Peer Review File

Article information: https://dx.doi.org/10.21037/atm-21-5315

Reviewer A

Comment 1: P5, Ln20: Please check with specificity is right word here? Specificity relates to a test's ability to correctly reject healthy patients without a condition.
Reply 1: Thank you for pointing out this inadequate word choice. We have updated the manuscript to correct this wording.
Changes in the text: Page 5, lines 19-20:
"Planar x-ray imaging is the prevailing fracture healing diagnostic tool; however, this technology has proven to be limited by clinician interpretation (14-17)."

Comment 2: P6, Ln16-17: This statement requires a reference.
Reply 2: References for page 6, lines 16-17 are located after the ensuing sentence (page 6 lines 17-19), as these references apply to both sentences (page 6 lines 16-19). We did not provide these references at the end of both sentences to prevent redundant citations in consecutive sentences.
Changes in the text: No changes have been made.

Comment 3: P6, Ln 20-22: There are technologies for instrumenting internal fixation devices and this should be mentioned here.
Reply 3: Technologies for instrumenting internal fixation devices are introduced at page 6, lines 23 – page 6, line 7.
Changes in the text: No changes have been made.

Comment 4: P8, Ln3: Is there a supporting reference, or is the same as the previous sentence (58), not clear
Reply 4: Thank you for pointing out this potential source of confusion. Page 8, line 3 is referring to the work contained within the current study. We have revised this sentence to clarify this point.
Changes in the text: Page 8, lines 1-3:
"New antenna calibration methods were developed in this study to enable antenna sensitivity (i.e., ARF shift per applied load) to be converted directly to fracture stiffness, as the latter metric is more clinically relevant".

Comment 5: P10, Ln3: What is ostectomy, or do you mean osteotomy? Not sure why middle screw hole needed to be filled. (see also ln14, page11)
Reply 5: Ostectomy and osteotomy are functionally synonymous in this context. The middle screw hole was filled to be consistent with previous testing methods, currently cited as (57) in Page 10, line 3. We have modified this sentence to improve the clarity of these methods.
Changes in the text: Page 10, lines 1-5:
"The middle screw hole was filled with a locking screw head (i.e., screw threads removed leaving only the locking head) to facilitate subsequent ostectomy at the LCP midspan and to maintain consistent methods to prior analogous tests (57) (Figure 1A)".

Comment 6: P10, ln19: The variation in antenna distances evaluated appear to be very small, how might antenna distance variations in clinical application influence this method?
Reply 6: The antenna distances referred to in page 10, line 19 refers to the distances used for antenna calibration. This distance does not require a large range as its purpose is to determine the antenna sensitivity (i.e., expected resonant frequency shift per change in implant-antenna distance) for the current testing configuration. The selected 10 μm calibration range is appropriate based on being of the same approximate order of magnitude of expected implant deflections, based on finite element results.
Changes in the text: No changes have been made.

Comment 7: P13, ln11: FEA model of comparable plate geometry, please elaborate. Ln13-14 was the plate straight or contoured to the anatomy, not clear?
Reply 7: The geometry of the FEA plate model was made to recapitulate that of the plate used in the corresponding real-world experiments of this study. We have clarified this point by changing the word "comparable" to "equivalent". The plate was contoured to the anatomy, as is described in page 13 lines 16-18 and depicted in Figure 3C.
Changes in the text: Page 13, lines 13 – 16:
"Three dimensional part files were also generated for a LCP and locking screws of equivalent geometry to those used in the prior cadaveric study, and for an intramedullary nail (IMN) (7 mm diameter, 147 mm length, two 3.5 mm pins for securing the proximal and distal ends)".

Comment 8: P14, ln2: What does a 1.0 index fracture callus mean, is this common termninology, if not why use it. If so, please reference.
Reply 8: Callus index is common terminology [Ref 1] and a contextual definition is currently provided when used within this manuscript at page 14, line 5 "A 1.0 index fracture callus (i.e., no periosteal callus) was modelled within the fracture gap".
[Ref 1]: Eastaugh-Waring SJ, Joslin CC, Hardy JR, Cunningham JL. Quantification of fracture healing from radiographs using the maximum callus index. Clin Orthop Relat Res. 2009 Aug;467(8):1986-91. doi: 10.1007/s11999-009-0775-0.
Changes in the text: No changes have been made.

