PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Osseointegrated Reconstruction and Rehabilitation of Transtibial Amputees: the Osseointegration Group of Australia surgical technique and protocol for a prospective cohort study
AUTHORS	Haque, Russel; Al-Jawazneh, Shakib; Hoellwarth, Jason; Akhtar, Muhammad Adeel; Doshi, Karan; Tan, Yao Chang; Lu, William; Roberts, Claudia; Al Muderis, Munjed

VERSION 1 – REVIEW

REVIEWER	Prof.dr. Maria Nijhuis-van der Sanden,PT,PhD e-mail: Ria.Nijhuis-vanderSanden@radboudumc.nl Radboud university medical center Radboud Institute for Health Sciences, IQ healthcare 114 Department of Rehabilitation P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
REVIEW RETURNED	01-May-2020

GENERAL COMMENTS	This is a relevant and interesting study. However, there are a number of methodological flaws (or flaws in the description of the methodology) that need to be addressed to increase the transparency of the study and to be able to interpret the future results. I used the STROBE checklist to point what needs to be described more in detail to be able to replicate the study. Especially the statistical analysis needs to be improved.

REVIEWER	Patrick Lin MD Anderson Cancer Center USA
REVIEW RETURNED	08-May-2020

GENERAL COMMENTS	The authors present their surgical technique and protocol for transtibial osseointegration. Since the tibia has genuinely unique anatomic considerations that are distinct from the femur, the study merits serious consideration. One cannot assume that the good results obtained in the femur can translate directly to the tibia. I have a few suggestions & comments regarding this manuscript: 1. The study began in 2014 and may end in 2022. It is relatively late to be publishing the protocol for the participating centers to ensure uniformity of technique. It may be helpful to have a sense of current accrual and remaining accrual. A power analysis for the cohort size was briefly alluded to but not described in detail. 2. Abstract p.5 - "The study has a relatively short follow-up period of

	2 years..." I believe this may be not such a bad concern, since some patients will have much more than 2 years. In fact, the range will likely be 2-10 years. Please clarify. 3. Introduction - It would be helpful to describe the soft tissue techniques of other surgeons to highlight the differences and to stress the key aspects of the authors' own technique. There is controversy as to how best to handle the soft tissues around the abutment. 4. Outcome p. 11 - "Functional outcome measures and conventional radiographs were also taken at baseline as well as at 12, 24 and yearly follow-up..." There is no unit of measure after 24. Does this refer to weeks? 5. Adverse events p. 11 - Infections that do not require hospitalization would ideally be recorded as well, particularly those involving the skin at the interface. Readers like myself would be interested in knowing how often patients are treated with oral antibiotics for what seem like minor skin irritations.

VERSION 1 – AUTHOR RESPONSE

Response to Reviewer 1.

Thank you for reviewing the manuscript in great detail. We really appreciate the suggestions made and have done our best to make necessary changes to the manuscript. Response to specific suggestions are tabulated below

In the abstract the background information is quiet long, while the methods and analysis lacks information on the start and endpoint of the study, the design, measurement timepoints, outcome measures and statistical analysis. It should be clear which primary and secondary outcome measures will be used. Moreover, it would be nice to define which outcomes are perceived as clinically relevant outcomes. Which outcomes need to be reached to accept this surgery and rehabilitation strategy as better than the common socket prothesis	Changes to the abstract made as per suggestions: details about start and endpoint of the study, the design, measurement timepoints, outcome measures and statistical analysis have been added. To identify factors affecting outcomes following added: “Multivariable multilevel logistic regression will be performed with a focus to identify factors associated with outcomes and adverse events, specifically infection, periprosthetic fracture, implant fracture, and aseptic loosening”
In the strengths and limitations section statements are formulated as already found positive outcomes for instance statement 2: this study would not only underline the feasibility of OI but also… It would be nice to formulate the statements as potential outcomes, this is a research project focussed on getting insight, not on already existing results	The statement has been re-phrased as suggested: “The findings of the study would assess whether osseointegration in transtibial amputees is feasible in terms of risks and benefits”
In the introduction section authors present the rationale of the study I did miss an overview	In addition to skin problems, problems related to fit, pistoning, proprioception, loss of

