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.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

ARTICLE DETAILS

TITLE (PROVISIONAL)	To what extent are surgery and invasive procedures effective beyond a placebo response? A systematic review with meta-analysis of randomized, sham controlled trials
AUTHORS	Jonas, Wayne; Crawford, Cindy; Colloca, Luana; Kaptchuk, Ted; Moseley, Bruce; Miller, Franklin; Kriston, Levente; Linde, Klaus; Meissner, Karin

VERSION 1 - REVIEW

REVIEWER	**Jacobs, Wilco** Leiden University Medical Center, Neurosurgery
REVIEW RETURNED	19-May-2015

GENERAL COMMENTS	This is a well-perforemd systematic review, with clear methodology and reporting. The subject is worthy of investigation and could help future trialists in deciding on study designs, but also help readers on the value of clinical trials with or without sham. There are however some issues with the design and interpretation that need to be considered. Abstract/summary/key messages/What this paper adds - The conclusion of the abstract is biased against surgical interventions having only sham responses and are too generalized. Overall design of study - I am unsure if the design can appropriately answer the second question (estimate the specific effects of over sham procedures). I would expect a meta-regression, where sham is a factor ad where there is a variety of studies using sham or not. Including only sham studies disregards the possibility that the interventions subjected to sham controlled studies are biased towards having no effect. In which case the sham intervention was appropriatey used. Participants studied - Most of the studies concern low back pain. This is a very heterogeneous group, where it is known that in presence of a specific cause conditions are more appropriate for surgery (disc herniation for short term effect, for example), while other conditions are not (inappropriately called non-specific LBP). Methods - Methods generally comply with PRISMA guideline. - Was the protocol registered? For example in Prospero?

	- The authors mention using random effects model because the studies are expected to be heterogeneous. This is incorrect, if studies are truly heterogeneous, pooling is out of order. This relates to the earlier mentioned more fundamental problem, what about the effect sizes of sham versus surgical procedures in cases where the surgery is not effective? - I would think that blinding (or even the success of blinding) is an essential part of the risk of bias assessment, and worthy of a sensitivity analysis. This goes beyond the allocation concealment. Results - The results are well presented and complete. - Aim (3) is not answered by the data. Interpretation and conclusions - The conclusions are too far fetched and too generalized, the included selection of surgical trials involve definitely a selection of surgical procedures that are less invasive. Moreover, the overall effect as well as most of the subgroups are statistically significant, indicating an effect of the surgery. Then why on page 15, line 51 the authors conclude that also obesity lacks evidence of effect over sham, while the effect is 0.52? - How do the sham responses compare to placebo responses in non-surgical trials? While placebo effects on objective outcomes in non-surgcal trials are generally absent or small, they exist for subjective, patient reported outcomes (Hrjobartson 2010). - The authors too easily discard surgical interventions as a field where good quality research is difficult if not impossible. This disregards the problems in other fields where bias is just as present, in similar (placebo response), or in other forms (conflict of interest, publication bias). It also disregards the need to assess evidence in the best possible way, with appropriate methods given the circumstances (Black, 1996). - I would propose to turn the reasoning around: in many cases where sham surgery was used as a control intervention, it was well applied as the intervention indeed was questionnable. As an example, the vertebroplasty trials, where there was indeed no effect of the surgery over sham treatment caused the procedure to be abandoned, at least in some countries. This is supported by the effect in more conventional surgical areas where the authors find a relatively small effect of sham on the total effect. References - I do not have obvious related additional references to be cited. - p6, L10, ref 14 refers to Migraine review, and does not deal with vertebroplasty, moreover the text discusses two trials, while only one ref is given. I would expect the trials of Kallmes and Buchbinder (11, 12) that are referenced earlier on a topic on anesthesia. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;312:1215-8. Hróbjartsson A, Gøtzsche PC; Gøtzsche (January 2010). Hróbjartsson, Asbjørn, ed. "Placebo interventions for all clinical conditions". Cochrane Database of Systematic Reviews (1): CD003974.

