PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effect of intrathecal morphine and epidural analgesia on postoperative recovery after abdominal surgery for gynecologic malignancy. An open-label randomized trial.
AUTHORS	Kjolhede, Preben; Bergdahl, Olga; Borendal Wodlin, Ninnie; Nilsson, Lena

VERSION 1 - REVIEW

REVIEWER	AJW Teunissen Maasstad Ziekenhuis The Netherlands
REVIEW RETURNED	09-Jun-2018

GENERAL COMMENTS	- Patients, staff and researcher were not blinded. This leads a lot of possibilities for bias. How was this prevented? - Difference in discharge possibilities on day 3 is not a good endpoint if the epidural stays in place until day 3 . If you stop a excellent pain treatment like epidural, patients have to adjust to pain. This takes time. Also patients will have difficulties in ambulation until the epidural stops. So discharge is hampered by an epidural no matter what treatment for pain the other group of patients receives. No explanation for the calculation of total opioid consumption [so how to repeat this study? or check the results?] If you do a single centre study -- rescue postoperative pain opioids should be the same in both groups. If you are bot allowed to give NSAIDS to the epidural group, than you should also withhold NSAIDS in the ITM group to make comparison possible The possibility that clonidine could be a contributing factor for the results is not discussed. The quality improvement is not proven [besides earlier discharge] so the tittle suggest something that was not the result of this study. - Literature since 2016 is hardly mentioned --- [intrathecal morphine ERAS] results in usefull hits

REVIEWER	Karl Kothbauer Luzerner Kantonsspital Division of Neurosurgery Lucerne, Switzerland
REVIEW RETURNED	13-Jun-2018

GENERAL COMMENTS	The authors have provided a readable and plausible manuscript about the benefits of single injection intrathecal Morphin vs. continuous postOP peridural analgesia in gynecologic abdominal

	surgery. In this day of ubiquitous pressure to reduce cost in healthcare, rapid recovery from major surgery may provide a substantial cost advantage, when it leads to shorter postOP Hospital stay. According to this manuscript, intrathecal morphine in the setting chosen accomplishes just that. There are no major points which from my perspective would require changing. The number and quality of illustrations is appropriate. Figure 1 contains minor editing errors (randomiz/s/ed, clinical praxis - change to practice). In table 4 the p-value is confusing: is it for all Grades or Grade 2 only? And if the latter what about the rest? The discussion is clear and compares the results well with evidence from the published evidence. It appears that the discussion is mildly cost-biased, as the medical issues about pain relief seem to get a bit less space. The fact that the pain in NRS is under 2 in the first two days in the epidural group but between 2 and 3 in the intrathecal group is well enough explained, but may deserve a little more attention. After addressing these rather minor issues I would recommend publication of this paper.

REVIEWER	Deena Harji University Hospital of North Tees
REVIEW RETURNED	02-Aug-2018

GENERAL COMMENTS	Congratulations to the authors for completing this RCT. I have some comments with regards to the trial design: 1. The aim of this trial is compare ITM and EDA and assess LoS and QoL. LoS is an important endpoint, however should not be the primary endpoint. It is disappointing to see that pain scores are not the primary endpoint given that the intervention being tested is an analgesic intervention and would be far more patient-centric and relevant. 2. In the EDA arm - the epidural intervention continues till Day 3. This is a flaw in trial design given that one of the primary endpoints is proportion of women discharged at Day 3. It is therefore not unsurprising that a higher proportion of women in ITM were discharged within 3 days compared to the EDA group. 3. The lack of patient and public involvement is also disappointing. Is LoS or LoS <3 days an important endpoint for patients? Collecting QoL is not a surrogate marker for PPI input. Are the QoL measures collected relevant and applicable to patients? 4. There is no cost-effectiveness data reported in this trial. Were there outcomes assessed and measured. Given these issues with the methodology in trial design and the results I'm not sure of the clinical impact of the results of this trial and whether they would be sufficient to change practice.