of al clinical complaints of patients with amputations. They only focus on the skin problems but these are not the only ones (see also review Leijendekkers et al., Disabil Rehabil. 2017 Jun;39(11):1045-1058, Disabil Rehabil. 2018 Jul;40(14):1732).	balance as well as asymmetrical gait leading to pain in different areas (from the referred paper) has been mentioned as "These include skin problems such as infections, and skin breakdown due to chronic irritation and thermal injury,8-11 mechanical problems such as suboptimal fit, pain and pistoning12 and lastly, problems with proprioception that leads to loss of balance and falling.13 Gait with a TSP has been found to be asymmetrical correlating with a weakness in the hip abductor muscles, which can explain the back pain and pain in other regions experienced by such users including ipsilateral and contralateral limb, buttocks, neck and shoulder.14 Socket prostheses users account for their poor quality of life mostly to physical disability, pain and decreased energy levels.5, 15"
In the objectives the authors describe that they want to describe and compare the outcomes of this study with preoperative data and TFA outcomes. However, in the statistical analysis it is not clearly described which statistical analyses will be used to compare…. Moreover, as the authors describe, it seems that there are two subgroups included so it seems to be logical toad subgroup analyses. Moreover, this is a longitudinal design with a large number of outcome measures but also a number of potential influencing factors like age, physical status presurgery, adherence to the rehab protocol etc. It has a lot of added value if this study with 4 measurement points will be used to get insight in influencing factors on the outcome? Which individuals benefit most (or least) of this new technique? Please formulate hypotheses about the expected outcomes? Which group will benefit most? Which factors will have a negative or positive influence on the functional outcomes or quality of life? Or which patients are the omst at risk for failure? Please add a statistician to the research group and think about more advanced analysis methods like logistic regression or multilevel analyses. This will be a large cohort of patients (n=100) so we can learn a lot more than only describe the outcomes for the total included patient group.	Major changes to Study Objectives and Statistical Analysis section has been made. In the study objectives section following have been added: "One of the primary objectives of this study is to identify the individual patient characteristics or factors that have a positive or negative influence in the outcomes mentioned above. This analysis in a regression model would help to identify the patients based on their characteristics who would be most or least benefitted with this novel procedure and who would be at a higher or lower risk of failure. The other question that is study will identify is the rate of additional surgical interventions as well as to identify factors associated with further surgery, specifically for infection, periprosthetic fracture, implant fracture, and aseptic loosening" Statistical analyses section has been completely re-written. (See below)
The key elements are the setting, the measurement time points and the primary outcome	Specific paragraph related to setting has been added. Details about data sampled, at

	which timepoints (T0, T1, T2, T3, T4…) and specific outcomes have been explicitly mentioned now. A Table has been added to clearly mention the parameters sampled.
Please start the methods section with the design of the study and the time point of measurement. The setting is not clearly described. The start and end of the study should be described at the start of the methods section. The inclusion started in 2014, so already 6 years!! Ago. Then it is common to describe the number of included patients at the moment this article is sent in and to describe how many patients expected to treat each year. Moreover, the authors describe that already a pilot was done, are these patients also included in the cohort? How many patients are referred and how many(%) receive surgery?	Method section started with design and time point of measurement. Setting has been added as separate paragraph. Details about patient accrual mentioned. The patients of pilot study have not been included in this study due to lack of standard protocol. (mentioned in text)
I do not understand the statement that patients are not involved in the study design etc. Mostly this is advised to be sure that the most relevant outcomes are measured. Moreover, participating patients need to be informed that the data are used for publication. Please, explain.. The methods section is written in the past tense, although this is an ongoing study this is rather confusing	Patient and Public involvement section is a mandatory requirement of BMJ open to be mentioned in the methods section to underline of they had any influence on the study design. All participating patients have been properly counselled about participation in the study and all have given written informed consent. The tense has been uniformly made to future.
Moreover, it would be nice to follow the most common way to describe a study, first design, participants (included the number that is planned), the current state, the inclusion and exclusion criteria, the recruitment, referral and inclusion procedure, the intervention protocol (in this case surgery and rehabilitation procedure), data sampling and measurement procedure, measurement instruments and data management, statistical analysis per research question and in the discussion the rationale for choices made with pro's and con's.	Changes have been made to follow the suggested outline.
It seems that patients are included in two steps first a selection by phone and afterwards a selection by screening. Who is involved in both procedures? How is guaranteed that no selection bias occurs. Is a checklist or systematic procedure used? Are data sampled to be able to fill the flowchart at the end of the study (inclusion and exclusion	Details about setting and patient selection has been explicitly discussed now. Inclusion and Exclusion criteria along-with reasons have been added as a table. Potential for selection bias discussed as separate paragraph. Another paragraph added later to underline the steps taken to