Comment 9: P14, ln12: Any idea how the values of fracture callus relate to difference stages/temporal evolution of fracture healing?
Reply 9: Fracture healing is a highly variable process, with the material and structural compositional progression of the fracture callus varying on a case by case basis. Thus, attempting to map the spatially and materially homogeneous callus elastic moduli used for FE simulations to a specific fracture healing time point would prove misleading. However, we believe this limitation to be permissible for the proof-of-concept nature of the current analysis; we justify this with the assumption that the structure / composition of the callus will feature an overall structural rigidity comparable to some point within the range of materials tested in this study (i.e., the rigidity falls some point between the 0.01% and 100% callus models tested).
These limitations and their justification are discussed in great detail from page 23, lines 13 – 22.
Changes in the text: No changes have been made.

Comment 10: P16, ln5: Why was bone strain used as validation parameter, when the quantities of interest are deflection of the implant. Why was implant deflection or implant strain measured?

Reply 10: We agree with your concern that direct validation with implant deflection would be the ideal parameter. However, direct real-world measurements of implant deflections resulting from comparable bending in a material testing machine is highly prone to measurement error / uncertainty due to the exceedingly small implant displacement magnitudes (i.e., on the order of tens of μm) which can be obfuscated by the fixture vibrations inherent to a servo-hydraulic loading device. Mechanical strain measurements are less prone to high measurement uncertainty and thus was selected as the mechanical metric for model validation. Similarly, accurate measurements of plate implant strains are arduous due to the high strain gradients resulting from screw-hole induced stress risers, which is why bone strain was used. Relative changes in implant deflections, rather than the absolute value at any given callus elastic modulus, are the data of interest; therefore, imperfect accuracy in FE absolute implant displacement predictions seems reasonably permissible. We have added this justification to the Methods subsection "Strain Gage Validation" (page 16, lines 15-18).

Changes in the text: Page 16, line 15-18:

"To validate the accuracy of the FE models, strain results were compared between in vitro and in silico tests; strain, rather than implant deflection, was selected for the mechanical validation parameter to obviate measurement uncertainties associated with accurate observation of exceedingly small implant deflections (i.e., on the order of tens of μm)".

Comment 11: P19, ln19: In the intro it was stated that average time to diagnose non-union was 6 months and here stated fracture stiffness can predict 2 weeks earlier. That does not appear to be clinically significantly earlier detection of impaired healing? Nor would it appear that it can be classified as early prediction of fracture healing outcome. Please clarify.

Reply 11: The 2.5-week improvement described in this sentence refers to the results from the referenced literature, and is not in discussion of the results of this study. While 2.5 weeks may not sound substantial, this study represents a very early state in the development of a novel technology. This sentence merely provides support for the postulation that mechanical diagnostic techniques may provide expedited prediction of fracture healing outcome. We have revised this sentence to emphasize that additional technology development is necessary for potentially achieving early fracture healing prediction in a clinical setting.

Changes in the text: Page 20, lines 21 – 23:

"Thus, the demonstrable accuracy of DEC quantifications of fracture stiffness supports the conclusion that continued study and development of this technology may provide clinical utility for early prediction of fracture healing outcome".

Reviewer B

Comment 1: Page 10. Line 15. How can calibration be done in clinical settings? Is it only used to check the FEM accuracy?

Reply 1: Antenna calibration methods function as a means to convert antenna resonant frequency shifts into a direct measure of fracture bending stiffness as described in the introduction (page 8, lines 1 – 3), within the methods section (page 10, line 16 – page 11, line 4), and graphically depicted in Figure 2. The calibration methods used in this study were developed specifically for clinical settings, and can be clinically implemented using the same methods described in page 10, line 16 – page 11, line 4.
Antenna calibrations are unrelated to FEM data; the FEM studies were performed as an *in silico* means to evaluate whether bending induced implant deflections will reduce with fracture healing, as these reductions are foundational to the use of DEC antenna measurements for diagnostic applications (please see page 12, line 21 – page 13, line 5).
Changes in the text: No changes have been made.

Comment 2: Page 11. Line 10. Explain in more detail the preload cycles.
Reply 2: Preload cycles standardly refers to application of conditioning loading cycles, of loads comparable to those used during the subsequent data collection, in order to prevent viscoelastic effects from impacting the mechanical properties measured in biological tissues. For clarification, we have changed "preload cycles" to "preconditioning load cycles".
Changes in the text: Page 11, lines 9 – 12:
"Cyclic four-point bending loads were applied to the sample via compressive displacement of the MT crosshead (1.0 – 3.0 N-m at 0.05 mm/s cross-head displacement rate, n = 5 preconditioning load cycles, and n = 5 data collection cycles per test)".