REVIEWER	Gabor Mihala School of Medicine, Griffith University, Australia
REVIEW RETURNED	16-Aug-2018

GENERAL COMMENTS	Thank you for the opportunity to review your manuscript (from statistical analysis point of view). Abstract, Objectives. It is confusing that QoL is mentioned, which is one of the secondary outcomes. Is QoL more important than the other sec. outcomes? Why is the other primary outcome (%) not mentioned here? Abstract, Results. Analgesics amount results should be here, as it was a study outcome. Abstract, Conclusions. Replace 'facilitates a short' with 'associated with shorter'. Introduction, first sentence. Suggested reference: Devlin, N.J., et al., Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Economics, 2018. 27(1): p. 7-22. Page 7, row 8. In general, I am wondering about the validity of SF6d and EQ5d for this setting and study population. The most important question was most likely about pain, which is a component in both tools, but most questions in these tools focus on aspects not relevant to inpatient population (e.g. work, social functioning, vitality, mobility). Why were both applied? Page 7, row 28. Power analysis or sample size calculation? I think the second. I did a quick check using Stata 15, and the group sizes appear to be a bit low (also had to guess the LOS in the control group). What were your parameters? Page 7, row 36. Please also present spread (inter-quartile range or standard deviation). Statistical analysis, in general. There were too many hypothesis tests (p-values) performed throughout the analysis, which increases the likelihood of error due to multiple comparisons. Perhaps, conclusions should be drawn from the results of primary outcome analyses, while the results of secondary outcomes could be presented descriptively only (no statistical analysis for these). Also, analysis of LOS could be dropped in favour of the dichotomous time outcome measure (considering the low sample size and usually skewed distribution of LOS with outliers). Clinical significance should also be considered and discussed (not just statistical). Considering the number of discharges on day 3 (37 in both groups), the multivariable model appears to be overloaded with covariates. Please provide more details on variable selection, model building and testing. In prospective trials, the probability of events occurring in the exposed group compared to the non-exposed group should be described with Rate Ratio (Relative Risk). Odds Ratio is used in case-control studies. RR can also be obtained with regression methods.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: AJW Teunissen

Institution and Country: Maasstad Ziekenhuis, The Netherlands Please state any competing interests or state ‘None declared’: none declared

Please leave your comments for the authors below

- Patients, staff and researcher were not blinded. This leads a lot of possibilities for bias. How was this prevented?

It was for obvious reasons not ethically possible to blind patients or staff. The researchers were only in a few cases directly involved in patient care. Care was given by staff that was not involved in the study planning or reporting. Both analgesic regimens were standard care at the time of the study, although a variant of ITM (without clonidine) was mostly used in benign abdominal gynecological surgery, thus the staff were well experienced in the care and standard monitoring of the vital parameters of patients having regional analgesics and consequently treated the participants similar based on their clinical experience. The subjective data were based on forms with self-reported information and participation in the trial was voluntary. Although there may be a risk for certain bias in our design we consider it difficult to avoid this and find it less likely that it will influence the results substantially. We have now commented on this issue in the Discussion.

- Difference in discharge possibilities on day 3 is not a good endpoint if the epidural stays in place until day 3 . If you stop a excellent pain treatment like epidural, patients have to adjust to pain. This takes time. Also patients will have difficulties in ambulation until the epidural stops. So discharge is hampered by an epidural no matter what treatment for pain the other group of patients receives.

We agree partly with the reviewer. In accordance with the trial registration the primary endpoint was length of hospital stay and we refer to this in the revised version. The proportion of patients discharged on day 3 is now categorized as a secondary post hoc outcome. Our argument for retaining the information about proportions of women discharged on the third postoperative day is to demonstrate a drawback of the EDA. The study purpose was to compare the two regional anesthetic regimens used at our department (ITM and EDA). The standard was to leave the EDA in place until the third postoperative morning. It was therefore necessary to obtain a measure for how large a proportion that de facto was discharged on the third day to examine whether the EDA model de facto delays discharge. All patients with EDA have full ambulation from postoperative day 1. The EDA catheter was connected to a small portable infusion pump. This is an important part of the ERAS program. We therefore do not believe that the reviewer’s considerations concerning difficulties with ambulation until the EDA stops is correct. The pain assessments demonstrate a drawback for the EDA on day 3. It seems that the ITM and the EDA react different at the time when the EDA stops which we already have commented on in the Discussion but developed it further.

No explanation for the calculation of total opioid consumption [so how to repeat this study? or check the results?]

All opioids, independent of administration route and including the ITM and the EDA, were converted into an equivalent iv. morphine dose. This procedure is already described in the Material and Methods part of the manuscript. The references to how the conversion into equivalent iv morphine was done is presented [19,20]

If you do a single centre study -- rescue postoperative pain opioids should be the same in both groups.

