirabegron and placebo were administered as daily oral tablets in the mirabegron trials. In order to produce a connected network for the purposes of the analyses, it was assumed that the placebo treatments would be equally effective when administered to the same populations.[45] We recognize that this assumption is potentially a source of bias in the network meta-analysis.

We are not aware of any published evidence suggesting that the effectiveness of onabotulinumtoxinA and mirabegron is different in males and females. One study in botulinum toxin found no difference in gender (PubMed ID 25582926); a further three studies (one in botulinum toxin and two in mirabegron) found the treatments worked well in men but did not compare outcomes in men and women.

c) Differences in percentage of women (see Table 2)
With a single exception (Sahai 2007, only 34 patients, 55.9% women), percentage of women is larger in BTX studies ranging from 84.5% to 100% as compared to MBG studies (68.7% - 89.3%). I cannot judge
whether this may have any influence on NMA results. Hence, we made the assumption of equal efficacy but recognize that this may represent a limitation of the analysis. We have added the following sentence to the Discussion, at the start of the paragraph discussing baseline severity issues (p15).

“There were differences in the proportion of females in the onabotulinumtoxinA studies (55.4% - 100%) compared to the mirabegron studies (68.7% - 89.3%), with all but one of the onabotulinumtoxinA studies involving a larger percentage of women. This NMA has made the assumption that sex does not affect response to treatment. Patients in the onabotulinumtoxinA trials had more severe urinary Incontinence and urgency symptoms at baseline than those in the mirabegron trials…”

In summary, the comparability of BTX and MBG studies is questionable and network meta-regressions to adjust for imbalances are based on a very limited amount of data. The abstract does not mention any of these limitation and is thus too optimistic.

In this indirect treatment comparison, baseline differences between the onabotulinumtoxinA and mirabegron studies were adjusted using network meta-regression. Results of this NMR suggest that, …”

*** Further comments ***

The literature search until end of August 2013 is outdated and should be updated. Please see first comment above in response to the editorial comments.

Excluding non-English studies is certainly a source of bias as correctly stated by the authors in the discussion. I would suggest to give some information on studies published in other languages in the discussion.

The protocol specifies English language only studies as an inclusion criterion. Hence the search strategy only searched for studies in English. We do not, therefore, know how many other studies in languages other than English might be relevant to this review. We have highlighted that this is a limitation. If the referee thinks there are some non-English language papers we have missed then it would be helpful to know. It is very common practice, although not ideal, for systematic reviews and NMAs to only review English language papers. We have revised the sentence relating to the searches (p16) as follows:

“The network meta-analysis was informed by extensive searches conducted across a range of information sources. Since only English language studies were specified in the protocol, studies published in other languages were not included in this review.”

Please report number of studies and patients for each outcome in the results section and/or in Figures 3 and 4.

The number of studies for each outcome are provided in the Results section in Table 3 Summary of outcome results for onabotulinumtoxinA (100 U) compared with mirabegron (50 mg and 25 mg) (p11-12). We have added the number of patients to this column.
Adverse events have not been considered in the paper which is OK in my view. However, I would expect some information on common and severe adverse events in the discussion.

The protocol did not specify adverse effects as an outcome so we did not extract the included studies with this in mind. The decision to exclude adverse effects was based on the fact that the safety profile for these medications is very different. We have added wording to this extent in the relevant section of the Discussion (p16).

Original text: While the adverse effect profiles for both drugs are different from those of antimuscarinic agents,[44] the current analysis did not address adverse events with either compound.

Revised text: Both onabotulinumtoxinA and mirabegron have side effects associated with treatment, albeit the adverse effect profiles are different from those of antimuscarinic agents.[44] Most common side effects with mirabegron include hypertension, nasopharyngitis, urinary tract infection and headache, while onabotA is associated with UTIs, dysuria, urinary retention, bacteriuria, and increased residual urine volume. Currently, there is little evidence comparing the safety of mirabegron and onabotulinumtoxinA with other interventions for OAB. The current analysis did not address adverse events with either compound, which represents a limitation of the review.

