A meta-analysis of temporal changes of response in the placebo arm of surgical randomized controlled trials: an update

Karolina A. Wartolowska¹,²*, Stephen Gerry²,⁴, Benjamin G. Feakins³, Gary S. Collins²,⁴, Jonathan Cook¹,², Andrew Judge¹,²,⁵ and Andrew J. Carr¹,²

Abstract

Background: Temporal changes in the placebo arm of randomized controlled trials (RCTs) have not been thoroughly investigated, despite the fact that results of RCTs depend on the comparison between arms.

Methods: In this update of our earlier systematic review and meta-analysis, we set out to investigate the effect of assessment time and number of visits on the magnitude of change from baseline in the placebo arm of these trials. We used linear mixed-effects models to account for within-trial correlations.

Results: Across all 47 trials the magnitude of response in the placebo arm did not change with time (β = -0.0070, 95% CI -0.024, 0.010) or visit (β = -0.033, 95% CI -0.082, 0.017) and remained significantly different from baseline for at least 12 months or seven follow-up visits. Change in the placebo arm in trials with subjective outcomes was large (β₀ = 0.68, 95% CI 0.53, 0.82) and relatively constant across time (β = -0.0042, 95% CI -0.024, 0.016) and visit (β = -0.029, 95% CI -0.089, 0.031), whereas in trials with objective outcomes the response was smaller (β₀ = 0.28, 95% CI 0.11, 0.46) and diminished with time (β = -0.030, 95% CI -0.050, -0.010), but not with visit (β = -0.099, 95% CI -0.30, 0.11). For trials with assessed outcomes, there was no significant effect of time (β = -0.0071, 95% CI -0.026, 0.011) or visit (β = -0.032, 95% CI -0.33, 0.26); however, these results should be interpreted with caution due to the small number of studies, and high clinical heterogeneity between studies. In trials with pain as an outcome, the improvement was significant (β₀ = 0.91, 95% CI 0.75, 1.07), but there was no effect of time (β = -0.013, 95% CI -0.06, 0.03) or visit (β = -0.045, 95% CI -0.16, 0.069), and pain ratings remained significantly different from baseline for 12 months or seven visits.

Conclusions: These results are consistent with our previous findings. In trials with subjective outcomes response in the placebo arm remains large and relatively constant for at least a year, which is interesting considering that this is an effect of a single application of an invasive procedure. The lack of effect of time and visit number on subjective outcomes raises further questions regarding whether the observed response is the result of placebo effect or the result of bias.

Keywords: Surgery, Placebos, Randomized controlled trials, Systematic review, Meta-analysis
Background

There is conflicting evidence regarding the change in magnitude of placebo response in randomized controlled trials (RCTs). Several studies have demonstrated that values in the placebo arm remain different from baseline for the duration of the follow up [1–3]. However, there is no consensus as to whether placebo response does [4–6] or does not [4, 7] significantly change with time, and as to whether placebo response has a time-effect curve with a peak and carry-over effect [2, 8]. Experimental studies, specifically designed to investigate temporal changes after placebo administration, typically use a very short observation time, e.g., under an hour [9, 10], so their results provide little insight into what happens in clinical trials, in which the follow up lasts weeks or months.

In the previous paper [3], we used data from our earlier systematic review of surgical RCTs with a placebo arm [11] to determine the effect of time on the effect size of the primary outcome at the primary assessment time point. [3] We observed that the effect in the placebo arm remained significantly different from baseline throughout blinded follow up, but that the timing of the assessment did not appear to affect the magnitude of effect.

Typically, meta-analyses focus on the effect size at the primary assessment time point and treat it as a single outcome measure of treatment efficacy; however, trials often report outcomes at multiple follow-up visits. In this paper, we have assessed the temporal characteristics of changes in the placebo arm using a meta-analysis of within-and between-study changes across all outcome time points for all analyzed surgical RCTs. Our aim was to verify the findings from our earlier paper by incorporating the data from all reported follow-up visits, accounting for within-study and between-study correlations, and to investigate the effect of the outcome type and number of visits on temporal changes in the placebo arm of surgical RCTs.

Methods

Search strategy, eligibility and data extraction

The details of the search strategy and data extraction have been published previously [11]. In brief, we searched the databases of MEDLINE, EMBASE, CENTRAL, and ClinicalTrials.gov for placebo-controlled RCTs investigating the efficacy of a surgical intervention. The databases were searched from their inception to October 2015.