Comment 3: Page 12. Line 8. Elaborate on the "individual data points"
Reply 3: Thank you for pointing out this potential source of confusion. This sentence was intended to portray that no data was omitted for outlier or experimental error reasons. Additional data analysis has been added to this manuscript at the request of another reviewer, thus necessitating outlier removal for that particular analysis. Please see page 12, lines 10 – 12 for the corresponding revisions, which are copied below for convenience.
Changes in the text: Page 12, lines 10 – 12:
"No data was omitted from when comparing MT and DEC stiffness values. Single outliers were removed from the DS and FF groups when calculating individual sample's percent differences between MT and DEC predicted bending stiffness values".

Comment 4: Page 14. Line 23. How were these smooth cylinders interfaces with the irregular bone surface?
Reply 4: Thank you for pointing out this detail omission. We have added additional details to page 15, line 8 to clarify the surface-to-surface constraint used for load transfer from the cylinders to the cortical bone.
Changes in the text: Page 15, lines 8 – 10:
"Hard contact properties were assigned between the cylinder and cortical surfaces to facilitate application and transfer of mechanical load via these contact points".

Comment 5: Page 15 Line 2. These boundary conditions do not seem to fully exclude the possibility of rigid body motion? Could this affect displacements and the analysis?

Reply 5: Thank you for pointing out this detail omission regarding the model boundary conditions. The model used boundary conditions to prevent rigid body motion, however this was not detailed in the original manuscript. We have added additional details to page 15, lines 2 – 5 to clarify the boundary conditions implemented to prevent rigid body movements.
Changes in the text: Page 15, lines 2 – 5:
"To prevent rigid body translations, a zero proximal-distal-direction translation boundary condition was applied to the metatarsal proximal-distal mid-plane. Similarly, a zero cranial-caudal-direction translation boundary condition was applied to the metatarsal's cranial-caudal mid-plane".

Comment 6: Page 15. Line 16. could you elaborate on the mesh refinement?
Reply 6: Mesh refinement definition has been added after the first use of this term (please see page 15, line 22).
Changes in the text: Page 16, line 1:
"Mesh refinement (i.e., reduction of the average element volume) beyond the …"

Comment 7: Page 16. Subsection "strain gauge validation". This effectively describes a small project (validation of the RF displacement detection method) within the main project but is missing details and illustrations. This section should be more elaborate. For example, the comparison of the FE and experimental results is mentioned and FE is deemed validated but there is no graphical presentation to prove that?
Reply 7: We agree with the critique that additional information would benefit the reviewer, but respectfully disagree that the manuscript is missing details regarding this experiment. We have edited the methods section to detail the exact strain values observed in both FE and real-world models (please see page 17, lines 11 – 14). Graphical representation of these values would be superfluous as such a plot would not provide additional information beyond the now included mean ± standard deviation strain values. Detailed methods for this sub-experiment are provided throughout page 16, line 15 – page 17, line 14. Graphical representation of these methods would not illustrate additional details beyond what are already shown in Figures 1E and 3D.
Changes in the text: Page 17, lines 11 – 14:
"For both fixation methods, mean FE strain values (68.0 ± 4.2 με, 72.1 ± 4.1 με, for plate & IMN; respectively) were within one standard deviation of experimental strains (81.6 ± 37.6 με, 101.6 ± 30.8 με, for plate & IMN; respectively), thus the FE models were deemed validated".

Comment 8: Page 16. Line 11. Could you provide a reference for the standard strain application techniques for bones?
Reply 8: We have updated this line to reflect that strain gage applications followed manufacturer recommended application techniques.
Changes in the text: Page 16, line 22:
"Following manufacturer recommended application techniques, stacked strain rosettes (C2A-06-062WW-350; Micro-Measurements; Raleigh, NC, USA) were adhered to …"