with reasons) e.g. demography, medical, prosthetic, how is compliance determined? Which psychological and pain issues are involved and measured? And how are expectations discussed? The multidisciplinary team involved is quiet large (I miss the physiotherapist?) Please describe each role in clinical practice and/or the research. For instance how many people are involved in the surgery? In the training? In the measurements? Which radiological decision rules? What is the time between enrolment and surgery for each patient? I miss the methods for follow-up: how are the measurements arranged? And again please provide the status at this moment. How many patients are enrolled? How many are measured at t1-t2-t3 and t4. Please provide information on the preoperative data both regarding decision-making for inclusion and for potential influencing factors (eg relationship between DEXA outcomes and the rehab outcomes or complications). *Which data are sampled by the physio and what training program was described? This is not transparent but relevant for the outcomes.*	reduce the risk of bias. Timeframe for surgery discussed separately. Multidisciplinary team has been described. Which parameters are measured, by whom and at which time points has been explicitly mentioned now. Present status of enrolment has been mentioned. All the potential influencing factors are being recorded in the database and at the time of statistical analyses influence of each factor on the outcomes in a regression model will be recorded and reported, as will be the coefficient of relative influence of each factor.
As far as I understand the cohort contains 2 subgroups please think about the consequences for the study analyses. *Page 9 line 12 and further: this is confusing: it is not clear if this technique is used in the study population or the described experience was before the inclusion. If so, then it would be more clear to put this previous experience in the introduction (so the rational for the study is determined by theoretical study findings and previous clinical expert experience). It is more clear to describe in the methods section only the implants which are used. Which surgical information will be sampled and reported in the study as potential influencing variables?* *In contrast to the surgery, the rehabilitation protocol is poorly described related to the content, the frequency and the dose. How is home based physiotherapy arranged and is this a structured protocol or not? Which decision rules are used? (the rehab adherence will be of influence on the outcomes). See for instance Leijendekkers et al, . Physiother Theory Pract. 2017;33(2):147-161.*	The details about previous experience has been removed from implants section. Surgical information sampled has been mentioned. Rehabilitation Protocol and Physiotherapy protocol has been described in details with help of 2 flowcharts.