Surgical intervention was defined as “any interventional procedure that changed the anatomy and required a skin incision or the use of endoscopic techniques. Dental studies and invasive procedures used to deliver a pharmacological substance or stem cells, or that aimed to alleviate symptoms by modulation, stimulation or denervation were excluded”. Placebo was defined as “sham surgery, or an imitation procedure intended to mimic the active intervention; including the scenario where a scope was inserted and nothing was done, but patients were sedated or under general anesthesia and could not distinguish whether or not they underwent the actual surgery” [11].

We only included trials that reported continuous data and presented data in a form that made it possible to calculate an effect size. From each eligible study, we extracted the lead author’s surname, the publication year, the number and timing of each treatment and follow-up visit, the primary outcome, and the mean and standard deviation (SD) of the primary outcome in the placebo arm of each trial. Outcomes were classified as “subjective”, i.e., patient-reported and depending on the patients’ perceptions and cooperation, “assessed”, i.e., subjective ratings judged by external assessors or “objective”, i.e., measured using devices or laboratory tests and independent of patients’ or observers’ perceptions, for example, weight. As most of the literature on placebo effect is related to pain studies, we also extracted data on pain outcomes, irrespective of whether these were primary or secondary outcomes, for a secondary analysis.

Study effect sizes were quantified as within-arm standardized mean differences (SMDs) between baseline and follow-up values using the Cohen’s d method [12]. A pretest-posttest correlation coefficient (r) of 0.5 was used to calculate the standard error of each SMD [3]. Where necessary, SMDs were sign-flipped to ensure that improvement was consistently presented as a reduction in SMD.

Model for analysis of multiple time points

In order to analyze the trajectory of change in the placebo arm, we adopted a similar method to that proposed by Ishak et al. [13], who conducted a meta-analysis of longitudinal studies reporting data at a series of fixed time points, and then used linear mixed-effects models to account for correlations, both within and between studies. We adapted Ishak’s method to conduct a meta-analysis of studies that report data at many different and irregularly spaced time points, while still accounting for correlations between data points within and between studies. All standardized data points from each individual study were included. We used a random coefficients model (also known as a random intercept and slope model) with continuous time, which allows for an arithmetic description of the relationship between the measurement of interest and time from baseline [14]. The models were based on polynomials of time by adding higher-order polynomials, first as fixed effects, then as random coefficients. Decisions on which terms to include in the final model were based on Akaike information.
were eligible for inclusion. The median duration of

Results
Study selection and characteristics
Information on the search strategy, Preferred Reporting
Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram, and characteristics of the identi-
fied and included trials have been reported previously [3], with an abridged version presented in Additional
file 1. A total of 47 trials, involving 1744 participants,
were eligible for inclusion. The median duration of

Main findings
Across all analyzed surgical RCTs, the magnitude of pla-
cebo response did not diminish with time or visit, and
remained significantly different from baseline for the
duration of the trial (Fig. 1, Tables 1 and 2).

When trials were subdivided by outcome type, time
was associated with statistically significant decreases in
placebo response in trials with objective outcomes. For
trials with subjective outcomes, neither time nor visit
were significantly associated with change in placebo re-
response; however, the response remained large and sig-
ificantly different from baseline for at least 12 months.

In analyses limited to trials with pain as the outcome,
the effects of time and visit were not significant (Fig. 3,
Table 5 and 6).

Discussion
Main findings
To the best of our knowledge, this is the first meta-
analysis modelling longitudinal changes in the magnitude
of effect in the placebo arm of surgical RCTs across all
follow-up visits. This study supports our earlier findings
that, across all analyzed trials, the magnitude of placebo
response does not significantly change with time; it also
shows that placebo response does not change with the
number of visits. However, when trials were subdivided
by outcome type, for trials with subjective outcomes, the
response in the placebo arm remained large and relatively
constant for at least 12 months or seven visits, while for
trials with objective outcomes, the improvement in the
placebo arm diminished with time but not with follow-up
visits. It should be noted that, unlike pharmaceutical trials,
in which there may be multiple applications of a placebo
regimen, all but three trials analyzed in this study used a
“one-off” placebo intervention (Additional file 1).

Strengths and limitations
The main strength of this paper is that, unlike our previous
analysis, we investigated the effect of time and visit number
across all follow-up visits. The main limitation of this study
is the paucity of data; both in the form of the number of eli-
geble studies, and the number of observations per study.

We were not able to investigate the extent to which
placebo response was caused by the so called “true
placebo effect”, i.e., the difference in effect between the
placebo and non-interventional arms [16], because only
one trial included a non-intervention group [17]. Analyses were limited to the assessment of "placebo response", i.e., the total change in the placebo arm between the baseline and follow-up visits. As a consequence of this, the magnitude of response may be affected by factors that are not related to placebo intervention itself, such as regression to the mean, fluctuations in disease severity, spontaneous improvement or report bias.