Comment 9: Page 17. Line 7. This naming scheme is a bit confusing as MT is referring to a very wide range of techniques, within which DEC is one method to measure displacement. The main point here is that you are comparing two different strain/displacement measurement techniques and both are part of MT methods?
Reply 9: The DEC technique is not a standard material testing technique, and refers to a novel diagnostic technique. We selected the two acronyms, DEC and MT, to reflect that we are comparing fracture stiffness values predicted by a novel and pre-existing standard technique. We do not think it is unreasonable to use the MT acronym given that its meaning within the context of this study is clearly defined and described on page 11, lines 6-13. However, we do agree that subsequent use of the MT acronym when describing material testing equipment is misleading; therefore, we have removed this acronym from use in this specific context.
Changes in the text: Page 11, line 7; Page 7, line 10

Comment 10: Page 17. Line 13. Shift the reference to fig 4 earlier in the paragraph.
Reply 10: We have moved the requested reference to three sentences earlier in the paragraph.
Changes in the text: Page 17, line 21

Comment 11: Page 18. Line 22. How could your tool be used in the clinic? How can the calibration be made in a clinical situation? Can the sensing resolution be maintained at a larger readout distance (though skin and muscle) when applied clinically - has to be applied non-invasively?
Reply 11: The use of DEC in a clinical setting can follow the same methods used in this study. This device was designed and built with clinical implementation in mind, and is a design iteration of a DEC device which has already been implemented in translational animal fracture studies, as referenced in the introduction (page 7, line 20). Understanding how this technology can perform through increasing depths of tissues is a necessary step for clinical implementation of this tool; however, predictions of this performance would be purely conjecture without experimental data to support our claims. We thus have recommended, throughout the discussion section, that follow up studies be performed to better characterize the DEC technology.
Changes in the text: No changes have been made.

Comment 12: Page 19. Line 1. Figure 4. Does it look like there is a bias/trend for lower displacements (higher stiffness) with the DEC-based sensing? Could you comment on that?
Reply 12: Statistical comparison strongly supports there to be no meaningful differences in fracture stiffness values obtained via MT versus DEC methods (from page 17, line 18: $p = 0.587$, $p = 0.985$, $p = 0.975$; for intact, DS, and FF comparisons; respectively). We therefore do not find it appropriate to include discussion of bias in the differences of these results as this would insinuate data interpretation which is not statistically supported.
Changes in the text: No changes have been made.

Comment 13: Page 20. Line 21. Instead of a DEC model, better refer to it as a physical model. The DEC is just a type of displacement gage?

Reply 13: We are uncertain to which line this critique refers. There is no mention of DEC near page 20, line 21 of the original manuscript (now located at page 21, line 22 of the updated manuscript), and we find no use of "DEC model" elsewhere in this manuscript.
Changes in the text: No changes have been made.

Comment 14: Page 21. Line 1. Was there a check to verify that all observed deflections and strains were within the linear range of the materials?
Reply 14: No formal check was performed to parametrically verify that observed deflections were within the linear range for each model permutation. Such analysis was deemed unnecessary based on the extremely low load magnitude relative to the implant stiffness. For example, for the titanium plate with reduced stiffness structure and minimum callus modulus (i.e., the model permutation with the lowest expected bending rigidity), maximum Mises stress at the implant midspan was approximately 10% of the material yield stress.
Changes in the text: No changes have been made.

Comment 15: Page 21. Line 17. Given the inter-sample variation (repeatability) of the DEC method, how much of this temporal dependence, shown in FE, is expected to be observed in practice?
Reply 15: DEC *intra*-sample variation is low; *inter*-sample variation is to be expected, regardless of whether measuring via DEC or MT methods, due to inherent differences in samples' stiffness resulting from anatomical variations, positioning of the implant hardware, etc. However, use of DEC in a clinical setting relies upon *intra*-sample stiffness progression to elucidate diagnostic trends, thus we do not see reason to expect DEC to deviate from the trends shown by the FE data.
Changes in the text: No changes have been made.

Comment 16: Page 22. Line 23. If the level of confidence is low, why was it used - provide some argumentation pro- your method?
Reply 16: No computational model can ever act as a perfect predictor of clinical progression due to fracture healing being a highly variable process, with the material and structure compositional progression of the fracture callus varying on a case by case basis. However, we believe this limitation to be permissible for the proof-of-concept nature of the current analysis; we justify this with the assumption that the structure / composition of the callus will feature an overall bending rigidity comparable to some point within the range of materials tested in this study (i.e., the rigidity falls some point between the 0.01% and 100% callus models tested). The FE models in this study thus still fill a pertinent knowledge gap, necessary for further development of DEC, without the need for excessive animal testing. The utility of this data, despite its current limitations, is discussed at length in page 21 lines 9 – 16 and page 23 line 13 – page 24 line 5.
Changes in the text: No changes have been made.