Are there instructions for the PT at home? Please provide data sampling management: which data are sampled at each point (pre, post, 12 and 24 months) by whom?	
In the statistical analysis only a pre-post comparison is escribed, however not using all data points leads to lost of power and information. Moreover, as described before, it seems to be logical to find out if there are differences between subgroups and to analyse influencing factors. I do not know if all outcomes are sampled at all time points please provide a clear overview. Why did the authors not sample data on pain? This because pain seems to be a determining factor in functional recovery.	Statistical Analysis section has been completely changed according to suggestions: "The primary questions this study aim to identify are 1. the individual patient characteristics or factors that have a positive or negative influence in the outcomes measured or in other words who would be most or least benefitted with this novel procedure and who would be at a higher or lower risk of failure? and 2. what are the rates of additional intervention for patients undergoing transtibial osseointegration, and for what reasons? This project will also aim to collect data which can allow investigation of diverse questions regarding transtibial osseointegration as further insight develops. Multivariable logistic regression will be performed with a focus to identify factors associated with further surgery, specifically for infection, periprosthetic fracture, implant fracture, and aseptic loosening. Additionally, factors associated with Daily prosthesis wear hours, Prosthetic wear satisfaction, SF-36 and mobility (6MWT, TUG, K level) will be evaluated. Separate regression models will be developed for short and long residuum TTOIs as well. A p value of 0.05 will be the cutoff of significance. The p value for each regression identifying significant predictors of dependent variable outcome will be reported, as will the coefficients of relative influence of each variable. The pre- versus post-operative continuous value data will be presented as mean and standard deviation and compared with Student's T-test or analysis of variance (ANOVA) if the data is normally distributed. Should the data not be normally distributed the median and interquartile ranges will be reported and comparison made using Wilcoxon test. For comparison of qualitative variables such as gender, laterality, or reason for amputation, frequency comparison will be performed using Chi-squared test or Fisher's Exact test, depending on the actual occurrence of each variable. P=0.05 will be considered statistically significant."
risk on bias in the inclusion procedure, measurement methods and differences in intervention	Reducing risk of bias section added:" In addition to reducing the risk of selection bias as described above, bias relating to surgeon expertise and protocol adherence is

	eliminated since all operations will be performed by a single primary surgeon. Bias related to data collection will be minimized by employing dedicated research assistants who will be unaware about details of patient demographic characteristics, surgical and implant details and previous recorded scores. Further, the results of functional outcome measures (6MWT, TUG, K-levels) depend on the patients' actual performance, while the results of subjective outcome measures are completely patient reported from surveys. In addition, the assessors will not be involved in data analysis"
The power analysis (n=100) is not clearly explained. Which outcome variable was used to calculate the power?	Power analysis details have been added: "Preliminary data and clinical experience has been obtained from an initial pilot study comprising 10 patients owing to absence of prior literature. Software G* Power was used to calculate an a priori sample size. Considering SF-36 physical component score as primary outcome measure, the pre-operative and 2 year post-operative scores were recorded. Comparing the means (37.62 and 44.83) and SDs (11.8 and 19.5) of these 2 groups respectively using Wilcoxon test, the effect size was calculated to be 0.36 and sample size was calculated to be 87 assuming α error to be 0.05 and in order to achieve a Power of 95 %. Considering a drop-out rate of 20%, a final sample size of 109 was decided upon. None of the patients of the pilot study have been included in this study due to absence of standard protocol."

Response to Reviewer 2.

Thank you for reviewing the manuscript in great detail. We really appreciate the suggestions made and have done our best to make necessary changes to the manuscript. Response to specific suggestions are tabulated below

1. The study began in 2014 and may end in 2022. It is relatively late to be publishing the protocol for the participating centers to ensure uniformity of technique. It may be helpful to have a sense of current accrual and remaining accrual. A power analysis for the cohort size was briefly alluded to but not described in detail.	Although the first patient was enrolled in 2014, in the initial few years, there were only few cases given that it was a relatively new procedure. It is only recently that more patients are getting enrolled and so description of surgical details, study parameters and rehabilitation protocol will benefit not only us but also other groups planning to start transtibial osseointegration surgery. Current accrual details (68 patients already enrolled) have been mentioned in manuscript. Power and Sample size analysis has been described in manuscript: "Preliminary data and clinical experience has been obtained from an initial pilot study comprising 10 patients owing to absence of

