PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Open reduction and internal fixation of humeral shaft fractures versus conservative treatment with a functional brace: a study protocol of a randomised controlled trial embedded in a cohort
AUTHORS	Rämö, Lasse; Taimela, Simo; Lepola, Vesa; Malmivaara, Antti; Lähdeoja, Tuomas; Paavola, Mika

VERSION 1 - REVIEW

REVIEWER	Kevin Tetsworth Royal Brisbane Hospital Department of Orthopaedic Surgery Brisbane, QLD, Australia
REVIEW RETURNED	04-Dec-2016

GENERAL COMMENTS	In the "Statistics" section, I have responded "no"; this is because i do not believe specialist statistical review is necessary for this study. With respect to query 13, i have also responded with a "no". Although the authors have done a nice job of adhering to the CONSORT guidelines, they have not included a CONSORT checklist. I would suggest they do so, to confirm all aspects of the CONSORT guidelines have been satisfied. Specific comments: (1) in several instances, the authors refer to "dislocated" fractures - I would prefer you use the more conventional term "displaced" (line 167 is one example, there is at least one other similar error) (2) line 63 in Abstract - the "declined cohort" is typically referred to as the "Intention to Treat" group, and in an identical fashion it is possible to conduct an analysis including these patients in the group they would have been assigned to, increasing the generalisability of the study results (3) I would also note the Abstract is formatted in an odd way; is it possible to revise this so that it takes a more normal appearance (sometimes referred to as IMRAD format). (4) Both the Introduction and the Abstract should contain a more explicit study hypothesis; preferably the identical sentence appears in both portions of the manuscript. (5) The primary outcome measure should be explicitly stated immediately before the study hypothesis. Although you have included this information in Table 3, together with secondary outcome measures, the primary outcome measure should be stated clearly in the text near the end of the introduction, immediately

	preceding the study hypothesis. (6) Line 178 - I find it difficult to believe a 4.5 mm plate will be used in all patients. I have many elderly or patients of Asian descent where a 4.5 mm plate is far too wide. I would suggest you consider using either a 3.5 mm or 4.5 mm plate at the surgeon's discretion, based on each patient's anatomy and size. Alternatively, you could consider using a 3.5 mm plate in all operative patients. (7) Line 196 - the word "primarily" is unnecessary (8) Line 207 - Define DASH and 15D the first time they are used in the manuscript (9) Lines 215-216 - awkward - please re-write, using articles as necessary and appropriate (10) Line 232- define NRS when first used in the manuscript (11) Line 240 - your primary outcome measure is the DASH score at 12 months - it is completely irrelevant and unnecessary to define a definition for recovery as you have done here - use this or another validated outcome instrument as your measure of recovery - I appreciate that you prefer to use a patient reported subjective measure, and that's fine (12) Lines 251-254 - section describing the numerical rating scale - needs to be placed before the sections on recovery, at line 230-249 (13) Lines 257-260 - awkward - please re-write, using articles as necessary and appropriate (14) Line 269 - awkward - please re-write, using articles as necessary and appropriate (15) Line 290 - awkward - please re-write - I would recommend "...suggests that operative exploration is not mandatory..." (16) Line 296 - the word "suturation" is unfortunately not an accepted word in English - please consider using "repair" instead (17) Line 323-330 - Sample size needs to be placed at the end of the statistics - seems out of place here - you have given some of the details without stating software package used - G-Power or similar? (18) Line 327 - I believe you are trying to demonstrate a clinically important difference between the two comparison groups - as this is currently written, you are already assuming ORIF will be superior - in an ideal study it is as likely nonoperative treatment will be superior, or that there will be no difference (19) This group you have called the "Declined Treatment Cohort", while laudable, is an unusual approach - most often we see an "Intention to Treat" analysis - could the authors please comment on why they have taken this approach (20) My only other comment with respect to the references is that the authors appear to be completely ignoring anything having to do with MIPO humeral fixation, using an anterior sub-muscular percutaneous plate - was this a conscious decision, to simplify the

	discussion and avoid any potential confounding influence? This should perhaps more correctly be considered a trial of ORIF vs non-operative management (21) I find the title difficult to read and lacking impact - please consider re-writing to make it clear this is a randomised controlled comparison of humeral shaft fractures treated either by open reduction and internal fixation or by non-operative means with a functional brace (22) How will you deal with failure of treatment? At least 30% of your non-operative patients will fail treatment in a brace, almost all studies appear to have similar rates for conversion to ORIF - even Sarmiento's paper lost over 30% to follow-up, I am assuming they were also unable to tolerate the brace or had unacceptably bad angular deformities or gross displacement uncontrolled by the brace - regardless, it was unclear from the manuscript how this issue would be addressed - could the authors please clarify

REVIEWER	Belloti, Joao Carlos Universidade Federal de SãoPaulo (UNIFESP) Escola Paulista de Medicina Brazil
REVIEW RETURNED	13-Jan-2017