Reviewer C
Comment 1: Figure 4 demonstrates the differences in stiffness between the Gold Standard MT methods and the proposed DEC. While it was useful to show the differences in stiffness

magnitude, it would be more meaningful to report the % difference in stiffness. This will indicate the expected error in measurements that may be meaningful when employing these techniques clinically.

Reply 1: Thank you for this insight. We agree that use of percent difference is meaningful for presenting measurement errors, as this data can become obfuscated when presenting through mean stiffness measurements alone. We have included a new box and whisker plot (please see Figure 4B, copied below for convenience) to present the percent difference in MT and DEC measurements for each sample. We have also modified the Results subsection "Ex Vivo Prediction of Fracture Stiffness" to introduce this data. Additionally, we have updated the second paragraph of the Discussion section to clarify the percent differences observed in this study, and to emphasize that relatively large percent errors in individual samples suggests that continued technology development is necessary to improve the reliability of this emerging technology.

Changes in the text: Figure 4:

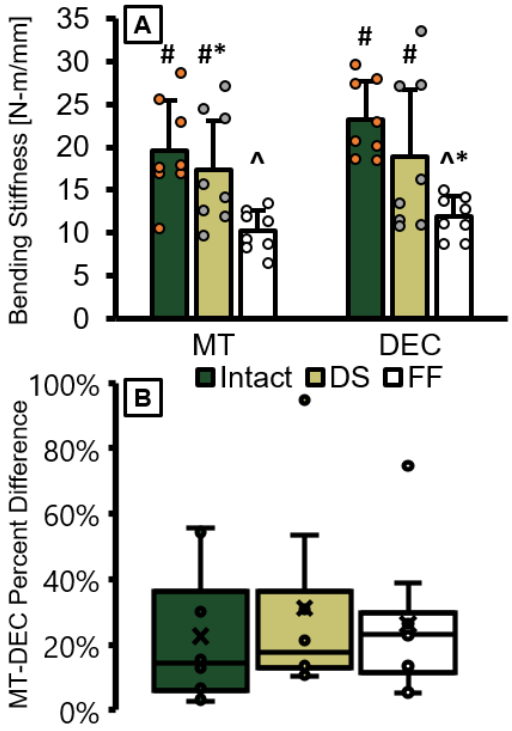

Figure 4: Fracture bending stiffness predictions. (A) Bending stiffness of cadaveric metatarsals of decreasing stability levels (i.e., intact, destabilized [DS], and fully fractured [FF] states) were measured using material testing (MT) methods and direct electromagnetic coupling (DEC) predictions. Bars depict mean + standard deviation, where overlaid points correspond to the mean values of each sample (n = 8). Bars which do not share symbols (#, *, ^) are significantly different ($p < 0.05$). (B) Percent differences of MT and DEC bending stiffness measurements for each sample (n = 8) represented as a standard box and whisker plot.

Page 18, lines 3 – 7:

"For each sample, percent differences between MT and DEC predictions of fracture stiffness were individually calculated for each fracture state; mean percent differences (±

standard deviation, after outlier removal) were 22.7 ± 21.8%, 22.1 ± 15.6%, and 19.5 ± 12.1% for the intact, DS, and FF states; respectively (Figure 4B). DS and FF states each had single outliers, 95% and 75% respectively, which were not included in the preceding mean percent difference calculations".

Page 19, line 20 – page 20 line 5:
"The data of this study, however, suggest DEC can predict fracture bending stiffness to within, on average, 23% difference of MT methods (Figure 4B), with no significant differences in values measured by each method for any fracture state (Figure 4A). Unlike MT devices, DEC is highly portable and non-invasive, and therefore well suited for use as a diagnostic tool in clinical or home tele-medicine applications. It should, however, be noted that individual samples did exhibit erroneously high percent differences in MT versus DEC predictions of fracture stiffness, thus suggesting that additional improvements are necessary to improve the reliability of this emerging technology".

Comment 2: It also would be meaningful to modify Figure 4 to show the differences between the MT and DEC for each paired test. For example: In line 2 of Pg 19, you state that the DEC can predict fracture bending stiffness within 25% difference of MT methods. Please refer to two data points in Figure 4: The orange data point at 11 N-m/mm at Intact MT vs. 18 N-m/mm at Intact DEC. That represents a 63% increase in stiffness with the DEC method. Percent error should be considered for all comparisons and discussion points.
Reply 2: We agree with your critique, and have addressed this during the corrections corresponding to your Comment 1. Please see Reply 1 and the corresponding "Changes in the text" for details on how we've rectified this concern.
Changes in the text: Same as Comment 1.