	prior literature. Software G* Power was used to calculate an a priori sample size. Considering SF-36 physical component score as primary outcome measure, the pre-operative and 2 year post-operative scores were recorded. Comparing the means (37.62 and 44.83) and SDs (11.8 and 19.5) of these 2 groups respectively using Wilcoxon test, the effect size was calculated to be 0.36 and sample size was calculated to be 87 assuming α error to be 0.05 and in order to achieve a Power of 95 %. Considering a drop-out rate of 20%, a final sample size of 109 was decided upon. None of the patients of the pilot study have been included in this study due to absence of standard protocol."
2. Abstract p.5 - "The study has a relatively short follow-up period of 2 years..." I believe this may be not such a bad concern, since some patients will have much more than 2 years. In fact, the range will likely be 2-10 years. Please clarify.	Correctly pointed out. Have rephrased the statement. This study has minimum 2 year follow-up, but it would range from 2-8 years with average follow-up time to be much higher.
3. Introduction - It would be helpful to describe the soft tissue techniques of other surgeons to highlight the differences and to stress the key aspects of the authors' own technique. There is controversy as to how best to handle the soft tissues around the abutment.	As suggested, mention has been made in the surgical technique section about the unique way we handle the periosteum.
4. Outcome p. 11 - "Functional outcome measures and conventional radiographs were also taken at baseline as well as at 12, 24 and yearly follow-up..." There is no unit of measure after 24. Does this refer to weeks?	Changes have been made in the manuscript and these time-points for follow-up have now been mentioned as "3, 6 and 12 months and yearly follow-ups thereafter"
5. Adverse events p. 11 - Infections that do not require hospitalization would ideally be recorded as well, particularly those involving the skin at the interface. Readers like myself would be interested in knowing how often patients are treated with oral antibiotics for what seem like minor skin irritations.	Infections requiring oral or intravenous antibiotics has been aimed to be recorded. Although it is challenging to record oral antibiotics when it has been prescribed by local GPs, we aim to keep GPs office informed about the same, when we enrol a new patient for the osseointegration procedure. Also the patients are informed at the outset that they need to let us know whenever having antibiotics for local stoma infections. We also try to record administration of oral antibiotics with help of survey forms when the patients come for follow-up and try to cross-check from GPs records. In case of any mismatch, we further investigate. All data that is sampled along with the time points of sampling has been made into a table for better illustration.

VERSION 2 – REVIEW

REVIEWER	M.W.G. Nijhuis-van der Sanden, PT,PhD Radboud university medical center Research Institute for Health Sciences Department of IQ Healthcare, Nijmegen, The Netherlands
REVIEW RETURNED	30-Jun-2020

GENERAL COMMENTS	The authors did a great job and improved the manuscript substantially. There are some small remarks: page 6 line 39-41. I understand what the authors want to say but please restructure the sentence: The other objective is to identify…..etc, In table 1 It is stated that patients can choose for a osteointegration, however in the same table as exclusion criterium it is mentioned that having no mobility, socket or skin problems is an exclusion criterium: this would mean that "new" amputees have more possibilities to choose than those who want to change the socket prothesis ? Does this mean that each amputee without contra-indications will be offered the choice: socket or osseo? This seems to be the case as described in the setting and screening part. However, how are patients informed about possible complications and long-term outcomes? This because going back to a socket prothesis after osseo is not a simple solution… So an ethical sound shared decision making process needs to be present with clear discussion on the different scenario's including risks and benefits… Potential selection bias: the selection of high income patients will influence the outcomes: we know that health status is higher, the comorbidities lower, and the adherence to the rehab program higher in this group. This means that the outcomes of the study are not generalizable to all countries and all populations. Data analysis section: please mention that a p-value of ≤ 0.05 will be considered as significant in stead of = The pre- versus post-operative data will be compared with a T-test or ANOVA: why not post-hoc analyses related to the longitudinal data analysis over the measurements at t0, t1 etc…? The authors describe a comparison of qualitative variables like gender etc? A comparison with what? The factors mentioned are potential independent variables influencing outcomes as expressed in the dependent variables like pain, functional capacity or negative outcomes like fractures or infection etc.

VERSION 2 – AUTHOR RESPONSE

Response to Reviewer 1.