GENERAL COMMENTS	- This is a well-designed multicentric ECR protocol that compares the surgical and non-surgical treatment of humerus diaphyseal fractures. - My main concern about this protocol is regarding the inclusion of patients with primary radial nerve injury, since this subgroup of patients has a very different clinical recovery from patients without nerve damage. I suggest that this subgroup of patients be analyzed in a parallel study. -I recommend that the authors review the description and classification of potential adverse events, since non-union and deep infection are not considered minor adverse events in the treatment of these fractures. There is no description of how adverse events will be accessed (SAEs and MAEs)

VERSION 1 – AUTHOR RESPONSE

We thank Kevin Tetsworth for the careful review of our paper and appreciate the insightful suggestions for improvement.

Reviewer comment 1:
In several instances, the authors refer to "dislocated" fractures - I would prefer you use the more conventional term "displaced" (line 167 is one example, there is at least one other similar error)

Authors' response 1:
We fully agree with this remark.

Authors' action 1:
We changed the term `dislocated´ to `displaced´.

Reviewer comment 2:
line 63 in Abstract - the "declined cohort" is typically referred to as the "Intention to Treat" group, and in an identical fashion it is possible to conduct an analysis including these patients in the group they would have been assigned to, increasing the generalisability of the study results

Authors' response 2:
Thank you very much for the comment, it helped us see that the parts concerning the declined cohort were written in a confusing fashion.

We believe that you are referring to the Zelen's design, in which patients are randomized before informed consent. In this design, because the group to which a given patient is assigned is already known, informed consent is sought conditionally. We agree that such randomisation without informed consent (RWIC) could potentially increase the number of trial participants, but it causes some major difficulties on the other hand. We acknowledge that cluster randomized designs and emergency research are routinely used with RWIC, with the justification that informed consent is infeasible in those settings. Other RWIC designs have raised serious concerns that they do not appropriately respect patient autonomy. Particularly a scheme with "experimental" and "standard" treatments also includes a risk for bias if eligible patients would be randomised prior to consent. If the patient was assigned to the "standard" therapy, it would be offered to the patient without the need for consent. If the patient was randomised to the "experimental" therapy, only then the patient would be asked for consent. If this patient refuses, however, then he/she is offered the "standard" therapy and an "intention-to-treat" analysis is performed based on the randomised assignment. Patients who are randomised to the "experimental" treatment and refuse will dilute the treatment difference at the time of data analysis and this introduces bias.

However, in our study, we have sought to enhance the external validity by adding two prospective cohort studies to run along our RCT proper. The "declined cohort" consists of patients who would be eligible to randomisation but are not willing to be randomised, usually due to having a strong preference to either treatment modality. In clinical practice, it is not feasible to define either of our treatment options as the "standard" therapy, since the current evidence leaves it to the treating surgeon and the patient to decide between surgical and non-surgical treatment.

More importantly, though, there are specific ethical problems also because the subjects are randomised to treatment without having been properly informed and without providing their consent, and subjects randomised to either of the options have been denied the chance of receiving the other option. We believe that this approach would not be allowed by any ethics review board in our country.

For these reasons, we chose to randomise the participants only after informed consent and to follow the "declined cohort" up in order to increase generalisability.

The reviewer comment helped us to understand that the relevant sections of the manuscript were unclear and confusing.

Author Action 2:
We have revised the description of the cohort(s) in the Abstract, and added the mention that the randomisation is done after and informed consent has been obtained. The sentence and the third paragraph of the Methods and analysis of the Abstract now read (lines 53-56):
80 patients from 18 years of age with humeral shaft fracture are randomly assigned after informed consent has been obtained to open reduction and internal fixation (ORIF) with locking plate or conservative treatment with functional bracing

and (lines 60-66)

Patients not willing to participate in the RCT are asked to participate in a prospective cohort follow-up study, “the declined cohort”. This cohort will be followed-up at the same time points as the randomised patients to assess the potential effect of participation bias on the results of the RCT and to enhance the external validity of the RCT. In one of the recruiting centres all co-operative patients with humeral shaft fractures, not eligible for randomisation, are asked to participate to a “non-eligible cohort” study.

We have also made a revision to the text by moving the description of the cohorts below section “Blinding” and made them a section of their own. The revised section is as follows (lines 363-400):

Surrounding cohort studies
The generalisability, or external validity of RCTs is often discussed and questioned due to ineligibility of many patients with the condition of interest, or the potential participation bias. To enhance the external validity of our trial and to study the participation bias, we introduced follow-up cohorts of the patients who decline to participate in the RCT and the patients with a humeral shaft fracture, but ineligible for the trial.

