Author’s response to reviews

Title: Accuracy of a non-invasive CT-based measuring technique for cement penetration depth in human tibial UKA.

Authors:

Christian Scheele (C-Scheele@web.de)

Peter Müller (peter.mueller@med.uni-muenchen.de)

Christian Schröder (christian.schroeder@med.uni-muenchen.de)

Thomas Grupp (thomas.grupp@aesculap.de)

Volkmar Jansson (Volkmar.Jansson@med.uni-muenchen.de)

Matthias Pietschmann (Matthias.Pietschmann@med.uni-muenchen.de)

Version: 1 Date: 01 Nov 2018

Author’s response to reviews:

*** Please also see attached file with identical content ***

Covering letter – Resubmission of revised form of BMIM-D-18-00114

Title: Accuracy of a non-invasive CT-based measuring technique for cement penetration depth in human tibial UKA.

Dear Dirk Krüger,

thank you very much for considering our paper as appropriate for the journal. We highly appreciate your interest in our work. As an aspiring orthopedic surgeon, I’m humbled and honored to be given the possibility to resubmit our revised paper.
In my opinion, pointing out the areas in need of clarification by the reviewers gave us the chance to add value to the work and inspired us for further projects (MRI scans, automated protocols, etc.)

Please see below our responses to the editor/reviewers’ comments on a point-by-point basis:

A) Editors’ comments:

1. **Author contributions:** Please use only initials.

Names were changed to initials in the section author contribution (declarations section, lines 26-33, page 10)

2. **Ethics approval:** This section should be named "Ethics approval and consent to participate"

Section was renamed to "Ethics approval and consent to participate" (declarations section, line 2, page 10)

3. **Rephrase your ethics statement**

The statement was rephrased. Section was renamed to "Ethics approval and consent to participate" (declarations section, line 2-5, page 10)

B) Comments from Reviewer #1 (Elise C. Pegg, Ph.D.):

**General aspects**

- The approach detailed will not be feasible to use clinically. 1) UKRs are not routinely CT scanned after surgery 2.) The time it would take to scan and then process the scans would be prohibitive
We do agree that CT scans are not routinely performed after UKA implantation (we would rather do X rays for routine check ups). However, in selected cases with long lasting pain, no evidence for periprosthetic infection or before we perform revision to TKA, we do perform CT scans (e.g., to specify malrotation, malpositioning, osseous lesions.) In these selected cases, the creation of a 3D virtual model of cementation can help to assess area-wide cementation quality or to look for signs of aseptic loosening. We do also agree, that the time needed with the semi-automated process, in which every slice is assessed, prohibits large scale clinical use. But we do think that the rapid development of software tools enables us to develop a much faster approach with reasonable effort. For now, we see major benefits on the research side. Based on non-invasive CT-based assessments, penetration can be measured with subsequent load-to-failure biomechanical testing (which might include destruction of the specimen).

• Furthermore, this research is based on the premise that loosening is a result of cementing technique. Although this can be true, there can be other causes of implant loosening, such as polyethylene wear particles causing osteolysis or excessive mechanical loads.

We do totally agree. However, polyethylene wear and osteolysis is difficult to simulate in in vitro cadaver studies. Concerning excessive mechanical loads, a non-invasive measuring technique of cement penetration allows for subsequent failure load biomechanical testing, which should be the endpoint in research on fixation of arthroplasty.

• I think the clinical need for this research and how the authors expect it to be used should be conveyed more clearly.

We changed the conclusion to meet this comment (Conclusion section, line 6-9. Page 8.). Thank you very much.

1. Is a sample size of 12 sufficient?

Obviously, the higher a sample size, the more accurate the results provided by the study. However, in order to meet real life conditions, such as sclerotic bone, variations in bone density (e.g., due to age and gender) and the impact of the implantation procedure, we decided to use human tissue. Alternatives such as sawbones have other levels of density, which is relevant for CT scans. Obviously, they do not allow to determine threshold values applicable for humans.

As cadaver material like human tibia is scarce and need to be treated with the highest level of responsibility, we limited our study to 12 specimens. These 12 specimens were tested using both techniques.

Based on our results, we calculated the required sample size using MEDCALC software (MedCalc Software; Ostend; Belgium). Applying a Type I-error (alpha; significance) of 5%, a Type II-error (beta) of 20 %, and our observed mean difference of ca. 0.005 mm, standard
deviation of the differences of ca. 0,30 mm and a maximum allowed difference of 1,5 mm (reasonable assumptions regarding that the standard thickness of slices in CT scans is 1,0 mm), the minimum required number of pairs would have been 8, which is well below the tested 12 pairs. Based on these assumptions, the presented study with 12 specimens provides a power of 97,5 %, which limits type II error to 2,5 % (see pictures below). In conclusion, we consider a sample size of as sufficient.