Page 5, last sentence: Please explain what is meant by "Studies that were considered similar ... were included in the network meta-analyses".

We have added in a link to Appendix C which contains the details, and slightly revised the text under Network meta-analysis (p5).

"Studies that were considered similar based on the modified PBAC tool (detailed in online supplementary appendix C) and reported sufficient data at 12 weeks were included in the network meta-analyses); those that neither compared key treatments nor provided indirect information did not contribute to the networks."

Page 9, Assessment of heterogeneity: The authors state that "... no suitable statistic is available to assess heterogeneity directly between the two bodies of evidence ...". I am not sure whether I understand this statement correctly, however, statistical methods to evaluate inconsistency are available, e.g. Higgins et al. (2012), Krahn et al. (2013).

Thank you for this suggestion. We agree that this statement could have been better explained. Statistical methods to evaluate inconsistency were considered, however there is no potential for inconsistency in the networks included in this manuscript. Inconsistency can arise in an NMA when there are comparisons that are informed by independent sets of direct and indirect evidence (see, for example, Lu and Ades, 2006). Our network does not contain any such comparisons. We have clarified the text in Assessment of heterogeneity (p9) by removing the sentence: "... no suitable statistic is available to assess heterogeneity directly between the two bodies of evidence ...".

The section now reads as follows: Pairwise comparisons of onabotulinumtoxinA 100 U versus placebo generally had minimal to low heterogeneity (I² < 5%), indicating that the results were similar across all studies. For urinary frequency, moderate heterogeneity (I² ≈ 50%) was observed, indicating some variation between studies above that explained by chance. The pairwise comparison of
onabotulinumtoxinA 100 U versus onabotulinumtoxinA 150 U for 100% reduction in UI episodes/day also had moderate heterogeneity ($I^2 \approx 60\%$). All comparisons of mirabegron versus placebo showed minimal heterogeneity (table 3). In addition, the networks had no potential for inconsistency between direct and indirect evidence.[43]


Discrepancy between Table 3 and Table E.1: Numbers of studies do not match in these tables for the first outcome: 6 vs 7

Thank you for pointing out this discrepancy between Table 3 and Table E.1 in terms of the number of studies for the outcome of 100% reduction in UIE, which also affects Table F. We have revisited the analyses and checked the number of studies involved.

The correct number of studies should be 7 (as reported in Table E.1); we have revised the entries in Table 3 and Table F accordingly.

**VERSION 2 – REVIEW**

**REVIEWER** Rufus Cartwright
Imperial College London, UK

**REVIEW RETURNED** 10-Dec-2015

**GENERAL COMMENTS** I think this SR is technically accurate, identifies the correct studies, and reaches a sound if fairly obvious conclusion. I have no concerns about publication in the present form.

**REVIEWER** Guido Schwarzer
Institute of Medical Biometry and Statistics
Medical Center - University of Freiburg
Germany

**REVIEW RETURNED** 07-Dec-2015

**GENERAL COMMENTS** The authors addressed my first major comment concerning potential industry bias as well as all minor comments in sufficient detail in the revision of the manuscript and / or the point-to-point reply.

However, with respect to my second major comment, comparability of BTX and MBG studies, neither is the point-to-point reply convincing nor are changes in the manuscript sufficient.

In reply to my point 2b), the authors state that "In NMA differences in the placebo response between studies are not problematic.”. However, in reply to my point 2a) and point 3) of reviewer #1, the authors say that "A meta-regression approach was taken to specifically address the issue of baseline differences between populations.”
These two statements clearly contradict each other. In my view, baseline differences between BTX and MBG studies are problematic as treatment results of individual studies depend on the results in the placebo group (there is less room for improvement under the experimental treatment in a study with a probability of 40% under placebo as compared to a study with only 10% under placebo). In principle, network meta-regression is a statistical method that can be useful to adjust for baseline imbalances. However, individual patient data from both BTX and MBG studies would be necessary to investigate this issue adequately.

I fully agree with reviewer #1 that the problem of baseline differences in severity seems intractable. It is unclear whether superior results for BTX are due to the treatment itself or due to treating more severely ill patients with BTX.