The Declined cohort
Eligible patients not willing to participate in randomisation are asked to participate to a follow-up study (referred to as the declined cohort). Usually the unwillingness to participate in the RCT is due to a strong preference for one of the treatment modalities and the resulting unwillingness to receive randomly assigned treatment. The patients receive usual care with the treatment method (conservative or ORIF) decided with the patient after information of both treatment methods is given. The cohort will follow the same follow-up protocol as the randomised trial (Figure 3 and 4). Analysis of the outcome measures is done separately from the randomised trial, and the results are compared to the results of the RCT. In addition to controlling for the potential effect of participation bias on external validity, this gives a possibility to evaluate potential effects of patient expectations to outcomes, when the patient has been able to participate in the decision of treatment method.

Non-eligible cohort
All compliant patients with fresh humeral shaft fracture, but not eligible to randomisation (reason being either fracture extending too proximally or distally or exclusion criteria is met, see Table 1), are asked to participate to a second prospective cohort study (later referred to as “the non-eligible cohort”) in one of the recruiting centres (Helsinki University Hospital). The fracture is treated with the discretion of treating physician either conservatively or operatively (ORIF or IMN) in the usual fashion. Patients fill the same baseline data as the randomised patients. The reason(s) for excluding from randomisation is recorded, the patient receives usual care and follow-up visits. The patients have a study follow-up visit at the outpatient clinic at 12 months, where the same questionnaires and measurements are performed as in the RCT (Figure 5). The results of the non-eligible cohort give us possibility to enhance the external validity of our trial. Non-compliant patients, fractures older than 2 weeks, end-stage malignancies, and patients with such severe trauma that baseline data is not possible to gather (typically multiple trauma patients with head trauma) are excluded from the non-eligible cohort.

We have also revised the last paragraph of section “Generalisability” in the Discussion, starting from line 583 to read as follows:
The current trial is designed to be pragmatic effectiveness trial, but still quite a number of patients with humeral shaft fracture will be excluded because of our exclusion criteria and due to fact that not all patients are willing to accept randomisation. To enhance the generalisability of our trial we have introduced the Declined and Non-eligible cohorts as described above. These cohorts give us a possibility to evaluate the external validity of the results of the randomised group.

Reviewer comment 3:
I would also note the Abstract is formatted in an odd way; is it possible to revise this so that it takes a more normal appearance (sometimes referred to as IMRAD format).

Authors' response 3:
Thank you for the comment. We have followed the general BMJ recommendations for manuscript formatting of protocol papers: "Abstract: this should be structured with the following sections. Introduction; Methods and analysis; Ethics and dissemination. Registration details should be included as a final section, if appropriate." (http://bmjopen.bmj.com/pages/authors/#studyprotocols). Moreover, we found the format we are using quite common in protocols of trials in BMJ Open, i.e.

-Bautista-Hernandez V, Cal-Purriños N, Arribas-Leal JM, et al Rapid Deployment Aortic Replacement (RADAR) Registry in Spain: a protocol BMJ Open 2017;7:e011437. doi: 10.1136/bmjopen-2016-011437
-Yeung J, Melody T, Kerr A on behalf of the TOPIC Study Investigators, et al Randomised controlled pilot study to investigate the effectiveness of thoracic epidural and paravertebral blockade in reducing chronic post-thoracotomy pain: TOPIC feasibility study protocol BMJ Open 2016;6:e012735. doi: 10.1136/bmjopen-2016-012735
-Rathod KS, Jones DA, Van-Eijl TJA, et al Randomised, double-blind, placebo-controlled study investigating the effects of inorganic nitrate on vascular function, platelet reactivity and restenosis in stable angina: protocol of the NITRATE-OCT study BMJ Open 2016;6:e012728. doi: 10.1136/bmjopen-2016-012728

Since this is a protocol article there are no results to be reported and thus the discussion paragraph is not feasible in abstract.

Reviewer comment 4:
Both the Introduction and the Abstract should contain a more explicit study hypothesis; preferably the identical sentence appears in both portions of the manuscript.

Authors' response 4:
Thank you for this insightful remark.

Authors' action 4:
We added our null hypothesis to both abstract and introduction as follows:
Our null hypothesis is that there is no clinically relevant difference in the primary outcome measure between the two treatment groups. We will consider a difference of minimum 10 points in DASH clinically relevant.

Reviewer comment 5:
The primary outcome measure should be explicitly stated immediately before the study hypothesis. Although you have included this information in Table 3, together with secondary outcome measures, the primary outcome measure should be stated clearly in the text near the end of the introduction, immediately preceding the study hypothesis.

Authors' response 5:
Thank you for this remark. We have followed the SPIRIT guideline as requested by the editor. In the SPIRIT format the primary outcome and rationale for its use is located later in the text than in intro (see SPIRIT checklist, item 12).

Reviewer comment 6:
Line 178 - I find it difficult to believe a 4.5 mm plate will be used in all patients. I have many elderly or patients of Asian descent where a 4.5 mm plate is far too wide. I would suggest you consider using either a 3.5 mm or 4.5 mm plate at the surgeon's discretion, based on each patient's anatomy and size. Alternatively, you could consider using a 3.5 mm plate in all operative patients.