Rescue opioids (iv. morphine, 0.5-1 mg, oxycodone 5 mg orally or iv. patient-controlled analgesia with morphine) were the same in both groups. We have clarified this in the revised manuscript.

If you are bot allowed to give NSAIDS to the epidural group, than you should also withhold NSAIDS in the ITM group to make comparison possible The possibility that clonidine could be a contributing factor for the results is not discussed.

This is a study comparing two multimodal analgesic regimens that were considered clinically relevant. For that reason, we aimed to make each regimen as optimal as possible. The aim was not to compare the intrathecal route to the epidural route per se. We have clarified this in the revised manuscript. We agree with the reviewer that the influence of clonidine needs further discussion and have added a sentence of that in the Discussion section. We are not sure if the reviewer also suggests a synergistic effect between clonidine and NSAIDs. To the best of our knowledge a small synergistic interaction in connection with spinal administration of the drugs in rats has been demonstrated (Anesthesiology. 1993 Aug;79(2):270-81). We have not discussed this in the revised manuscript.

The quality improvement is not proven [besides earlier discharge] so the tittle suggest something that was not the result of this study.

That is debatable but we comply with the reviewer and change the title to a more neutral: The effect of intrathecal morphine and epidural analgesia on postoperative recovery after abdominal surgery for gynecologic malignancy. An open-label randomized trial.

- Literature since 2016 is hardly mentioned --- [intrathecal morphine ERAS] results in usefull hits

We appreciate the reviewers suggestion and have once again searched the literature and up-dated the reference list with one publication from 2017 and four recent publications from 2018.

Reviewer: 2

Reviewer Name: Karl Kothbauer

Institution and Country: Luzerner Kantonsspital, Division of Neurosurgery, Lucerne, Switzerland
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below The authors have provided a readable and plausible manuscript about the benefits of single injection intrathecal Morphin vs. continuous postOP peridural analgesia in gynecologic abdominal surgery. In this day of ubiquitous pressure to reduce cost in healthcare, rapid recovery from major surgery may provide a substantial cost advantage, when it leads to shorter postOP Hospital stay. According to this manuscript, intrathecal morphine in the setting chosen accomplishes just that. There are no major points which from my perspective would require changing. The number and quality of illustrations is appropriate.

We appreciate the reviewers nice summary and assessment.

Figure 1 contains minor editing errors (randomiz/s/ed, clinical praxis - change to practice).

Sorry, for the mistakes which now have been corrected.

In table 4 the p-value is confusing: is it for all Grades or Grade 2 only? And if the latter what about the rest?

We understand the lay-out problem and has therefore deleted the column with the p-value in the table. Instead, the p-value is noted in the subtext together with the test used. We have also clarified the p-value by even adding it in the Results section.

The discussion is clear and compares the results well with evidence from the published evidence. It appears that the discussion is mildly cost-biased, as the medical issues about pain relief seem to get a bit less space. The fact that the pain in NRS is under 2 in the first two days in the epidural group but between 2 and 3 in the intrathecal group is well enough explained, but may deserve a little more attention.

We agree, and have commented on this in the Discussion.

After addressing these rather minor issues I would recommend publication of this paper.

Reviewer: 3

Reviewer Name: Deena Harji

Institution and Country: University Hospital of North Tees Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below Congratulations to the authors for completing this RCT.

I have some comments with regards to the trial design:

1. The aim of this trial is compare ITM and EDA and assess LoS and QoL. LoS is an important endpoint, however should not be the primary endpoint. It is disappointing to see that pain scores are not the primary endpoint given that the intervention being tested is an analgesic intervention and would be far more patient-centric and relevant.

We understand the point being raised by the reviewer, but do not completely agree. In the trial registration the LOS was already our primary endpoint and can at this point not be changed. We agree that the intervention tested is analgesic, but the ITM or EDA regimen also could influence ambulation, PONV, sleep, QoL, urinary retention etc. We argue that LOS is an endpoint that is dependent on several factors and as such, an important measure and well suited as primary outcome measure. We do not think that pain score is more patient-centric. It is important to recognize that it is not only how you experience pain, but also how you cope with it, that really matter. We regard QoL as a compound measure for how you experience recovery. By interpreting it in a qualitative way the latter would give more sense. Thus pain score as primary outcome would, for obvious reasons, be less favourable in this trial. We have presented both elements, i.e. QoL and pain scores, as these are complementing each other.