REVIEWER	**Benedetti, Fabrizio**

	University of Turin Medical School, Public Health
REVIEW RETURNED	24-May-2015

GENERAL COMMENTS	In this study the authors perform a systematic review/meta-analysis aimed at assessing the efficacy of different surgical procedures compared to sham procedures. In general, this study is interesting, timely and potentially important. In fact, the authors found that several invasive procedures are not justified, particularly for pain and obesity, for sham procedures may lead to similar therapeutic outcomes. There are however some points that, if addressed, I believe may improve the paper. Blinding is crucial in placebo surgery. To my knowledge, there are only a few procedures that can be defined placebos. Indeed, to devise a placebo invasive procedure is very difficult and the authors should address this point more in depth. This a major flaw in many meta-analyses of this kind, but particularly in this meta-analysis on sham vs verum surgery. Too many “Unclear” are present in Appendix 1, and this raises too many questions and doubts. To calculate SD (when not reported) from pre SD and post SD using r=0.5 for the product-moment correlation between pre and post measures may lead to wrong interpretations. This kind of approach is useful to include as many trials as possible, but I wonder if it is correct within this context. The authors should specify how many trials and/or outcome measures of this kind were included in their study. Patients and conditions should be better stratified. The diagnoses of low back pain, abdominal pain, migraine are very general. Moreover, obesity can be due to many factors, and at least the type of obesity should be specified. I would suggest that the authors include a more detailed diagnosis of all these conditions. Overall, I think this paper will have an important impact on both the scientific and medical community.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Recommendation:

Comments:
This is a well-perforemd systematic review, with clear methodology and reporting. The subject is worthy of investigation and could help future trialists in deciding on study designs, but also help readers on the value of clinical trials with or without sham. There are however some issues with the design and interpretation that need to be considered.

Abstract/summary/key messages/What this paper adds
- The conclusion of the abstract is biased against surgical interventions having only sham responses and are too generalized.
We have adopted the conclusion accordingly: “The non-specific effects of surgical and invasive procedures are generally large. Particularly in the field of pain-related conditions, more evidence from randomized placebo-controlled trials is needed to avoid continuation of ineffective treatments. Certainly, no conclusions can be drawn for surgical interventions that have not been tested in placebo-controlled clinical trials”.

Overall design of study
- I am unsure if the design can appropriately answer the second question (estimate the specific effects of over sham procedures). I would expect a meta-regression, where sham is a factor ad where there is a variety of studies using sham or not. Including only sham studies disregards the possibility that the interventions subjected to sham controlled studies are biased towards having no effect. In which case the sham intervention was appropriatey used.
To collect all available RCTs for surgical interventions is an interesting approach for a follow-up study. In order to quantify the specific effects of surgery above sham treatment, however, we think that our approach is more straightforward. Certainly, any conclusions from our review are restricted to the interventions under investigation and cannot be generalized to all surgical interventions. We had addressed this issue in our discussion section: "Certainly, not all invasive procedures warrant sham-controlled comparisons; for example, when results demonstrate indisputable changes in objective parameters the risks of sham procedures would be excessive."

Participants studied
- Most of the studies concern low back pain. This is a very heterogeneous group, where it is known that in presence of a specific cause conditions are more appropriate for surgery (disc herniation for short term effect, for example), while other conditions are not (inappropriately called non-specific LBP).
All placebo-controlled studies on back pain in our dataset were performed in patients with chronic pain conditions. Therefore, the available data do not allow conclusions regarding acute back pain due to disc herniation (see our comment above). We have also complemented Appendix 1 with more detailed information on the diagnoses.

Methods
- Methods generally comply with PRISMA guideline.
- Was the protocol registered? For example in Prospero?
The study protocol was approved by all members of our study group. The protocol has not been registered.

- The authors mention using random effects model because the studies are expected to be heterogeneous. This is incorrect, if studies are truly heterogeneous, pooling is out of order. This relates to the earlier mentioned more fundamental problem, what about the effect sizes of sham versus surgical procedures in cases where the surgery is not effective?
According to the Cochrane Handbook, a random-effects meta-analysis may be used to incorporate heterogeneity among studies. We did not identify substantial heterogeneity among results of studies, especially when one outlier was excluded. We added sensitivity analyses with and without this outlier to the results section. Moderate heterogeneity is very usual in such types of meta-analyses (compare, for example, Hróbjartsson, Ashbjørn, and Peter C. Gøtzsche. "Placebo interventions for all clinical conditions." Cochrane Database Syst Rev 1.1 (2010)).