Comment 3: If the applied load/moment is consistent among experimental configurations, fracture stiffness and implant deflections are synonymous metrics. However, stiffness is reported in the experimental portion of the study, but maximum deflection is reported in the computational part. Please report the computational fracture stiffness instead of the midsection span deflection. It will elucidate the applicability of using the DEC system over the course of healing time when comparing the computational data to your experimental data.
Reply 3: We agree that bending stiffness is a useful metric which can be extracted from the computational data; however, the primary purpose of this analysis was to 1) act as proof of concept regarding the clinical feasibility for DEC to quantify changes in the fracture site & 2) assist in subsequent development of DEC. DEC functions by quantification of deflections, which are subsequently converted to stiffness. However, it is known that DEC techniques have a lower limit on the minimum detectable implant deflection; thus, computational prediction of implant deflections is a far more important metric for the ongoing development of DEC devices and/or techniques for clinical implementation.
Changes in the text: No changes have been made.

Comment 4: Pg 10, Line 5: Refer to your 56 and 57 references here.
Reply 4: We have included reference 57 at the requested location; reference 56 follows a slightly modified loading procedure and is thus not applicable here.

Changes in the text: Page 10, lines 6-8:
"Samples were placed in a custom pneumatic DEC loading fixture that controlled four-point bending loads applied to the sample, thus inducing implant deflections towards the DEC antenna positioned at the LCP midspan (Figure 1B) (57)".

Comment 5: Pg 11, Line 3 and Figure 2: Looks like there are multiple data points taken on Figure 2 for repeatability. Please report the uncertainty in the calibrated DEC measurements.
Reply 5: Figure 2 appears to have data from multiple testing cycles, but is in fact a representative sample of a single data collection cycle. Within a single data collection cycle, multiple data points were collected at each value of the dependent variable in order to lessen effects from singular noisy / erroneous data measurements. The data of interest from any given subplot of Figure 2 are the slopes of linear fits applied to the collected data, and thus reporting of uncertainty is not applicable. Furthermore, the data presented in Figure 2 is from a single random representative sample, therefore all specific values from these plots, including uncertainty, are arbitrary.
Changes in the text: No changes have been made.

Comment 6: Pg 11, line 9: Why was a maximum of 3.0 N-m moment used vs 2.5 in the DEC measurements? It may not matter IF the stress/strain curve does not exhibit a significant toe in region.
Reply 6: Increased bending loads were used as a superfluous precaution to ensure sufficient data collection beyond any potential toe regions. The data of interest, for both MT and DEC methods, were slopes of linear fits applied to data beyond potential toe regions. Given the mechanical loading for both of these methods was well within the elastic region of the materials / structures, the slope magnitudes are reasonably assumed to be invariant to slight deviations in the load magnitude.
Changes in the text: No changes have been made.

Comment 7: Pg 12, line 4: What is meant by animal confounders? Between species? Between sample?
Reply 7: This line was included to follow the journal required ARRIVE checklist. We have updated this sentence to clarify this point.
Changes in the text: Page 12, lines 5 – 7:
"In the current study, pre-sacrifice inter-specimen animal confounders were not thought to be substantive for a comparative cadaveric ex vivo study, and thus were not controlled".

Comment 8: Pg 12, line 8: For "individual data points", do you mean individual "stiffness" data points we not omitted because of the "singular exclusion criterion" or because no outliers were detected?
Reply 8: Thank you for pointing out this potential source of confusion. We have revised this sentence accordingly.
Changes in the text: Page 12, lines 10 – 12:
"No data was omitted from when comparing MT and DEC stiffness values. Single outliers were removed from the DS and FF groups when calculating individual sample's percent differences between MT and DEC predicted bending stiffness values".