2. In the EDA arm - the epidural intervention continues till Day 3. This is a flaw in trial design given that one of the primary endpoints is proportion of women discharged at Day 3. It is therefore not unsurprising that a higher proportion of women in ITM were discharged within 3 days compared to the EDA group.

We do not agree with the reviewer and refer to our answers to reviewer 1 and the AE comments.

3. The lack of patient and public involvement is also disappointing. Is LoS or LoS <3 days an important endpoint for patients? Collecting QoL is not a surrogate marker for PPI input. Are the QoL measures collected relevant and applicable to patients?

We have taken the reviewers opinion ad notam. We do believe the questions are complex and cannot be answered in a simple way, but overall we believe that the LOS and LOS<3 days are meaningful endpoint even for patients. We have commented on the question about the relevance of the LOS<3 days as well as QoL measures in the Discussion.

4 . There is no cost-effectiveness data reported in this trial. Were there outcomes assessed and measured.

Correct. Health-economic was not a part of the study objectives.

Given these issues with the methodology in trial design and the results I'm not sure of the clinical impact of the results of this trial and whether they would be sufficient to change practice.

We argue that it is important to scientifically investigate and report different ways to handle postoperative analgesia in an ERAS setting. If a technically more simple method (ITM) that also is less resource demanding on the ward proves equal in many aspects to the prevailing routine (EDA), we believe that practise might be changed for that reason. Indeed, it should also encourage other research groups to improve different parts of the described regimens further through clinical trials.

Reviewer: 4

Reviewer Name: Gabor Mihala

Institution and Country: School of Medicine, Griffith University, Australia Please state any competing interests or state ‘None declared’: None.

Please leave your comments for the authors below Thank you for the opportunity to review your manuscript (from statistical analysis point of view).

Abstract, Objectives. It is confusing that QoL is mentioned, which is one of the secondary outcomes. Is QoL more important than the other sec. outcomes? Why is the other primary outcome (%) not mentioned here?

The primary outcome is according to the trial registration length of stay. This has now been clarified and simultaneously the secondary outcome variables in accordance with trial registration (pain assessment) have been elucidated.

Abstract, Results. Analgesics amount results should be here, as it was a study outcome.

We comply with the reviewers request and have now added the total amount of opioids used postoperatively the first week in the Abstract as well as in the Results.

Abstract, Conclusions. Replace 'facilitates a short' with 'associated with shorter'.

This has been done.

Introduction, first sentence. Suggested reference: Devlin, N.J., et al., Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Economics, 2018. 27(1): p. 7-22.

We thank the reviewer for this reference. As the reference describes a new value set of the EQ-5D we have chosen to expand on the discussion part relating to QoL and have added this reference there.

Page 7, row 8. In general, I am wondering about the validity of SF6d and EQ5d for this setting and study population. The most important question was most likely about pain, which is a component in

both tools, but most questions in these tools focus on aspects not relevant to inpatient population (e.g. work, social functioning, vitality, mobility). Why were both applied?

Pain is one dimension in health related quality of life measurement. We were not interested in only the pain-related Qol, but in the overall experienced QoL. If we only were interested in the pain we would of course have chosen a symptom specific form. The dimensions in the EQ-5D (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) are to a large extent relevant to the inpatient population. One can of course question the relevance of the dimension “usual activities (e.g. work, study, housework, family and leisure activities)” during hospital stay, but as measurements were undertaken repeatedly during the study until 6 weeks after surgery the vast majority of the measured times occurred when the patients were at home. The SF-36 was for obvious reasons only recorded twice, before surgery and at 6 weeks after surgery. The eight sections are all relevant. We agree that there is no evidence of content validity for our patient group, both forms are widely used in all areas of medicine and allow comparisons with population norms. We have commented on this in the Discussion. Basically, EQ-5D may be used to detect fast changes in QoL, such as day-by-day, whereas the SF-36 form in general detect changes over a longer period, one to four weeks. We choose to use both forms, EQ-5D and SF-36, in order to be complementary and to look for short term (day-by-day) changes as well as over a four weeks period.