- I would think that blinding (or even the success of blinding) is an essential part of the risk of bias assessment, and worthy of a sensitivity analysis. This goes beyond the allocation concealment.
We have now included a sensitivity analysis for risk of bias due to unblinding: The results were essentially the same (see Figure 2).

Results
- The results are well presented and complete.
- Aim (3) is not answered by the data.
*Study aim (3) was to estimate the contribution of the **surgical ritual and other nonspecific factors** to outcomes from these procedures. We would like to emphasize that we explicitly estimated the contribution of these nonspecific effects to the total treatment effects (see page 13 and Figure 4 of our manuscript): "Thus, on average, the changes in the sham groups accounted for 65% of the overall improvement from the treatments. This proportion of specific to non-specific treatment effects was larger in pain-related conditions (78%) and obesity (71%) than in GERD (57%) and other conditions (57%), and was considerably smaller in classical surgery trials (21%) than in endoscopic trials (73%) and those using percutaneous procedures (64%). (Figure 4) Changes in the sham groups accounted for 89% and 82% of overall improvement in intermediate and late pain outcomes."*

Interpretation and conclusions

- The conclusions are too far fetched and too generalized, the included selection of surgical trials involve definitely a selection of surgical procedures that are less invasive. Moreover, the overall effect as well as most of the subgroups are statistically significant, indicating an effect of the surgery. Then why on page 15, line 51 the authors conclude that also obesity lacks evidence of effect over sham, while the effect is 0.52?
As mentioned above, we certainly can only draw conclusions for surgical interventions that have been tested in placebo-controlled clinical trials. We also rephrased parts of the discussion section in order to avoid overgeneralization. Furthermore, we changed our conclusions regarding the studies on obesity on page 16: "There is evidence to support these interventions for GERD, and limited evidence to support the use of balloon insertion for obesity."

- How do the sham responses compare to placebo responses in non-surgical trials? While placebo effects on objective outcomes in non-surgcal trials are generally absent or small, they exist for subjective, patient reported outcomes (Hrjobartson 2010).
There is a major difference between our study and the meta-analysis by Hrobjartsson and Gotzsche: while these authors estimated the effect of placebo treatment versus no treatment, we estimated the specific effect of surgery versus placebo treatment. Please note that only one of the studies in our dataset included a no-treatment group.
Part of our data, however, can be compared to results of Krogsboll et al., who compared the changes from baseline in the placebo groups and no treatment groups to changes in the verum groups. Their subgroup analyses for studies on 'chronic pain' and 'obesity' showed smaller percentages of nonspecific effects to the active treatment effects than we found in the surgical trials (chronic pain, 40% vs. 78%; obesity, 33% vs. 71%). We complemented our discussion section accordingly (page 14): "While nonspecific effects accounted for approximately 65% of the effects from all invasive procedures, they made up to 78% of the active treatment effects in chronic pain conditions and 71% of the active treatment effects in obesity. These percentages are substantially higher than have been observed in non-surgical trials, namely 40% for chronic pain conditions and 33% for obesity. The higher contribution of non-specific effects in surgical trials could well be the result of higher placebo effects, although the lack of no-treatment groups in our dataset allows no firm conclusion."*

**Krogsbøll, Lasse T., Asbjørn Hróbjartsson, and Peter C. Gøtzsche. "Spontaneous improvement in randomised clinical trials: meta-analysis of three-armed trials comparing no treatment, placebo and active intervention." BMC medical research methodology 9.1 (2009)*