In my view, in order to be acceptable for publication, the following changes are mandatory.

Abstract - Results:
- state that severe baseline differences in severity exist between BTX and MBG studies

Strength and limitations of this study:
- network meta-analyses are only based on BTX studies and results are extrapolated to MBG studies
- magnitude of treatment difference between BTX and MBG is questionable due to substantial baseline differences in severity

In summary, I do not think about network meta-analysis that "... in the absence of a head-to-head comparison this is the highest level of evidence that can be generated based on the available data." There are the alternatives to not pool the data or to widen the network by including additional treatments for overactive bladder. If the authors do not choose any of these alternatives, the limitations of the conducted analyses must be clearly stated (in Abstract and list of strengths and limitations).

VERSION 2 – AUTHOR RESPONSE

<table>
<thead>
<tr>
<th>Reviewer Comments</th>
<th>Authors’ response (Yellow highlights represent revised/additional text to the original response/manuscript)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reviewer #1</strong></td>
<td></td>
</tr>
<tr>
<td>1. I think this SR is technically accurate, identifies the correct studies, and reaches a sound if fairly obvious conclusion. I have no concerns about publication in the present form.</td>
<td>No further action required.</td>
</tr>
<tr>
<td><strong>Reviewer #3:</strong></td>
<td></td>
</tr>
<tr>
<td>(1) The authors addressed my first major comment concerning potential industry bias as well as all minor comments in sufficient detail in the</td>
<td>No further action required.</td>
</tr>
</tbody>
</table>
| Revision of the manuscript and / or the point-to-point reply | We realise that we may have caused confusion by the wording of our responses. By ‘baseline differences’ we specifically meant differences between the BTX and MBG studies in the average number of urinary incontinence episodes per day and the average number of urgency episodes per day at week 0. We were not referring to differences in how the patients responded to baseline treatment (i.e. placebo).

The statement “In NMA differences in the placebo response between studies are not problematic” was meant as a general statement about NMA. It was not meant to refer to cases where differences in the placebo response may also impact the treatment effect. We agree that, for change from baseline in episodes of urinary incontinence and urgency, the different responses in the placebo arm are related to the differences in baseline severity.

To reduce the potential for confusion, we have adjusted our original response (please see below).

**Authors’ original responses** (italicized):
(2a) A meta-regression approach was taken to specifically address the issue of baseline differences between populations. Like all statistical models meta regression requires assumptions: in this case, we assumed that there is a linear relationship between baseline severity and change from baseline. Unfortunately individual patient data (IPD) were not available for the mirabegron trials and we only had IPD from onabotulinumtoxinA trials to work with. Thus we took a pragmatic approach rather than one of the alternatives, i.e. to completely ignore the differences between the studies, or not to conduct an NMA at all. We have therefore expanded on the Discussion (pp 15-16) as follows: “Patients in the onabotulinumtoxinA trials had more severe urinary incontinence and urgency symptoms at baseline than those in the mirabegron trials and all patients in the onabotulinumtoxinA trials had failed antimuscarinic therapy. … The NMR model that we used adjusts for the baseline severity by assuming a linear relationship between baseline severity and change from baseline. This assumption was tested using individual patient data from two onabotulinumtoxinA trials. The assumption held up in this test but should still be considered a limitation of the analysis given that such patient level data were not available for mirabegron trials. Based on the data available, this approach was the most pragmatic approach possible.”

We recognize the importance of ensuring the reader is fully aware of the limitations of the analysis and any impact these may have on the results. In view of the reviewer’s comments that results from the network meta-regression would be of limited value given that IPD were only available for BTX studies, we have brought this to the reader’s attention by incorporating a further bullet point in the list of ‘Strengths and limitations of the study’:

**Network meta-regression was applied to the changes from**

---

| (2) With respect to my second major comment, comparability of BTX and MBG studies, neither is the point-to-point reply convincing nor are changes in the manuscript sufficient. In reply to my point 2b), the authors state that "In NMA differences in the placebo response between studies are not problematic." However, in reply to my point 2a) and point 3) of reviewer #1, the authors say that "A meta-regression approach was taken to specifically address the issue of baseline differences between populations."