2. For the CT scan a thresholding approach was used to identify the cement region, but for the cut surfaces a 'magic wand' algorithm was used (I assume this is based on region-growing). Would it not be more rigorous to use the same/similar algorithm for both techniques? The most important goal for us was to set a reliable, valid gold standard, that precisely measures the true thickness of the cement layer. I do believe that the assessment of the 9 cut surfaces per specimen provides this high quality comparison for the CT-based measuring technique. Furthermore, CT scans are DICOM data sets, with Hounsfield Units (transferred into grey scale), whereas the cut surfaces were scanned and saved as jpegs (color scale). This difference in the underlying data sets (greyscale vs. color scale) is the first and most important limitation to the use of exactly the same algorithm. The second limitation is the fact, that this difference in data sets leads to a difference in software for editing and assessing the data. Moreover, the magic wand tool in Photoshop is a thresholding approach as well. It is a selection tool that automatically selects pixels based on tone and color. The resulting irregularly shaped selections were evaluated by the examiner and manually adjusted wherever necessary. This approach seems feasible, as it is easily possible to distinguish between bone cement and trabecular bone on the high resolution 100 pixel/mm scans of the cut surface. So we basically used a threshold based algorithm for both assessment, and are confident to have a powerful gold standard with the cut surfaces.

3. Would the CT scan HU thresholding approach work reliably in a clinical environment on patients with a variety of bone densities?

The problem of varying cortical and trabecular bone HU in patients due to gender, age or diseases (e.g., osteoporosis) as well as the potential overlap between bone cement density and cortical or trabecular bone density is a potential limitation to the presented technique and deserves further research (e.g., artificial intelligence based calculation of individual thresholds for each patient; (p. 7, line 33). We made these limitations and implications transparent in our conclusion (p. 7, lines 24-30).
However, to meet the challenge of varying bone densities we decided to test our technology on human tibae from donors of both sexes, aging between 53 and 90 years with varying bone densities (mean HU of cortical bone in predefined sections between 92 and 253 HU; see table below) and put a lot of effort in the determination of the appropriate threshold value. In conclusion, the technology presented proofed to deliver reliable results even in this real life situation and demonstrated the feasibility of CT-based cement measurement.

4. Although there was no significant difference found between the two measurement techniques, given the small sample size and high variability is it possible that this result is due to Type II error?

Yes, as any experimental study, we cannot rule out a type-II error. Due to the scarcity of cadaver legs available for biomechanical testing, we could not test our methodology on a large scale basis (100+ specimen). However, based on our assumptions (please see also answer to question No. 1), our study has a considerable power and therefore a limited probability of type-II error. We used graphical methods to make our results comprehensible and tangible.

5. Should Bland-Altman plots not have units on the axes?

Sure, we added [mm] as units on both axes.

6. A radiographer, with practise, can attain X-rays which are in line with the UKR tray (I know this because I have seen this done reliably). From an aligned X-ray of a UKR, would it not be possible to measure the cement penetration depth to a resolution sufficient to make a clinical judgement?

We agree with your statement. However, the quality of X-rays are highly depended on the experience of the radiographer. From our clinical experience, the majority of X rays would not offer a reliable basis for this kind of analysis. However, even with optimal x-rays, it is impossible to distinguish between the anterior and the posterior part of the prosthesis. Due to limited access to the back of the joint via a minimally-invasive approach, it is difficult to clean the bone bed in the dorsal tibial plateau. Furthermore, it might be difficult to put sufficient pressure in this posterior region of the joint during impaction. As a result, penetration is often less in the posterior area of the tibial plateau. This problem, sufficient penetration in the anterior part with insufficient penetration in the posterior part, is hardly visible on a.p. x-rays and the design of the prosthesis often prohibits lateral assessment (see picture below)
B) Comments from Reviewer #2 (Peer reviewer):

1. A table or schematic diagram to elaborate the methodology section would be helpful.

   All 12 specimens were implanted, CT scanned and cut into slices. Each specimen was measured via both techniques. There are no groups. We added a lot of pictures to make the methodology comprehensive and transparent. Concerning the high number of pictures and diagrams already in the manuscript, we decided not to add another diagram.

2. The Discussion could be more elaborate--what could be the other clinical applications this algorithm can be helpful for, etc.

   As Reviewer #1 stated, due to time consuming process and the shortage of postoperative CT scans, large scale clinical application needs further development to automate processes. Furthermore, it should be examined whether MRI scans offer the same opportunity without radiation. Please see altered Conclusion section, line 6-9. Page 8.

Thank you very much for your critical questions concerning the paper. I tried my very best to answer them as open and transparent as possible. I enjoyed the opportunity of rethinking our work.

Best,

Christian Scheele