- The authors too easily discard surgical interventions as a field where good quality research is difficult if not impossible. This disregards the problems in other fields where bias is just as present, in similar (placebo response), or in other forms (conflict of interest, publication bias). It also disregards the need to assess evidence in the best possible way, with appropriate methods given the circumstances (Black, 1996).
Please see our comments above. We have altered parts of the discussion section and now avoid overgeneralization of the results. However, we think that a review on the nonspecific effects of surgery is timely and highly warranted, even more when looking at our results for pain-related conditions. In endometriosis-associated pain, for example, surgical procedures are still frequently performed even though there is only one placebo-controlled study, which showed no significant effect of surgery above sham surgery (see Fig. 2). We are confident that our results, in addition to the results of a former meta-analysis on this issue published in the BMJ (Wartolowska et al., 2014), will open the way for the development of more efficient treatment approaches for some conditions in the field of chronic pain instead of discontinuing ineffective invasive treatments.

- I would propose to turn the reasoning around: in many cases where sham surgery was used as a control intervention, it was well applied as the intervention indeed was questionable. As an example, the vertebroplasty trials, where there was indeed no effect of the surgery over sham treatment caused the procedure to be abandoned, at least in some countries. This is supported by the effect in more conventional surgical areas where the authors find a relatively small effect of sham on the total effect.
See our comments above. We do not think that all surgical procedures are ineffective. Rather, we are confident that our study may help to reduce the application of ineffective surgical interventions, such as in certain chronic pain conditions.

References
- I do not have obvious related additional references to be cited.

- p6, L10, ref 14 refers to Migraine review, and does not deal with vertebroplasty, moreover the text discusses two trials, while only one ref is given. I would expect the trials of Kallmes and Buchbinder (11, 12) that are referenced earlier on a topic on anesthesia.
We have corrected the respective references.

Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;312:1215-8.
Hróbjartsson A, Gøtzsche PC; Gøtzsche (January 2010). Hróbjartsson, Asbjørn, ed. "Placebo interventions for all clinical conditions". Cochrane Database of Systematic Reviews (1): CD003974.

Additional Questions:
Please enter your name: Wilco Jacobs

Job Title: Senior scientist, Epidemiologist

Institution: Leiden University Medical Center; The Health Scientist

Reviewer: 2

Recommendation:

Comments:
In this study the authors perform a systematic review/meta-analysis aimed at assessing the efficacy of different surgical procedures compared to sham procedures. In general, this study is interesting, timely and potentially important. In fact, the authors found that several invasive procedures are not justified, particularly for pain and obesity, for sham procedures may lead to similar therapeutic outcomes. There are however some points that, if addressed, I believe may improve the paper.

Blinding is crucial in placebo surgery. To my knowledge, there are only a few procedures that can be defined placebos. Indeed, to devise a placebo invasive procedure is very difficult and the authors should address this point more in depth. This a major flaw in many meta-analyses of this kind, but particularly in this meta-analysis on sham vs verum surgery. Too many “Unclear” are present in Appendix 1, and this raises too many questions and doubts.
In the majority of our studies(87%) we judged the risk of bias due to unblinding as low (page 12). Furthermore, we have added a sensitivity analysis for this risk of bias in our results section, which showed a comparable effect size to the main analysis (Figure 2). In contrast, the risk of bias regarding allocation sequence concealment (listed in Appendix 1) was “unclear” in a lot of studies. However, excluding studies with unclear or high risk of bias regarding allocation concealment did not reveal different results (Figure 2).

To calculate SD (when not reported) from pre SD and post SD using r=0.5 for the product-moment correlation between pre and post measures may lead to wrong interpretations. This kind of approach is useful to include as many trials as possible, but I wonder if it is correct within this context. The authors should specify how many trials and/or outcome measures of this kind were included in their study.
The SD was calculated from pre SD and post SD using r=0.5 in 17 out of 39 studies (44%). We therefore had performed sensitivity analyses with r=0.3 and r=0.7 (Figure 2). Since these analyses did not reveal different results (Figure 2), we are confident that our conclusions are not biased due to this approach. We have now added the number of studies for which the SD was calculated to our manuscript.

Patients and conditions should be better stratified. The diagnoses of low back pain, abdominal pain, migraine are very general. Moreover, obesity can be due to many factors, and at least the type of obesity should be specified. I would suggest that the authors include a more detailed diagnosis of all these conditions.
We have now included more detailed diagnoses of these conditions in Appendix 1.