Comment 9: Pg 14, line 13: Reference 83 is refers to human and bovine bone material properties. Please provide a reference for the values in Table 1.
Reply 9: The values provided in Table 1 were obtained from cited reference 83. Bone, at a structural scale, exhibits a relatively wide range of material property values depending on the individual, location within the bone, and the species. The material properties used within this study have considerable overlap with those reported in literature for both ovine and human specimens, thus we believe the material properties used within this study to be appropriate [Ref 2, Ref 3].
[Ref 2]: Morgan, E.F., Unnikrisnan, G.U., Hussein, A.I., 2018. Bone Mechanical Properties in Healthy and Diseased States. Annu Rev Biomed Eng. 20, 119–143. doi:10.1146/annurev-bioeng-062117-121139.
[Ref 3]: Grant, C.A., Wilson, L.J., Langton, C., Epari, D., 2014. Comparison of mechanical and ultrasound elastic modulus of ovine tibial cortical bone. Med Eng Phys 36, 869–874. doi:10.1016/j.medengphy.2014.03.012
Changes in the text: No changes have been made.

Comment 10: Pg 16, line 23: Please report the strain values or differences in the strain values in the results or in supplementary material. This will help prove the validity the FE models.
Reply 10: We have revised the methods section accordingly.
Changes in the text: Page 17, lines 11 – 14:
"For both fixation methods, mean FE strain values (68.0 ± 4.2 με, 72.1 ± 4.1 με, for plate & IMN; respectively) were within one standard deviation of experimental strains (81.6 ± 37.6 με, 101.6 ± 30.8 με, for plate & IMN; respectively), thus the FE models were deemed validated".

Comment 11: Pg 17, line 8: Please report p-values for the statistical differences between MT and DEC values.
Reply 11: Please see page 17, lines 18 – 21.
Changes in the text: No changes have been made.

Comment 12: Pg 17, line 13: The 82.2 reduction in mean implant bending deflection was in response to what imparted moment? 2.5 N-m?
Reply 12: Thank you for pointing out this potential source of confusion. This value corresponds to a 2 N-m imparted load. We have revised this sentence to clarify this detail.
Changes in the text: Page 18, lines 10-11:
"Calibrated DEC results, for an applied 2 N-m four-point bending load, predicted an 82.2 μm reduction in mean implant bending deflection from intact to FF states".

Comment 13: Pg 17, line 17-22: The deflections reported are without context. I believe that the 82.2 um reduction as described in Line 13 is of importance to understand the state of healing (i.e. intact vs FF). Much more meaningful results would describe which parameters (size, material, etc.) are able to predict over the 82.2 threshold or a stiffness threshold as described. You elude to this on pg 21, first paragraph, but provide quantities IF you believe the 82.2 threshold is the most important.

Reply 13: We respectfully disagree that the results reported on page 17, lines 17-22 of the original manuscript (now page 18, lines 14 – 22 of the updated manuscript) are reported without context. The exact model permutations which led to the reported deflection reductions are explicitly stated. We have amended this section to reference the corresponding subplots in Figure 5 which graphically depict these data in more granular detail. Fracture fixation permutations which exhibited implant reductions near or greater than 82.2 μm have been identified in page 23, lines 3 – 12.
Changes in the text: Page 18, lines 18 – 22: Figure references added throughout this paragraph.

Comment 14: Pg 18, Line 4-11: Please put figure letters on Figure 5 and refer to these in the text. In addition, does the 74.4 um deflection correspond to a 2.0 N-m load while the 82.2 correspond to a 2.5 N-m load?
Reply 14: We have updated Figure 5 to include letters, and have made updates throughout the text to refer to these labelled figures. Clarification regarding the bending loads for the 82.2 value were addressed in response to Comment 12. We do not believe that additional clarification is required at the current location (page 21, lines 13 – 16 of the updated manuscript), as this sentence states that these values are from "comparable fracture fixation models".
Changes in the text: Page 18, lines 10 – 11:
"Calibrated DEC results, for an applied 2 N-m four-point bending load, predicted an 82.2 μm reduction in mean implant bending deflection from intact to FF states".
References to Figure 5 have been added throughout the Results subsection "Finite Element Predictions of Implant Deflections" (please see page 18, lines 18 – 22 & page 19, line 8) and in the 9th paragraph of the discussion (please see page 22 lines 18 & 23).

Comment 15: Pg 19, Line 10: Please cite studies regarding previous antenna versions
Reply 15: Studies regarding previous antenna versions are cited in the preceding sentence (please see page 20, lines 6-7 of the updated manuscript). We thus have not added additional citations to avoid redundant references in consecutive sentences.
Changes in the text: No changes have been made.