These two statements clearly contradict each other. In my view, baseline differences between BTX and MBG studies are problematic as treatment results of individual studies depend on the results in the placebo group (there is less room for improvement under the experimental treatment in a study with a probability of 40% under placebo as compared to a study with only 10% under placebo). In principle, network meta-regression is a statistical method that can be useful to adjust for baseline imbalances. However, individual patient data from both BTX and MBG studies would be necessary to investigate this issue adequately.

I fully agree with reviewer #1 that the problem of baseline differences in severity seems intractable. It is unclear whether superior results for BTX are due to the treatment itself or due to treating more severely ill patients with BTX.

**Reviewer’s original comments**
(2a) Baseline differences in disease severity / network meta-regression

This point was already mentioned and discussed by the authors. Network meta-regression based on IPD can be an adequate approach to adjust for imbalances in treatment groups. However, network meta-regressions conducted by the authors are only based on IPD from two studies comparing BTX with placebo whereas no IPD are available for MBG studies. In this case where differences in baseline severity are related to the differences in baseline severity, the regression approach was taken to specifically address the issue of baseline differences between populations.

---

We realise that we may have caused confusion by the wording of our responses. By ‘baseline differences’ we specifically meant differences between the BTX and MBG studies in the average number of urinary incontinence episodes per day and the average number of urgency episodes per day at week 0. We were not referring to differences in how the patients responded to baseline treatment (i.e. placebo).

The statement “In NMA differences in the placebo response between studies are not problematic” was meant as a general statement about NMA. It was not meant to refer to cases where differences in the placebo response may also impact the treatment effect. We agree that, for change from baseline in episodes of urinary incontinence and urgency, the different responses in the placebo arm are related to the differences in baseline severity.

To reduce the potential for confusion, we have adjusted our original response (please see below).

**Authors’ original responses** (italicized): (2a) A meta-regression approach was taken to specifically address the issue of baseline differences between populations. Like all statistical models meta regression requires assumptions: in this case, we assumed that there is a linear relationship between baseline severity and change from baseline. Unfortunately individual patient data (IPD) were not available for the mirabegron trials and we only had IPD from onabotulinumtoxinA trials to work with. Thus we took a pragmatic approach rather than one of the alternatives, i.e. to completely ignore the differences between the studies, or not to conduct an NMA at all. We have therefore expanded on the Discussion (pp 15-16) as follows: “Patients in the onabotulinumtoxinA trials had more severe urinary incontinence and urgency symptoms at baseline than those in the mirabegron trials and all patients in the onabotulinumtoxinA trials had failed antimuscarinic therapy. … The NMR model that we used adjusts for the baseline severity by assuming a linear relationship between baseline severity and change from baseline. This assumption was tested using individual patient data from two onabotulinumtoxinA trials. The assumption held up in this test but should still be considered a limitation of the analysis given that such patient level data were not available for mirabegron trials. Based on the data available, this approach was the most pragmatic approach possible.”
situation, results from network meta-regression are of very limited value.

2b) Results in patients receiving placebo
On page 12, the following percentages are reported for patients receiving placebo:
- complete continence (BTX: 6.5% - 10.7%, MBG: 36.8% - 40.5%)
- 50% reduction in daily UI episodes (BTX: 28.9% - 33.2%, MBG: 59.2% - 60.1%)

These numbers clearly demonstrate that results for untreated patients are different in BTX and MBG studies (which is an indication that BTX and MBG populations are in general different which contradicts the comparability assumption).

Nevertheless, the authors state on page 15: “For the purposes of the analyses it was assumed that placebo treatment would be equally effective.”

baseline in urinary incontinence episodes and urgency episodes, in order to adjust for differences in the baseline values of these outcomes. Because individual patient data were available for onabotulinumtoxinA trials, but not for mirabegron trials, it was assumed that the relationship between baseline values and change from baseline was the same across all included trials.