Overall, I think this paper will have an important impact on both the scientific and medical community.

Additional Questions:

Please enter your name: Fabrizio Benedetti

VERSION 2 – REVIEW

REVIEWER	Valerie Durkalski-Mauldin Medical University of South Carolina, USA
REVIEW RETURNED	28-Sep-2015

GENERAL COMMENTS	The authors present an interesting meta-analysis of sham-controlled studies in a variety of disease areas with the goal of quantifying the sham-controlled trials of invasive procedures and estimating the treatment effects. This is certainly a topic that needs more attention in the literature, however, I do have a few comments for the authors to consider. 1. Page 7, line 5: You mention on the previous page (Pg6, line38) the limitations with trials comparing invasive procedures to drug therapy without a sham arm. You then state on page 7 that you will be looking at studies that use a sham procedure. Is there data available so you can comment in the discussion about your findings as compared to those trials that only used a drug arm as the comparator (as mentioned in reference 16)? It would be interesting to know the treatment effects seen in those studies compared to what you found. 2. Page 9, first paragraph on QA: Can the authors provide a bit more rationale for why they used a publication if the miOM was defined by method 3 (clearly most relevant outcome determined by the reviewers)? If a manuscript did not clearly state the primary outcome, then maybe it was not a well-designed trial to examine. How often was method 3 implemented? 3. Page 10, first sentence: Why did the authors decide to exclude manuscripts with dichotomous endpoints? It seems that several trials use a binary outcome for success/failure in these settings. I could not derive from the appendix how many papers were excluded due to this requirement. Was there a sufficient amount to do a separate meta-analysis? 4. Page 11 first paragraph: How did the authors define 'low risk of allocation concealment'? Similarly, on page 12, how was 'adequate blinding of patients and outcome assessors' defined? 5.Figure 2; The statistical heterogeneity is relatively high, even when using random effects and subgroup analyses. Can the authors expand on this heterogeneity in the strength and weakness section of the manuscript? This is worth highlighting since the audience will be highly variable and may not fully understand what these means when interpreting your results. 6. Page 14, line 45: The authors state they found 'positive through modest overall effect' How are the authors defining modest? According to interpretation of Cohen's d or according to clinical relevance in the surgical/invasive procedure setting? 7. The authors should include how many publications of the 55 had a sample size >=100. The total sample size across these studies was reported in a table but not the number of studies and is difficult

	to pull out from the appendix table. 8.Page 15, line 45: How did the authors know if the studies were successfully blinded? 9. Page 15, line 52: Although it is a general sense that negative papers do not get published as much as positive papers, do the authors know if that is true for the surgical/invasive trials? 10. Page 17, line 29: I think the authors should clarify that 'it is noteworthy that only 55 sham-controlled studies that met our criteria for the study, have been conducted to date'. There most likely are other sham-controlled studies that you did not capture due to your criteria for inclusion. 11. The 'Grey literature' is described in the methods but it is not clear how it was used in the interpretation of the results. Minor: Page 6, line 42: correct the wording for, 'Comparative trials do not allow to estimate the specific.....' Page 7, line 3: Rather than 'To address these issues', a more appropriate term may be, 'To better understand these issues...' Page 11, line 39: Rather than stating 'the majority (25) of studies', it should read about half of the studies. Page 14, line 47: the symbol should be greater than or equal to 100.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name

Valerie Durkalski-Mauldin

Institution and Country

Medical University of South Carolina, USA

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

Please leave your comments for the authors below The authors present an interesting meta-analysis of sham-controlled studies in a variety of disease areas with the goal of quantifying the sham-controlled trials of invasive procedures and estimating the treatment effects. This is certainly a topic that needs more attention in the literature, however, I do have a few comments for the authors to consider.

1. Page 7, line 5: You mention on the previous page (Pg6, line38) the limitations with trials comparing invasive procedures to drug therapy without a sham arm. You then state on page 7 that you will be looking at studies that use a sham procedure. Is there data available so you can comment in the discussion about your findings as compared to those trials that only used a drug arm as the comparator (as mentioned in reference 16)? It would be interesting to know the treatment effects seen in those studies compared to what you found.