Authors' response 6:
Thank you for this important comment. While the choice of 4.5 mm plate might be problematic in some regions, we do not believe that to be the case in Finland. We have over 10 years of experience of using the (narrow) 4.5 mm LCP plate for humeral shaft fractures in the Finnish hospital setting in high-volume university clinics, and we have not encountered problems with the plate being too wide. We fully agree that the broad 4.5 mm LCP plate (used mainly for certain femoral shaft fractures) is not suitable in most of the patients with humeral shaft fractures. We would like to retain the narrow 4.5 mm LCP plate as the only option, but if there is a deviation in the implants chosen this will be reported in the final paper containing the results of our study.

Authors' action 6:
We added the word "narrow" to the intervention paragraph to be more exact.

Reviewer comment 7:
Line 196 - the word "primarily" is unnecessary

Authors' response 7:
We think the word "primarily" is needed, since we are also using objective outcome measures not reported by the patients, i.e. the objective measurement section of the Constant Score.

Reviewer comment 8:
Line 207 - Define DASH and 15D the first time they are used in the manuscript

Authors' response 8:
Thank you for this remark which we fully agree with.

Authors' action 8:
We added the explanations of these terms to the correct part of the text and also to the abstract which included the same terms.

Reviewer comment 9:
Lines 215-216 - awkward - please re-write, using articles as necessary and appropriate

Authors' response 9:
Thank you for this comment. We agree on this.

Authors' action 9:
We re-wrote this sentence as follows:
DASH is a widely used and validated tool assessing upper-extremity related deficits and symptoms in the daily life reported by the patient.

Reviewer comment 10:
Line 232- define NRS when first used in the manuscript

Authors' response 10:
Thank you for this relevant remark.

Authors' action 10:
We changed the order of the secondary outcome measures so that now NRS is defined properly when first used.

Reviewer comment 11:
Line 240 - your primary outcome measure is the DASH score at 12 months - it is completely irrelevant and unnecessary to define a definition for recovery as you have done here - use this or another validated outcome instrument as your measure of recovery - I appreciate that you prefer to use a patient reported subjective measure, and that's fine

Authors' response 11:
We highly appreciate this remark, which prompted us to re-evaluate this issue. We realised that this section and its rationale in our manuscript was not very clear. Our aim was to carry out a responder analysis in order to assess the proportions of participants who achieve an acceptable symptom state (Patient Acceptable Symptom State, PASS). When evaluating response to treatment, PASS may be considered better than minimal clinically important improvement (MCII) because it is less sensitive to baseline levels of symptoms compared. Yet we admit that there are no universally accepted criteria for PASS in the context of the outcome of treatment of humeral fractures.
In addition, we are assessing the proportion of patients achieving the pre-injury level of DASH (i.e. "safe" estimate of recovery) and proportion of patients scoring pre-injury level of DASH plus 10 points. Since we have considered 10 points in DASH a minimal clinically important difference in our sample size calculation, we consider this a relevant responder analysis.

Authors' action 11:
The paragraphs 'Recovered patients' and 'Poor recovery' were revised as follows:

Responder Analysis
We will carry out a responder analysis, in which the proportions of patients reaching the patient acceptable symptom state (PASS) will be determined. We will analyse the cut off limit for PASS based on our primary outcome, DASH, using a receiver operating characteristic (ROC) curve analysis. Patient's global assessment of satisfaction with the treatment outcome will be used as an anchoring item. Patients who report very satisfied or satisfied in the patients' global assessment of satisfaction with treatment outcome will be categorized as "responders".
In addition, we will assess the proportion of patients scoring equal to or less than their pre-injury DASH-score plus 10 points, which we have chosen as the primary definition of clinical recovery to pre-injury level since the minimal clinically important difference (MCID) of DASH is 10 points. However, the DASH score has been shown to have a ceiling effect. In a population generally asymptomatic prior to the fracture the subjects may perceive residual disability with outcomes that fall within the MCID range of the pre-injury status of the instrument. Hence, we will also conduct a similar analysis of the patients who reach their reported pre-injury DASH score or less, which we consider as a definition of a conservative or "safe" estimate of recovery to pre-injury status.

Reviewer comment 12:
Lines 251-254 - section describing the numerical rating scale - needs to be placed before the sections on recovery, at line 230-249

Authors' response 12:
We agree with this comment and this is covered with our action to comment 10.

Reviewer comment 13:
Lines 257-260 - awkward - please re-write, using articles as necessary and appropriate

Authors' response 13:
Thank you for this comment we agree with.

Authors' action 13:
We re-wrote this sentence as follows:
The Constant Score (CS) is widely used scale for shoulder function containing both clinician-assessed physical examinations and patient-reported assessments for pain, activities of daily living (ADL) functions and range of motion (ROM) in affected and non-affected shoulder. We added ROM of elbow to the physical examination, since it has shown to be affected by humeral shaft fractures. ROM is measured with a goniometer and strength with calibrated spring balance.