(2b) We recognize that the two placebo treatments are different - one is a placebo injection and one is a placebo tablet – but in order to make a connected network we need to assume that these two types of placebo are equivalent, i.e. we are assuming that if the two different placebos were given in the same study they would have the same effect. However, we don’t necessarily expect the two placebos to have the same effect in different studies (and the model does not force the placebo response to be equal across studies). In general, in NMA, differences in the placebo response between studies are not problematic because comparators are evaluated based on treatment effect of the intervention vs placebo rather than the absolute responses. With any mixed treatment comparisons there will be some difference between the trials but in the absence of a head-to-head comparison this is the highest level of evidence that can be generated based on the available data.

To further clarify in the text, we have made the following changes in the Discussion (p16):
The treatments themselves also differ in their mode of administration: onabotulinumtoxinA and the placebo in the onabotulinumtoxinA trials were administered as a single procedure, while both mirabegron and placebo were administered as daily oral tablets in the mirabegron trials. In order to produce a connected network for the purposes of the analyses, it was assumed that the placebo treatments would be equally effective when administered to the same populations.[45] We recognize that this assumption is potentially a source of bias in the network meta-analysis.

We should have written:
To further clarify in the text, we have revised the Discussion (p16) so that the sentence “For the purposes of the analyses it was assumed that placebo treatment would be equally effective” now reads
“In order to produce a connected network for the purposes of the analyses it was assumed that the placebo treatments would be equally effective when administered to the same populations.”
In my view, in order to be acceptable for publication, the following changes are mandatory.

(a) Abstract - Results:
- state that severe baseline differences in severity exist between BTX and MBG studies

(b) Strength and limitations of this study:
- network meta-analyses are only based on BTX studies and results are extrapolated to MBG studies
- magnitude of treatment difference between BTX and MBG is questionable due to substantial baseline differences in severity

In line with reviewer 3’s comments we have revised sections of the manuscript, as requested, and hope these amendments will address the reviewer’s concerns appropriately.

(a) Abstract: We have inserted a sentence into the Results to highlight that the two patient populations differed in terms of the underlying severity of their symptoms.

(b) Strength and limitations of this study: We have added a further bullet to the list (max 5 points allowed) to provide further explanation of the limitations of the analysis.

|bullet| Network meta-regression was applied to the changes from baseline in urinary incontinence episodes and urgency episodes, in order to adjust for differences in the baseline values of these outcomes. Because individual patient data were available for onabotulinumtoxinA trials, but not for mirabegron trials, it was assumed that the relationship between baseline values and change from baseline was the same across all included trials. |

In summary, I do not think about network meta-analysis that “... in the absence of a head-to-head comparison this is the highest level of evidence that can be generated based on the available data.” There are the alternatives to not pool the data or to widen the network by including additional treatments for overactive bladder. If the authors do not choose any of these alternatives, the limitations of the conducted analyses must be clearly stated (in Abstract and list of strengths and limitations).

Since we did take the route of pooling the data, we have followed reviewer 3’s suggestion to clearly state the limitations of the analysis in the results of the abstract and the list of strengths and limitations (see response to comment 3).

---

**VERSION 3 - REVIEW**

| REVIEWER | Guido Schwarzer  
Institute for Medical Biometry and Statistics, Freiburg  
Medical Center - University of Freiburg  
Germany |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>REVIEW RETURNED</td>
<td>25-Jan-2016</td>
</tr>
</tbody>
</table>

**GENERAL COMMENTS** The authors addressed my remaining concerns in the revision.
Comparative assessment of onabotulinumtoxinA and mirabegron for overactive bladder: an indirect treatment comparison

Nick Freemantle, David A Ginsberg, Rachael McCool, Kelly Fleetwood, Mick Arber, Kristin Khalaf, Clara Loveman, Quanhong Ni and Julie Glanville

BMJ Open 2016 6:
doi: 10.1136/bmjopen-2015-009122

Updated information and services can be found at:
http://bmjopen.bmj.com/content/6/2/e009122

These include:

References
This article cites 35 articles, 1 of which you can access for free at:
http://bmjopen.bmj.com/content/6/2/e009122#BIBL

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Health economics (370)
Urology (77)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/