Page 7, row 28. Power analysis or sample size calculation? I think the second. I did a quick check using Stata 15, and the group sizes appear to be a bit low (also had to guess the LOS in the control group). What were your parameters?

We have changed as suggested to “sample size calculation” and have given the standard deviation for LOS that was used for the calculation of the sample size.

Page 7, row 36. Please also present spread (inter-quartile range or standard deviation).

We comply with the reviewer’s request and have changed range to inter quartile range throughout the manuscript.

Statistical analysis, in general. There were too many hypothesis tests (p-values) performed throughout the analysis, which increases the likelihood of error due to multiple comparisons.

We appreciate the reviewer’s considerations. Concerning too many p-values and increased likelihood of error due to multiple testing, that is of course theoretically correct, but at the same time, we present the p-values and leave it to the reader to judge the level and importance by themselves.

We have further reconsidered the statistics in the manuscript and discussed it thoroughly with our statistician. Since the randomization was well balanced and none of the clinical parameters that could confound the results differed significantly between the groups we redraw all multivariate analyses. Based on this, some of the concerns raised by the reviewer are solved automatically and we therefore do not give a point-by-point response to these issues below.

Perhaps, conclusions should be drawn from the results of primary outcome analyses, while the results of secondary outcomes could be presented descriptively only (no statistical analysis for these).

We do not quite understand the reviewer here.

Also, analysis of LOS could be dropped in favour of the dichotomous time outcome measure (considering the low sample size and usually skewed distribution of LOS with outliers). Clinical significance should also be considered and discussed (not just statistical).

In accordance with the response to the AE as well as other of the reviewers on the question about primary and secondary outcomes we prefer to still have LOS on the continuous scale as outcome.

The LOS data are analysed by means of a non-parametric test, which is a robust way that almost completely eliminate the effect of outliers. The clinical significance has now been commented on in the Discussion.

Considering the number of discharges on day 3 (37 in both groups), the multivariable model appears to be overloaded with covariates. Please provide more details on variable selection, model building and testing. In prospective trials, the probability of events occurring in the exposed group compared to the non-exposed group should be described with Rate Ratio (Relative Risk). Odds Ratio is used in case-control studies. RR can also be obtained with regression methods.

See our response to reviewer 4's question above about statistical analysis, in general.

VERSION 2 – REVIEW

REVIEWER	Aart Jan W. Teunissen Maasstadziekenhuis The Netherlands
REVIEW RETURNED	08-Nov-2018

GENERAL COMMENTS	I liked how you have improved the article and made the conclusions appropriate and the discussion correct and enough reflection on design shortcomings. Some small remarks: If diclofenac is not given to one group on day 0,1 and 2 and is given to the other group there must me a statistical difference between the groups regarding non-opioids. The result: "whereas the use of non-opiate use was similar between the groups" should be restricted to day 3 and de days after day 3. It is true for those days, but it simply cannot be true for day 0-1-2. Figure 2 does also not show this conclusion. Just some change in the result section will make it correct. it is all right in the discussion part of the article 10. Chopra P, Talwar V. "Low dose intrathecal clonidine and fentanyl added to hyperbaric bupivacaine prolongs analgesia in gynecological surgery". J Anaesthesiol Clin Pharmacol. 2014;30:233-7. Does not proof that there is a rational for adding clonidine to intrathecal morphine only a rational for using intrathecal clonidine for postoperative pain Sites BD1, Beach M, Biggs R, Rohan C, Wiley C, Rassias A, Gregory J, Fanciullo G. Anesth Analg. 2003 Apr;96(4):1083-8; "Intrathecal clonidine added to a bupivacaine-morphine spinal anesthetic improves postoperative analgesia for total knee arthroplasty ". Does give a rational for adding clonidine to morphine intrathecally So why not use a referral like this for the discussion part of your article?

REVIEWER	Gabor Mihala School of Medicine, Griffith University
REVIEW RETURNED	04-Nov-2018

GENERAL COMMENTS	Thank you for the detailed response. I need to make a further 3 comments here. Sample size calculation: please also specify the control group mean entered into the sample size calculator (or other parameters as necessary). The calculation appears to be correct when I use 4.3. Regarding my earlier comment "Perhaps, conclusions should be drawn from the results of primary outcome analyses, while the results of secondary outcomes could be presented descriptively only (no statistical analysis for these)". This suggestion was intended to reduce the number of testing (p-values) throughout, by analysing LOS for statistical and clinical significance, and analyse the other outcomes for clinical significance only. It was an idea for you to consider. Statistical methods in general: it appears that the ANCOVA and logistic regressions was replaced with unadjusted repeated measures ANOVAs in this revision. Did you check that the assumptions of ANOVA were met, i.e. appropriate for the type of outcomes you used it for?