Reviewer comment 14:
Line 269 - awkward - please re-write, using articles as necessary and appropriate

Authors' response 14:
Thank you for this remark. We think there was a misspelling in our manuscript. However, we think the entire sentence can be left out since the issue is dealt with in the section 'Use of health care services and cost-effectiveness'. (Lines 341-350).

Authors' action 14:
We removed the entire sentence from the end of the section 15D.

Reviewer comment 15:
Line 290 - awkward - please re-write - I would recommend "...suggests that operative exploration is not mandatory..."

Authors' response 15:
We agree with the comment.

Authors' action 15:
We changed the sentence as follows:
Current literature suggests that operative exploration is not mandatory in primary radial nerve palsy in association with humeral shaft fracture.

Reviewer comment 16:
Line 296 - the word "suturation" is unfortunately not an accepted word in English - please consider using "repair" instead

Authors' response 15:
We thank for this remark.

Authors' action 16:
We changed this from `suturation´ to `nerve repair´.

Reviewer comment 17:
Line 323-330 - Sample size needs to be placed at the end of the statistics - seems out of place here - you have given some of the details without stating software package used - G-Power or similar?

Authors' response 17:
Thank you for this remark. As in comment 5 we stated, we have used the SPIRIT checklist. The Sample Size (item 14) is placed now as suggested in SPIRIT checklist.

Authors' action 17:
G*Power version 3.1 was used and we re-wrote the first sentence of this paragraph as follows:
The sample size calculation was done using G*Power 3.1 and is based on DASH as the primary outcome measure in this trial.

Reviewer comment 18:
Line 327 - I believe you are trying to demonstrate a clinically important difference between the two comparison groups - as this is currently written, you are already assuming ORIF will be superior - in an ideal study it is as likely nonoperative treatment will be superior, or that there will be no difference

Authors' response 18:
We thank the reviewer for this very relevant comment.

Authors' action 18:
We changed this sentence to a neutral approach as follows:
Using these assumptions the required sample size is 35 per group with 80% power to show clinically important difference between the treatment methods with two-sided type I error rate of 5%.

Reviewer comment 19:
This group you have called the "Declined Treatment Cohort", while laudable, is an unusual approach - most often we see an "Intention to Treat" analysis - could the authors please comment on why they have taken this approach

Authors' response 19:
Please, see our response to comment 2.

Reviewer comment 20:
My only other comment with respect to the references is that the authors appear to be completely ignoring anything having to do with MIPO humeral fixation, using an anterior sub-muscular percutaneous plate - was this a conscious decision, to simplify the discussion and avoid any potential confounding influence? This should perhaps more correctly be considered a trial of ORIF vs non-operative management

Authors' response 20:
Thank you for this insightful comment. At the time of the planning of our trial, the MIPO-plating of humeral shaft fractures was not common. The technique is also more demanding expertise-wise, and we sought to keep our trial pragmatic and applicable to most centres treating these fractures. Also, the primary evidence gap in the treatment of these fractures is, in our opinion, whether operative treatment is beneficial in the first place, not which operative method would yield the best results. We agree that our trial is a trial of ORIF vs non-operative management.

We are aware of one trial (Matsunaga et al, 2013, ISRCTN24835397) comparing MIPO to bracing, the results of which appear to have not been published yet. We think that there is still lack of evidence supporting MIPO over ORIF as a surgical option when running a somewhat pragmatic trial as ours. Matsunaga's results may have effect on this in the future. In the discussion, we intentionally left speculation on MIPO vs ORIF out to simplify things.

Authors' action 20:
We have now added one reference about the MIPO-technique to the introduction (An et al, 2010).

Reviewer comment 21:

I find the title difficult to read and lacking impact - please consider re-writing to make it clear this is a randomised controlled comparison of humeral shaft fractures treated either by open reduction and internal fixation or by non-operative means with a functional brace

Authors' response 21:
Thank you for this very important comment.

Authors' action 21:
We clarified the title to specify the surgical treatment method as follows:
Open reduction and internal fixation of humeral shaft fractures versus conservative treatment with a functional brace: a study protocol of a randomised controlled trial embedded in a cohort

Reviewer comment 22:
How will you deal with failure of treatment? At least 30% of your non-operative patients will fail treatment in a brace, almost all studies appear to have similar rates for conversion to ORIF - even Sarmiento's paper lost over 30% to follow-up, I am assuming they were also unable to tolerate the brace or had unacceptably bad angular deformities or gross displacement uncontrolled by the brace - regardless, it was unclear from the manuscript how this issue would be addressed - could the authors please clarify

Authors' comment 22:
Thank you for the remark. The paragraph "Safety considerations" is dealing with these matters. In our manuscript, we describe that we consider the fracture a non-union if after 26 weeks it does not show clinical and/or radiological consolidation. We have also tried to emphasize our approach to radial nerve palsy but we find it very difficult to have a predefined protocol to all serious or minor adverse event. The main principle is that if there seems to be a treatment failure, all the necessary interventions will be done to help the patient get better.