Thank you for reviewing the manuscript in great detail. We really appreciate the suggestions made and have done our best to make necessary changes to the manuscript. Response to specific suggestions are tabulated below

The authors did a great job and improved the manuscript substantially. *There are some small remarks: page 6 line*	Thank you for the appreciation. The sentence has been restructured as advised:

39-41. I understand what the authors want to say but please restructure the sentence: The other objective is to identify…..etc,	"The other objective is to identify the rate of additional surgical interventions as well as to identify factors associated with further surgery, specifically for infection, periprosthetic fracture, implant fracture, and aseptic loosening."
In table 1 It is stated that patients can choose for a osteointegration, however in the same table as exclusion criterium it is mentioned that having no mobility, socket or skin problems is an exclusion criterium: this would mean that "new" amputees have more possibilities to choose than those who want to change the socket prothesis ? Does this mean that each amputee without contra-indications will be offered the choice: socket or osseo? This seems to be the case as described in the setting and screening part. However, how are patients informed about possible complications and long-term outcomes? This because going back to a socket prothesis after osseo is not a simple solution… So an ethical sound shared decision making process needs to be present with clear discussion on the different scenario's including risks and benefits…	Thank you for pointing out the mistake in Table 1. From the exclusion criteria, the point mentioning exclusion of amputee with no mobility, socket, or skin problems has been removed as all amputees do have a choice, as has been correctly pointed out. As with any surgeon and surgery, counselling is complex. As the Reviewer has identified, patient counselling is indeed a shared decision making process. All patients are presented with the positive aspects as well as the risks of osseointegration. It does include a dynamic assessment and discussion of the benefits (mobility, lifestyle, etc) as well as the risks (infection, fracture, further surgery including full removal or further amputation) as discussed in this article and others. We do explain to patients the relative novelty of this surgery and the impact that it has on fully understanding the immediate and long term risk/benefit profile. Following has been added to the patient recruitment section "All patients are counselled extensively by the team which includes a dynamic assessment and discussion of the benefits (mobility, quality of life, etc) as well as the risks (infection, fracture, further surgery including full removal or further amputation, etc) of osseointegration. The patients are fully explained about the relative novelty of this surgery and that the immediate and long term risk/benefit profile is still not very well defined so that an ethical, sound and shared decision making process is achieved."
Potential selection bias: the selection of high income patients will influence the outcomes: we know that health status is higher, the comorbidities lower, and the adherence to the rehab program higher in this group. This means that the outcomes of the study are not generalizable to all countries and all populations.	This is indeed a limitation of this study, and has been identified separately in patient recruitment section. However, we will document the co-morbidities, adherence to rehab program, etc for each patient and analyse if and how they affect the outcomes. The following has been added to the patient recruitment section with heading Potential Selection Bias: "One of the limitations of this study is possibility of selection bias to exclude low

	income patients. Osseointegration is an expensive surgery and thus is not covered by the standard government insurance for our country. It is covered by more premium insurance plans. Thus we counsel patients to enrol in these top level insurance plans so that not only will the surgery itself be provided but any additional surgery for an adverse event will be covered, so long as they maintain their coverage. Due to this limitation the results of the study may not be generalizable to all countries and all populations."
Data analysis section: please mention that a p-value of ≤ 0.05 will be considered as significant in stead of = *The pre- versus post-operative data will be compared with a T-test or ANOVA: why not post-hoc analyses related to the longitudinal data analysis over the measurements at t0, t1 etc…?*	Thank you for pointing the mistake with p value. It has been corrected. An additional line mentioning about post-hoc analyses has also been added.
The authors describe a comparison of qualitative variables like gender etc? A comparison with what? The factors mentioned are potential independent variables influencing outcomes as expressed in the dependent variables like pain, functional capacity or negative outcomes like fractures or infection etc.	Thank you for pointing out the ambiguity caused by the word comparison. The section has been re-structured as following: "The influence of various factors such as patient gender, age, and cause of amputation on dependent variables relating to potential risks (infection, fracture, further surgery, etc) or benefit (mobility, QOL outcomes, etc) will be assessed. Multivariable logistic regression will be performed with a focus to identify factors associated with further surgery, specifically for infection, periprosthetic fracture, implant fracture, and aseptic loosening. Additionally, factors associated with Daily prosthesis wear hours, Prosthetic wear satisfaction, SF-36 and mobility (6MWT, TUG, K level) will be evaluated. Separate regression models will be developed for short and long residuum TTOIs as well. A p value of ≤ 0.05 will be considered as significant."