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: AJW Teunissen

Institution and Country: Maasstad Ziekenhuis, The Netherlands Please state any competing interests or state ‘None declared’: none declared

Some small remarks:

If diclofenac is not given to one group on day 0,1 and 2 and is given to the other group there must me a statistical difference between the groups regarding non-opioids.

The result: "whereas the use of non-opiate use was similar between the groups" should be restricted to day 3 and de days after day 3.

It is true for those days, but it simply cannot be true for day 0-1-2.

Figure 2 does also not show this conclusion.

Just some change in the result section will make it correct. it is all right in the discussion part of the article

10. Chopra P, Talwar V. “Low dose intrathecal clonidine and fentanyl added to hyperbaric bupivacaine prolongs analgesia in gynecological surgery”. J Anaesthesiol Clin Pharmacol. 2014;30:233-7.

Does not proof that there is a rational for adding clonidine to intrathecal morphine only a rational for using intrathecal clonidine for postoperative pain

Sites BD1, Beach M, Biggs R, Rohan C, Wiley C, Rassias A, Gregory J, Fanciullo G.

Anesth Analg. 2003 Apr;96(4):1083-8;

“Intrathecal clonidine added to a bupivacaine-morphine spinal anesthetic improves postoperative analgesia for total knee arthroplasty “.

Does give a rational for adding clonidine to morphine intrathecally So why not use a referral like this for the discussion part of your article?

We appreciate the clarifications suggested by the reviewer and have changed the text according to the request of the reviewer concerning non-opioids. Besides, we have recalculated the statistics so the period of comparison consists of Day 3 to Day 42. Thus, the days when the protocol requirements of non-opioids were unequal are excluded, i.e. Day 0 to Day 2. The reference has been changed according to the reviewer’s suggestion.

Reviewer: 4

Reviewer Name: Gabor Mihala

Institution and Country: School of Medicine, Griffith University, Australia Please state any competing interests or state ‘None declared’: None.

Please leave your comments for the authors below Thank you for the opportunity to review your manuscript (from statistical analysis point of view).

Sample size calculation: please also specify the control group mean entered into the sample size calculator (or other parameters as necessary). The calculation appears to be correct when I use 4.3.

We are not sure we understand what professor Mihala means by “control group mean”. Our calculation was based on the equation for sample size calculations for independent means:

n/group = 2* (((zα/2+zβ)*SD)/Δ) 2

where zα/2 + zβ = 1.96+0.842 (for a two-sided test. α = 5% and β= 80%).

SD is the standard deviation of the outcome and Δ is the mean difference between the groups that is considered clinically important.

Thus N =2 * n.

In the text, we already have stated the SD and the minimum clinical important difference, Δ, in means. Thus, it seems that all prerequisites for calculating the sample size seems to be available.

Regarding my earlier comment "Perhaps, conclusions should be drawn from the results of primary outcome analyses, while the results of secondary outcomes could be presented descriptively only (no statistical analysis for these)". This suggestion was intended to reduce the number of testing (p-values) throughout, by analysing LOS for statistical and clinical significance, and analyse the other outcomes for clinical significance only. It was an idea for you to consider.

This is a great idea that we now have adopted. We have deleted the column with p-values in Table 2 and commented on the clinical significance of the differences between the groups in the Result-text (or actually, the lack of clinical significance in any of the items in Table 2).

Statistical methods in general: it appears that the ANCOVA and logistic regressions was replaced with unadjusted repeated measures ANOVAs in this revision. Did you check that the assumptions of ANOVA were met, i.e. appropriate for the type of outcomes you used it for?

The answer is yes. We analyzed the criteria for using the repeated measures ANOVA. The histograms revealed acceptable normal distributions of the outcomes. However, the sphericity, as measured by the Mauchly Sphericity test, was violated in all outcomes. Adjustments were therefore done by means of the Greenhouse-Geisser correction method.