Authors' action 22:
Prompted by this valuable comment, we added a short paragraph after the non-union paragraph to clarify our reaction to certain treatment failures as follows:
If at any point an imminent problem in healing is noticed warranting a change in the treatment regimen this will be done with the discretion of the treating physician regardless of the initial treatment allocation. This will include, but is not limited to, deep infection, fracture threatening the skin integrity, malunion causing subjective problems and a re-fracture.

Please find below our responses to the valuable comments received from Joao Belloti:

Reviewer comment 1:
My main concern about this protocol is regarding the inclusion of patients with primary radial nerve injury, since this subgroup of patients has a very different clinical recovery from patients without nerve damage. I suggest that this subgroup of patients be analyzed in a parallel study.

Authors' response 1:
Thank you for this thoughtful comment. We think that according to current literature humeral shaft fracture with primary radial nerve palsy is not an absolute indication for surgical approach. We agree, that radial nerve palsy has an effect on the recovery and this is the reason we are stratifying the patients according to radial nerve status to spread the patients with palsy evenly between the two groups. Moreover, randomisation stratification factors will be included as covariates in the statistical analyses.

Reviewer comment 2:
I recommend that the authors review the description and classification of potential adverse events, since non-union and deep infection are not considered minor adverse events in the treatment of these fractures. There is no description of how adverse events will be accessed (SAEs and MAEs)

Authors' response 2:
Thank you for this very valuable comment. We agree that deep infection is a potentially severe adverse event. We consider secondary radial palsy as severe adverse event only when it is permanent. Most of the palsies recover within 3 months and we consider these recovering palsies minor adverse events. Yet we find it very difficult to have a predefined protocol to all serious or minor adverse event. The main principle is that if there seems to be a treatment failure, all the necessary interventions will be done to help the patient get better.

Authors' action 2:
We changed deep infection to SAEs. After this valuable comment, we added a sentence to the end of the first paragraph of our safety considerations section as follows:
Adverse effects will be registered from the medical records and during follow-up visits. All adverse events are treated in study hospitals by or under the supervision of an experienced orthopaedic trauma surgeon.

We feel these clarifications and corrections were very important and that the changes have clearly improved our paper. We hope that the revised manuscript now fulfils satisfactorily your expectations and will be accepted by the BMJ Open. We look forward to hearing from you soon and will be pleased to discuss our manuscript again if needed.

VERSION 2 – REVIEW

REVIEWER	Kevin Tetsworth MD FRACS Royal Brisbane Hospital Department of Orthopaedic Surgery Brisbane, QLD, Australia
REVIEW RETURNED	05-Mar-2017

GENERAL COMMENTS	I would like to first thank the authors for carefully addressing many of the issues identified when I originally reviewed this manuscript. Most, if not all, of the changes suggested have been implemented, and it is a better manuscript. I have two remaining concerns: (1) the use of the term "gold standard" when referring to ORIF on lines 118-19 and again on 560 - the point being there are many surgeons who believe IMN is equally good, and there is a rapidly growing body of evidence that tells us quite convincingly that MIPO humerus plating is better than ORIF or IMN (over 40 published studies support MIPO over ORIF/IMN) - you have chosen to almost completely ignore humeral MIPO, and that is a good choice for this protocol - it would be best if you simply state there is no clear advantage to any of the other available methods of surgical stabilization, and that ORIF provides clinical outcomes arguably as good as or better than the alternatives (2) I have a concern regarding your sample size calculation, and perhaps this will need to be adjudicated by the editors or an expert

	statistician - We know that loss to follow up for humerus fractures is very close to 30% (Sarmiento and others) - this means you will need 50 participants per group - of more concern, Wang, et al (2015) completed an a priori analysis that indicated 179 patients per group would be necessary to have adequate power to resolve a difference between ORIF vs MIPO for humeral shaft fractures - in my opinion, Wang et al have over estimated the sample size, and you have underestimated the sample size

REVIEWER	Joao C Belloti Universidade Federal de Sao Paulo Brazil
REVIEW RETURNED	01-Mar-2017

GENERAL COMMENTS	Regarding my first comment, I maintain the recommendation to the authors not to include patients with primary radial nerve injury. " there is no absolute indication of a surgical approach for fractures of the humeral shaft with primary radial nerve injury", however, in high-energy closed fractures, there is indication of the early surgical approach of the nerve. In addition, the low incidence of primary radial nerve injury (11- 16%) makes the sample too small and insufficient to evaluate the effectiveness of the two interventions in the subgroups (with and without primary radial nerve injury). Regarding my second comment, I think the authors confused non-union with the primary radial nerve palsy. I recommend that the non-union be categorized to serious adverse events (SAEs). I recommend including the following references: 1- Rocchi M, Tarallo L, Mugnai R, Adani R. Humerus shaft fracture complicated by radial nerve palsy: Is surgical exploration necessary? Musculoskelet Surg. 2016 Dec;100(Suppl 1):53-60. Review. PubMed PMID: 27900704 2- Li Y, Ning G, Wu Q, Wu Q, Li Y, Feng S. Review of literature of radial nerve injuries associated with humeral fractures- an integrated management strategy. PLoS One. 2013 Nov 8;8(11):e78576. doi: 10.1371/journal.pone.0078576.