Comment 16: Pg 19, Line 12: Because of the clinical applicability, would you be able to comment on the moments applied to the bones. Would these be viable in terms of minimal pain for individuals? Again, this may be out of the scope of your experimental design, but you could again point to the ability to implement this system clinically without adverse events.
Reply 16: The loading mechanism used in this study is equivalent to that used in previous *in vivo* studies in an ovine fracture model. We have clarified this point by citing this reference when describing the load application method (please see page 10, line 18 of the updated manuscript).
Changes in the text: Page 10, lines 16 – 18:
"Bending loads were applied to the intact sample (1.0 – 2.5 N-m, in 0.25 N-m increments, n = 5 preload cycles, and n = 5 data collection cycles per test, Figure 1B-D) while measuring the resultant change in ARF shift (Figure 2A) (57)".

Comment 17: Pg 19, Line 18: Should this be "N-m/mm" since your study was in "N-m/mm"? If not, is there a means to convert this number to compare to your study?
Reply 17: This value was obtained from literature and thus we do not have the geometric parameters to convert this value to be comparable to our study. Even if this value could be converted, direct comparison of these two values is unlikely to be applicable given that the literature value refers to humans rather than the ovine samples used in this study. We included this value from literature with the intent of showing that fracture stiffness has shown viability as a diagnostic metric, rather than for direct comparison to the values measured in our study.
Changes in the text: No changes have been made.

Comment 18: Pg 19, Line 20-22: More quantitative context is needed relative to your DEC system such that the system has enough sensitivity to predict union/non-union conditions. For instance, can you demonstrate that your measurement errors are insignificant as compared to the clinical meaningful differences?
Reply 18: We agree that additional analysis is necessary for clinical diagnostic efficacy of this device. At this time, the requested quantitative analysis cannot be performed without comparable clinical data with this device. We have accordingly revised the line in question to better reflect the need for continued study and development of this technology to achieve clinical utility.
Changes in the text: Page 20, lines 21 – 23:
"Thus, the demonstrable accuracy of DEC quantifications of fracture stiffness supports the conclusion that continued study and development of this technology may provide clinical utility for early prediction of fracture healing outcome".

Comment 19: Pg 20, Line 6-7: What additional translational studies do you propose? Within humans?
Reply 19: Thank you for pointing out this potential source of confusion. We have updated this sentence to specify that *in vivo* large animal and preclinical fracture studies are necessary to better understand this technology.
Changes in the text: Page 21, lines 6 – 8:
"Additional *in vivo* large animal and preclinical fracture studies will therefore be pursued to better establish the efficacy of DEC in a clinical setting".

Comment 20: Pg 21, Line 7-9: Can you provide relative differences (like % increases or decreases) to quantify the changes in magnitude from implant type and callus mechanics?
Reply 20: 96 permutations of fracture and treatment type were analyzed in these parametric finite element analyses, thus there are 9,120 total percent differences which can be reported for these models. We therefore believe the most appropriate way to present this data is to highlight the most meaningful results, while graphically displaying all data (Figure 5), as has been done in page 18, line 14 – page 19, line 8.
Changes in the text: No changes have been made.

Comment 21: Pg 21, Line 10-11: Can you comment what you expect the effect of hetero-structural/material properties would be on your results? Do you propose a future study on this?

Reply 21: Predicting the exact impact of a materially and structurally inhomogeneous fracture callus cannot be done with any level of confidence due to fracture healing being a highly complex process which is patient and case specific in its temporal progression. However, we believe this limitation to be permissible for the proof-of-concept nature of the current analysis; we justify this with the assumption that the structure / composition of the callus will feature an overall bending rigidity comparable to some point within the range of materials tested in this study (i.e., the rigidity falls some point between the 0.01% and 100% callus models tested). The FE models in this study thus still fill a pertinent knowledge gap, necessary for further development of DEC, without the need for excessive animal testing. The utility of this data, despite its current limitations, is discussed at length in page 21 lines 9 – 16 and page 23 line 13 – page 24 line 5.
Changes in the text: No changes have been made.

Comment 22: Pg 23, Line 1: As with the previous comment, do you propose any future studies that would address "the highly variable process". Can an FE model provide some insight?
Reply 22: As stated in the reply to the previous comment, development of computational models for this purpose is a highly advanced and non-trivial process which is outside the scope of this work. More useful insights can instead be gained through *in vivo* testing, which has been recommended in page 21, lines 6-8.
Changes in the text: No changes have been made.