Seventy-nine percent of the trials used or allowed the use of standard or rescue medication [3]; therefore, some of the effect observed in the placebo group may be the result of concomitant pharmacological treatment or lifestyle modifications. We did not include any information about concomitant medication in our analysis because of the limited reporting of this characteristic in the analyzed trials.

Analyses were based on summary data, as reported by the authors of each trial. This limited the power of our analyses and did not allow us to investigate the effect of patient-related factors; therefore, we were unable to draw inference about individual patient responses (significant relationships at the population level may not hold true at the individual level and vice versa, i.e., the ecological fallacy).

The statistical modelling carried out in this study relied on several choices and assumptions. We chose to use SMDs for the outcome measure, as the reported outcome types and measures were heterogeneous, rendering standardization necessary to combine results. In a study with baseline and multiple follow-up outcome assessments, we would typically adjust for the magnitude of the outcome at baseline, but this was not appropriate here. There is also the possibility that the standard error of each study estimate was not appropriately accounted for, since the baseline value was incorporated in the SMD and the weighting at each time point; however, removing the weighting from the model made little difference to the findings.

**Interpretation**
This study confirms our previous findings that placebo response does not seem to have a response-curve and that it persists for at least 12 months [3].

The "longevity" of placebo response has been reported previously [1, 2] and has been interpreted by some authors as regression to the mean [18–20]. In this study, we could not investigate whether placebo response was caused by regression to the mean because of the lack of information on the true population mean for the outcome measures. However, regression to the mean could explain many of the characteristics of placebo response, for example, that it tends to be larger for more extreme values and for more unreliable measures. [20].

**Table 1** The effect of time (in months) on placebo response across all trials

| Outcome type | Intercept $\beta_0$ (95% CI) | $P$ value | Coefficient $\beta$ (95% CI) | $P$ value |
|--------------|-------------------------------|-----------|-----------------------------|-----------|
| All          | 0.55 (0.42, 0.68)             | <0.0001   | -0.0070 (-0.024, 0.010)     | 0.40      |

**Table 2** The effect of visit on placebo response across all trials

| Outcome type | Intercept $\beta_0$ (95% CI) | $P$ value | Coefficient $\beta$ (95% CI) | $P$ value |
|--------------|-------------------------------|-----------|-----------------------------|-----------|
| All          | 0.56 (0.42, 0.71)             | <0.0001   | -0.033 (-0.082, 0.017)      | 0.16      |
Using longitudinal analysis showed that outcome type affects not only the magnitude of response in the placebo arm, but also its temporal changes; with the effect on objective outcomes decreasing with time, and the effect on subjective outcomes, including pain, not changing with time and remaining significantly different from baseline values. This is in line with the results of our original paper [3]. Findings in previously published studies, investigating temporal changes in the placebo arm have been inconsistent. Some studies reported that the effect of time is not significant [4] or that trial duration may explain some of the variation in improvement [21]. A meta-analysis of acupuncture trials [2], showed a peak in placebo response at 12 weeks with a subsequent drop at 52 weeks; however, the follow-up visits were divided into six groups according to the assessment timing, allowing a different number of trials to contribute to the effect in each group. Kaptchuk et al. [5] reported a continuous reduction in pain over 8 weeks in a sham acupuncture trial; however, they re-applied the treatment throughout

### Table 3

| Outcome type | Intercept $\beta_0$ (95% CI) | $P$ value | Coefficient $\beta$ (95% CI) | $P$ value |
|--------------|--------------------------------|-----------|------------------------------|-----------|
| Subjective   | 0.68 (0.53, 0.82)              | <0.0001   | -0.0042 (-0.024, 0.016)      | 0.67      |
| Assessed     | 0.37 (-0.070, 0.81)            | 0.086     | -0.0071 (-0.026, 0.011)      | 0.38      |
| Objective    | 0.28 (0.11, 0.46)              | 0.0037    | -0.030 (-0.050, -0.010)      | 0.0084    |

### Table 4

| Outcome type | Intercept $\beta_0$ (95% CI) | $P$ value | Coefficient $\beta$ (95% CI) | $P$ value |
|--------------|--------------------------------|-----------|------------------------------|-----------|
| Subjective   | 0.67 (0.50, 0.83)              | <0.0001   | -0.029 (-0.089, 0.031)       | 0.28      |
| Assessed     | 0.36 (-0.079, 0.80)            | 0.091     | -0.032 (-0.33, 0.26)         | 0.51      |
| Objective    | 0.32 (0.15, 0.50)              | 0.0024    | -0.099 (-0.30, 0.11)         | 0.10      |
the study, and it is known that administering treatment multiple times reinforces the placebo effect [22].