VERSION 2 – AUTHOR RESPONSE

We thank Kevin Tetsworth for the careful review of our paper and appreciate the insightful suggestions for improvement.

Reviewer comment 1:
The use of the term "gold standard" when referring to ORIF on lines 118-19 and again on 560 - the point being there are many surgeons who believe IMN is equally good, and there is a rapidly growing body of evidence that tells us quite convincingly that MIPO humerus plating is better than ORIF or IMN (over 40 published studies support MIPO over ORIF/IMN) - you have chosen to almost

completely ignore humeral MIPO, and that is a good choice for this protocol - it would be best if you simply state there is no clear advantage to any of the other available methods of surgical stabilization, and that ORIF provides clinical outcomes arguably as good as or better than the alternatives

Authors' response 1:
Thank you for this remark. We agree that there is a variation in surgical treatment modalities and current literature warrants the use of either ORIF, IMN or MIPO. In our view, the current literature supports "noninferiority" of these different modalities compared to each other, but possible superiority is an open question. (Heineman et al, Acta Orthop 2010 and 2012 and Kim et al, J Orthop Trauma 2015) Accordingly, we will refrain from using the term "gold standard" in our manuscript. However, in Finland, ORIF is by far the most widely used surgical treatment method for this type of humeral fractures and for which the competence of our surgeons is high. We want to use ORIF as the surgical method in our study therefore as we believe that we will obtain the best results by choosing a method which the surgeons in our study units are familiar with. Introducing a new method, for example MIPO, would have confounded the results by learning curve effects.

Authors' action 1:
We revised the mentioned sentences as follows:
Lines 117-120.
According to current literature open reduction and internal fixation (ORIF) with plating gives comparable or better results in surgical fracture treatment of humeral shaft fractures compared to other surgical treatment options.

Lines 565-567:
We chose the ORIF with plating as the surgical method in this trial since it is the most widely used surgical treatment method of humeral shaft fractures in Finland.

We also added two references to MIPO papers with randomised setting.

Reviewer comment 2:
I have a concern regarding your sample size calculation, and perhaps this will need to be adjudicated by the editors or an expert statistician - We know that loss to follow up for humerus fractures is very close to 30% (Sarmiento and others) - this means you will need 50 participants per group - of more concern, Wang, et al (2015) completed an a priori analysis that indicated 179 patients per group would be necessary to have adequate power to resolve a difference between ORIF vs MIPO for humeral shaft fractures - in my opinion, Wang et al have overestimated the sample size, and you have underestimated the sample size

Authors' response 2:
Thank you for this remark. We have already run the trial since late 2012 and already over 50 patients have participated in the 1 year follow-up. So far, our loss to follow up is 6% (3 patients). Moreover, since the recruiting units are responsible for the treatment of this kind of trauma in their regions, almost all the patients have come to all follow up visits. Also, loss to follow-up has been remarkably low in other prospective orthopaedic trials run in Finland and in trials investigating surgical treatment of humerus fractures (Dai et al. J Orthop Sci 2013).

Author Action 2:
We repeated the sample size calculation. With 80% power and alpha level of 0,05 the MCID of 10 points in DASH with the assumption of 14,7 points standard deviation as stated in the paragraph Sample Size, 35 patients per group is needed. Assuming 12,5% loss to follow up we should have 40 patients per group.

Please find below our responses to the valuable comments received from Joao Belloti:

Reviewer comment 1:
Regarding my first comment, I maintain the recommendation to the authors not to include patients with primary radial nerve injury. “there is no absolute indication of a surgical approach for fractures of the humeral shaft with primary radial nerve injury”, however, in high-energy closed fractures, there is indication of the early surgical approach of the nerve. In addition, the low incidence of primary radial nerve injury (11-16%) makes the sample too small and insufficient to evaluate the effectiveness of the two interventions in the subgroups (with and without primary radial nerve injury).

Authors’ response 1:
Thank you for this thoughtful comment. We fully agree that a patient with radial nerve palsy due to open fracture, crush injury with severe soft tissue damage or sharp penetrating injury should be operated. These patients are excluded from our trial. Generally speaking, in the Finnish society we very seldom see open fractures or solitary unilateral humeral shaft fractures caused by high-energy injury mechanism. We started our trial already in 2012 and have recruited 70 patients of the 80 patients so far: there has not been a single case with solitary humeral shaft fracture caused by high energy injury. All the patients with high energy trauma have had other injuries, resulting as an exclusion of the patient from our randomised trial. The current literature (as stated by the suggested reference Li et al, PLoS One 2013), however, does not support exploration of the radial nerve in closed, low energy fractures.