There is one recent systematic review summarizing results from 27 comparative surgical trials (Papanastassiou et al., 2012*). The authors concluded that vertebroplasty and kyphoplasty are more effective in reducing pain from vertrebral compression fractures than do non-surgical treatments. We now include this information in the introduction:

“Two more recent controlled studies of vertebroplasty for painful osteoporotic vertebral fractures reported similar degrees of pain relief from sham procedures involving only superficial anaesthesia compared to the more invasive active procedures.12 ,13 In contrast, a systematic review comparing surgical with non-surgical treatments for painful osteoporotic vertebral fractures came to the conclusion that vertebroplasty and kyphoplasty are superior to non-surgical treatments.14 Since invasive interventions frequently go along with larger non-specific effects than non-invasive treatments 15,16 surgical trials that do not include a sham surgery arm may give biased results. Thus, the efficacy of invasive procedures, for example for chronic pain conditions, remains controversial. 14,17”

*Papanastassiou ID, Phillips M, Van Meirhaeghe J, Berenson JR, Andersson GB, Chung G, Small BJ, Aghayev K, Vrionis FD. Comparing effects of kyphoplasty, vertebroplasty, and non-surgical management in a systematic review of randomized and non-randomized controlled studies. European Spine Journal 2012; 1826-1843.

2. Page 9, first paragraph on QA: Can the authors provide a bit more rationale for why they used a publication if the miOM was defined by method 3 (clearly most relevant outcome determined by the reviewers)? If a manuscript did not clearly state the primary outcome, then maybe it was not a well-designed trial to examine. How often was method 3 implemented?

Eighteen studies implemented method 3. There were very few disagreements among reviewers regarding the selection of the most important outcome. For example, weight loss was selected as miOM in a study on obesity.

We also would like to emphasize that we ran a sensitivity analysis including only studies with clearly defined POM; this analysis provided comparable results to our main meta-analysis (SMD, 0.29 [0.07; 0.50] vs. 0.34 [0.20; 0.49]; see Figure 2).

3. Page 10, first sentence: Why did the authors decide to exclude manuscripts with dichotomous endpoints? It seems that several trials use a binary outcome for success/failure in these settings. I could not derive from the appendix how many papers were excluded due to this requirement. Was there a sufficient amount to do a separate meta-analysis?

Twelve trials were excluded from the primary meta-analysis because they provided dichotomous outcome(s) only. Our decision to exclude trials without binary outcomes was based on our analysis plan and accounted for the finding that most outcomes in our dataset were continuous and that primary outcomes were defined more frequently for continuous outcomes than for dichotomous outcomes (13 versus 5).

We now present an additional sensitivity analysis based on these 12 dichotomous outcomes in the manuscript (see Supplementary Figure 2) and complemented the methods, results and discussion sections accordingly:

Additions to the methods section:
“According to our analysis plan, the meta-analyses focused on continuous outcomes.”

“Trials reporting only a dichotomous outcome measure (responder data) are noted in Supplementary Table 1, and a sensitivity analysis was computed for these outcomes (see below).”

“An additional sensitivity analysis was performed for dichotomous outcomes of 12 studies that provided no continuous outcome (Supplementary Figure 1).”

Additions to the results section:
“Twelve studies provided only a dichotomous outcome measure (Supplementary Table 1). Sensitivity analyses showed an overall effect of surgery compared to sham surgery (risk ratio 1.54, 95% CI, 1.11 to 2.15; p=0.01), while heterogeneity was large (I2=59%, p=0.005). Subgroup analyses according to condition revealed a significant effect of surgery vs. sham surgery for pain studies (n=9; risk ratio 1.60, 95% CI, 1.11 to 2.30; p=0.01; I2=59%, p=0.01) but not for other studies (n=3; risk ratio 2.19, 95% CI, 0.44 to 10.84; p=0.33; I2=60%, p=0.08) (Supplementary Figure 1).”