In our meta-analysis, visit number did not have a significant effect on the magnitude of placebo response. This contradicts the findings of Vase et al., who reported that a larger number of face-to-face visits was associated with a larger placebo response [23]. However, Vase et al. used individual patient data from several pharmacological trials, in which a placebo was administered multiple times during the trial, whereas we analyzed trial-level data from surgical trials, which tend to involve a “one-off” intervention.

Some of the improvement in the placebo arm might have been related to the effect of concomitant treatment. One plausible explanation for improvement is the use of rescue medications, such as pain-killers, or the introduction of lifestyle modifications, such as diet. Another possible explanation is that patients who were allowed to continue their standard pharmacological treatments throughout the trial might have improved due to better adherence to these treatments when in the trial.

Implications

There are two main implications of this study. First, for the researchers and clinicians designing placebo-controlled trials of invasive procedures using subjective outcomes, it is useful to know that the effect in the placebo arm does not seem to change significantly with time. This is important because in a placebo-controlled trial treatment effect is measured as the difference between the change in the active and placebo arms. Second, patients report significant improvement for a prolonged period of time after a single surgical placebo intervention, i.e., a procedure chosen to have no active therapeutic effect. Therefore, surgeons and therapists should be aware that large and sustained improvement in subjective outcomes after a procedure does not always mean that the crucial surgical element of the treatment is effective.

Conclusions

This study found that for surgical RCTs with subjective outcomes, the effect size in the placebo arm persists for at least a year, and does not significantly diminish with follow-up visits. Only in trials with objective outcomes did the magnitude of placebo response diminish with time, but not with number of visits. These findings further reinforce the value of controlling for potential bias, especially in trials with a subjective key outcome. Further research is needed to explore the extent to which the observed response in the placebo arm is explained by response bias and other potential sources of bias.

Table 5 The effect of time (in months) on placebo response in trials with pain as the outcome

| Outcome | Intercept (β0, 95% CI) | P value | Coefficient (β, 95% CI) | P value |
|---------|------------------------|---------|------------------------|---------|
| Pain    | 0.91 (0.75, 1.07)      | <0.0001 | -0.013 (-0.06, 0.03)    | 0.48    |

Table 6 The effect of visit on placebo response in trials with pain as the outcome

| Outcome | Intercept (β0, 95% CI) | P value | Coefficient (β, 95% CI) | P value |
|---------|------------------------|---------|------------------------|---------|
| Pain    | 0.94 (0.72, 1.15)      | <0.0001 | -0.045 (-0.16, 0.069)   | 0.36    |
Additional file

**Additional file 1**: Characteristics of surgical randomized controlled trials with a placebo arm included in this analysis. (PDF 47 kb)

**Funding**
This study was funded by the NIHR Oxford Musculoskeletal Biomedical Research Unit. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

**Availability of data and materials**
The datasets supporting the conclusions of this article are included within the article.

**Authors’ contributions**
KAW, SG, BGF, JC, GSC and AJC conceived and designed the study. KAW, SG, BGF, JC, GSC and AJC designed the statistical analysis. KAW, SG, BGF, GSC, JC and AJ analyzed the data. KAW drafted the manuscript and is the guarantor of the study. KAW, SG, BGF, GSC, JC, AJ and AJC contributed to the writing of the manuscript and critically revised the manuscript for important intellectual content. All authors agree with the manuscript results and conclusions. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
KAW, JC, AJ and AJC are involved in a placebo-controlled surgical trial on shoulder pain (NCT01623011).

**Publisher’s Note**
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**
1NIHR Oxford Musculoskeletal Biomedical Research Unit, Old Road, OX3 7LD Oxford, UK. 2Botnar Institute of Musculoskeletal Sciences, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Old Road, OX3 7LD Oxford, UK. 3Nuffield Department of Primary Care Health Sciences, University of Oxford, Woodstock Road, OX2 6GG Oxford, UK. 4Centre for Statistics in Medicine, University of Oxford, Windmill Road, OX3 7LD Oxford, UK. 5MRC Lifecourse Epidemiology Unit, University of Southampton, University Road, SO17 1BJ Southampton, UK.