The stratification based on primary radial nerve palsy is used to ensure even distribution of these patients to the two treatment groups. Our intention is to analyse the results within the whole treatment group, however, using the stratification variables as factors in the statistical model. In other words, we are not intending to perform subgroup analysis by radial nerve palsy between operated and conservatively treated patients. Neither are we intending to compare patients with and without radial nerve palsy, regardless of treatment method.

We expect to have about 5-10 patients with radial nerve palsy in our trial, of which we expect 80-90% to heal uneventfully. Thus, there is a possibility of having one to two patients with a need for nerve grafting or tendon transfer and having significant deficit in activities of daily living. We expect the rest of the patients with primary radial nerve palsy to have similar scores in the primary outcome measures at 12 months with the patients without radial nerve palsy.

Authors’ action 1:
We will maintain a stratification according to radial nerve status after the injury to ensure even distribution of patients to both treatment groups.

We changed the sentences regarding inclusion of patients with radial nerve palsy on lines 289-291 as follows:
Patients having a primary radial nerve palsy are also included in the study. Current literature suggests that operative exploration is not mandatory in primary radial nerve palsy in association with closed low-energy humeral shaft fracture.

Reviewer comment 2:
Regarding my second comment, I think the authors confused non- union with the primary radial nerve palsy. I recommend that the non-union be categorized to serious adverse events (SAEs).

Authors’ response 2:
Thank you for this comment. We have given this suggestion a lot of thought and we unfortunately must disagree. Based on the literature and our experience, we consider a non-union of humeral shaft

fracture as an undesirable outcome of treatment in 5-20% of conservatively treated patients rather than as a serious adverse event. Also, there is 7,5% possibility of non-union in surgically treated patients (meta-analysis by Heineman et al, Acta Orthop 2010). In our patients, non-union is operated at 6 months if the patient is willing to accept the risks of an operation. We expect the risks of an operation for non-union to be a bit higher compared to the risks for operating fresh fractures mainly because a possibility of an iatrogenic radial nerve lesion in the former. The time in hospital is practically the same with the patients treated because of non-union and for patients operated directly after the trauma. We expect no difference in patient reported outcome measures at 12 months between patients operated in acute phase or patients operated at 6 months because of a non-union. Also, compared to events listed as SAEs, even a persisting pseudoarthrosis of the humerus is clinically a much more benign situation, discouraging us from classifying a non-union as a SAE.

Authors' action 2:
We are keeping non-unions in minor adverse events.

Reviewer suggestion 3:
I recommend including the following references:
1- Rocchi M, Tarallo L, Mugnai R, Adani R. Humerus shaft fracture complicated by radial nerve palsy: Is surgical exploration necessary? Musculoskelet Surg. 2016 Dec;100(Suppl 1):53-60. Review. PubMed PMID: 27900704
2- Li Y, Ning G, Wu Q, Wu Q, Li Y, Feng S. Review of literature of radial nerve injuries associated with humeral fractures-an integrated management strategy. PLoS One. 2013 Nov 8;8(11):e78576. doi: 10.1371/journal.pone.0078576.

Authors' response 3:
Thank you pointing out these valuable references.

Authors' action 3:
We included these references to the paragraph 'Safety considerations'.

We feel these clarifications and corrections were once again very important and that the changes have improved our paper. We hope that the revised manuscript now fulfils satisfactorily your expectations and will be accepted by the BMJ Open. We look forward to hearing from you soon and will be pleased to discuss our manuscript again if needed.

VERSION 3 – REVIEW

REVIEWER	Kevin Tetsworth, MD FRACS Royal Brisbane Hospital Australia
REVIEW RETURNED	03-Apr-2017

GENERAL COMMENTS	Well done - thank you for addressing the various issues and conditions previously identified - these have all been addressed to my satisfaction - I look forward to seeing the results of this study when completed!

REVIEWER	Joao C Belloti UNIVERSIDADE FEDERAL DE SÃO PAULO (UNIFESP) ESCOLA PAULISTA DE MEDICINA
REVIEW RETURNED	10-Apr-2017

GENERAL COMMENTS	I was disappointed to hear that the authors have started the study in 2012 and have already included 70 of the 80 patients. Patients with diaphyseal fractures of the humerus with associated radial nerve injury have a completely different clinical evolution, when compared to ones with no nerve injury. The sample size therefore does not allow proper evaluation of outcomes of these two different groups. The primary objective of humeral fracture treatment is to obtain consolidation, with no significant deformities and with minimal functional limitation. The authors consider nonunion as a minor adverse effect. This condition is actually a failure of the treatment and therefore cannot be seen as a minor adverse effect. We had previously made these same considerations to the authors, but they unfortunately didn't agree.