Additions to the discussion section:
“However, please note that the analysis of dichotomous outcomes showed a somewhat larger specific effect for pain studies (Supplementary Figure 1).”

4. Page 11 first paragraph: How did the authors define 'low risk of allocation concealment'? Similarly, on page 12, how was 'adequate blinding of patients and outcome assessors' defined?
As part of the quality assessment of the individual studies ‘low risk of allocation concealment’ was defined according to the Cochrane Collaboration’s tool for assessing risk of bias. The same was true for ‘adequate blinding of patients and outcome assessors’. We already refer to this tool in the methods section:

“The methodological quality of the individual studies (sequence generation, allocation concealment, was assessed independently by three reviewers using the Cochrane Risk of Bias (ROB) tool.18”

5.Figure 2; The statistical heterogeneity is relatively high, even when using random effects and subgroup analyses. Can the authors expand on this heterogeneity in the strength and weakness section of the manuscript? This is worth highlighting since the audience will be highly variable and may not fully understand what these means when interpreting your results.

We now refer to heterogeneity in the strengths and weakness section:

“This is also indicated by the moderate-to-large heterogeneity in our meta-analyses, indicating more variation of effect sizes than would be expected by chance.”

6. Page 14, line 45: The authors state they found 'positive through modest overall effect' How are the authors defining modest? According to interpretation of Cohen's d or according to clinical relevance in the surgical/invasive procedure setting?

We refer to Cohen's d here. We have complemented the respective sentence:

“…, we found a positive though modest overall effect size (Cohen’s d) from the invasive procedures included in the analysis”

7. The authors should include how many publications of the 55 had a sample size >=100. The total sample size across these studies was reported in a table but not the number of studies and is difficult

to pull out from the appendix table.

Ten studies had a sample size >=100. We now included this information in the results section.

8.Page 15, line 45: How did the authors know if the studies were successfully blinded?

We used the Cochrane Collaboration's tool for assessing risk of bias (including blinding; see our response to comment 4).

9. Page 15, line 52: Although it is a general sense that negative papers do not get published as much as positive papers, do the authors know if that is true for the surgical/invasive trials?

Our funnel plot was asymmetric, which might be due to the lack of studies with negative outcomes. We have included this as a possible reason for funnel plot asymmetry:

"Inspection of the funnel plot suggests the presence of biases, such as small studies bias or publication bias (Figure 3)"

10. Page 17, line 29: I think the authors should clarify that 'it is noteworthy that only 55 sham-controlled studies that met our criteria for the study, have been conducted to date'. There most likely are other sham-controlled studies that you did not capture due to your criteria for inclusion.

We changed the wording and now state that "it is noteworthy that we could identify only 55 sham-controlled studies in the literature."

11. The 'Grey literature' is described in the methods but it is not clear how it was used in the interpretation of the results.

We edited the text on page 15 to clarify this issue:

"We also conducted a thorough search of the grey literature, as described above, and had input by experts in placebo research, increasing the likelihood of capturing all studies in this area. This activity allowed for a cross-check in the end to ensure we captured most of the relevant published randomized controlled trials for this review. We did not find any unpublished reports that met our inclusion criteria appropriate for this review, however there were some publications that were not readily accessible through the search engines commonly accessed that we were able to capture through these methods."

Minor:
Page 6, line 42: correct the wording for, 'Comparative trials do not allow to estimate the specific.....'

We corrected the wording (see response to comment 1).

Page 7, line 3: Rather than 'To address these issues', a more appropriate term may be, 'To better understand these issues...'

We have corrected the wording as suggested.

Page 11, line 39: Rather than stating 'the majority (25) of studies', it should read about half of the studies.

We have corrected the wording as suggested.

Page 14, line 47: the symbol should be greater than or equal to 100.

Thank you, we have corrected this error.

VERSION 3 – REVIEW

REVIEWER	Valerie Durkalski-Mauldin Medical University of South Carolina, USA
REVIEW RETURNED	05-Nov-2015

GENERAL COMMENTS	The authors addressed all of my comments and I appreciate the additional information and edits they provided in this resubmission.