Received: 13 April 2017 Accepted: 27 June 2017
Published online: 12 July 2017

**References**
1. Rana JS, Mannan A, Donnell-Fink L, et al. Longevity of the placebo effect in the therapeutic angiogenesis and laser myocardiad revascularization trials in patients with coronary heart disease. Am J Cardiol. 2005;95:1456–9. doi:10.1016/j.amjcard.2005.02.013.
2. Koop YH, Jung WY. Time course of placebo effect of acupuncture on pain: a systematic review. ISRN Pain. 2012;2013:1–8. doi:10.1155/2013/204108.
3. Wartolowska KA, Feakins BG, Collins GS, et al. The magnitude and temporal changes of response in the placebo arm of surgical randomized controlled trials - a systematic review and meta-analysis. Trials. 2016;17:389. doi:10.1186/s13063-016-1720-3.
4. Walach H, Maidhof C. Is the placebo effect dependent on time? In: Kirsch I, editor. Expectancy, experience, and behavior. Washington, DC: American Psychological Association; 1999. p. 321–32.
5. Kapchak TJ, Stason WB, Davis RB, et al. Sham device v inert pill randomized controlled trial of two placebo treatments. BMJ. 2006;332:391–7. doi:10.1136/bmj.38726.603310.55.
6. Agid O, Siu CO, Portin SG, et al. Meta-regression analysis of placebo response in antipsychotic trials, 1970–2010. Am J Psychiatry. 2013;170:1335–44. doi:10.1176/appi.ajp.2013.12030315.
7. Curie A, Yang K, Kirsch I, et al. Placebo responses in genetically determined intellectual disability: a meta-analysis. PLoS One. 2015;10:1–16. doi:10.1371/journal.pone.0133116.
8. Turner JA, Deyo RA, Loeser JD, et al. The importance of placebo effects in pain treatment and research. JAMA. 1994;271:1609–14. doi:10.1001/jama.199403510440069036.
9. Vase L, Robinson ME, Verne GN, et al. The contributions of suggestion, desire, and expectation to placebo effects in inurable bowel syndrome patients. An empirical investigation. Pain. 2003;105:17–25.
10. Levine JD. Role of pain in placebo analgesia. PNAS. 1979;76:3528–31.
11. Wartolowska K, Judge A, Hopewell S, et al. Use of placebo controls in the evaluation of surgery: systematic review. BMJ. 2014;348:g3253. doi:10.1136/bmj.g3253.
12. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org.
13. Ishak KJ, Platt RW, Joseph L, et al. Meta-analysis of longitudinal studies. Clin Trials. 2007;7:525–39.
14. Brown H, Prescott R. Applied mixed models in medicine. 3rd ed. United Kingdom: Wiley; 2015. http://onlineibrary.wiley.com/book/10.1002/9781118778210.
15. Hebjørgbonn A, Gatzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med. 2001;344:1594–602. doi:10.1056/NEJM200105243442106.
16. Ernst E, Reisch KL. Concept of true and perceived placebo effects. BMJ. 1995;311:151–3. doi:10.1136/bmj.311.7004.151.
17. Schwartz MP, Wellink H, Gooszen HG, et al. Endoscopic gastroplication for the treatment of gastro-oesophageal reflux disease: a randomised, sham-controlled trial. Gut. 2007;56:30–8. doi:10.1136/gut.2006.096942.
18. Rück A, Šylvén C. ‘Improvement’ in the placebo group could be due to regression to the mean as well as to sociobiologic factors. Am J Cardiol. 2006;97:152–3. doi:10.1016/j.amjcard.2005.08.015.
19. Morton V, Torgerson DJ. Regression to the mean: treatment effect without the intervention. J Eval Clin Pract. 2005;11:59–65. doi:10.1111/j.1365-2753.2004.00505.x.
20. McDonald CJ, Mazzuca SA, McCabe GP. How much of the placebo ‘effect’ is really statistical regression? Stat Med. 1982;2:417–27. doi:10.1002/sim.4780020401.
21. Walach H, Sadagiani C, Dehm C, et al. The therapeutic effect of clinical trials: understanding placebo response rates in clinical trials – a secondary analysis. BMC Med Res Methodol. 2005;5:26. doi:10.1186/1471-2288-5:26.
22. Colloca L, Petrovic P, Wagner TD, et al. How the number of learning trials affects placebo and nocebo responses. Pain. 2010;151:430–9. doi:10.1016/j.pain.2010.08.007.
23. Vase L, Vollert J, Finnerup NB, et al. Predictors of the placebo analgesia response in randomized controlled trials of chronic pain: a meta-analysis of the individual data from nine industrially sponsored trials. Pain. 2015;156:719–802. doi:10.1097/j.pain.0000000000000217.