Reviewer 1 (Mark Robinson)

In the manuscript I received there were two subtly different versions of the abstract. The version that includes the quantification of the JCF data is more informative.

**Thank you for pointing this out. The version without data was included unintentionally. The revision now has only one version (the one with data).**

L65-67 The authors suggest that “abrupt changes in gait mechanics” could “overwhelm the adaptive response of cartilage”. Are the authors aware of any studies that might be able to support this statement? Over what time period does cartilage adapt to changes in mechanical stimulus?

**There is evidence that extended periods of unloading reduce the glycosaminoglycan (GAG) content of articular cartilage (Owman et al., 2014). GAGs affect the compressive stiffness of cartilage. Similar results are seen for measured cartilage stiffness in animal models (Jurvelin et al., 1986). Decreased stiffness should result in greater internal strain from a given JCF. The time period over which cartilage adapts to mechanical unloading-then-loading is not well studied, but the Owman et al. (2014) study found that GAG content was still below baseline a year after six weeks of unloading.**

**This topic was briefly addressed in the original Discussion, but associated text and citations were added in the Introduction (Line 71).**

As this sentence essentially reflects the authors “mechanical overloading” hypothesis, can the authors also justify their JCF variables analysed (peak, loading rate and impulse) in the context of “overwhelming the adaptive response”.

**It is unknown which particular feature of loading cycles (e.g. peak, rate, impulse) is most important mechanistically in cartilage failure. For this reason we chose to analyze several features that have previously been associated with OA. This information was added (Line 82).**

L72 The hypothesis you state is not precisely the hypothesis you test. Your hypothesis implies that you statistically examine the entire gait cycle when in fact your results examine three discrete representations of the JCF; the peak, the loading rate and the impulse. I recommend that you either make your hypothesis more specific to these 0D (discrete) variables or consider a more suitable statistical approach that would evaluate the JCF over the entire gait cycle (a 1D analysis e.g. statistical parametric mapping www.spm1d.org or functional data analysis).

**The hypothesis was revised to be more specific to the outcome variables (Line 79)**

L75-77 The rationale for examining the medial knee is based on the prevalence of medial knee OA. As the study of Wise et al. (2012) is not based on amputees and as the subsequent studies using the external KAM do not have convincing outcomes can this choice be strengthened for example by either by referring to other studies observing other features of amputee gait (e.g. trunk kinematics or knee abduction angles) or the prevalence of medial knee OA in veterans.

**This is a good question. We added references on the incidence/prevalence of knee OA in the young military population in general and in those with limb loss specifically (Line 52). To the knowledge of the authors, there are no studies showing the prevalence of medial knee OA specifically in military populations with unilateral limb loss (previous studies on the military population have reported knee OA in general; medial vs. lateral not specified). Some studies on medial knee bone mineral density and WORMS score suggesting risk for medial knee OA in the limb loss population were added (Line 86).**

Methods:

L95: If the authors have the variability data for the time walking independently, then please can this be added.

**These data were not available unfortunately. The gait analyses were performed as close to two months as was practically feasible given participant availability in a rehab setting.**

L175: The authors report standardised effect sizes, to quantitatively judge the size of the effect observed. However my preference would be to present the confidence intervals of the difference along with some indication of what magnitude of difference between the two groups the authors would consider to be meaningful. Ideally this change would be based on an understanding of what magnitude of JCF difference might “overwhelm the adaptive response” that therefore in turn support the “mechanical overloading” hypothesis.

**95% CIs were added to the text as requested (Line 187). Effect sizes were retained as well; they were not really used in this study as an interpretive aid, they are just included as an additional conservative check on reporting “differences” given the small sample sizes.**

**The question on minimum meaningful difference is a good and important one. Unfortunately this value (if it exists) is unknown. We speculate it is highly subject-specific. The requirement of a large effect size (d > 0.8) seems a reasonable place to start. Effect sizes for studies on medial joint loading and OA initiation (Amin et al., 2004) and progression (e.g. Miyazaki et al., 2002) in older adults are typically much smaller than 0.8, so we think this is a fairly conservative setup. These studies are on the adduction moment, not on contact forces, but there is not a lot of data to use for this purpose at the moment. Text on this topic was added (Line 189).**

**Relatedly, some recent studies have reported “minimum detectable changes” in medial knee joint loads from repeated-measures designs (between-day reliability, essentially). This value is not necessarily the “minimum biologically meaningful difference” but it is relevant to this question. Discussion on this topic was added (Line 306).**

L181: Would it be possible to add the model input parameters into the supplementary material to increase transparency and aid easy replication?

**We added a table to the manuscript with the baseline parameter values of the joint contact model (Lines 164, 586).**

L187-189: I very much liked the idea of the sensitivity analysis of the model parameters to further evaluate the results but at the end of the paragraph describing the analysis I was not clear how the outcome of the analysis was to be interpreted. Please add a brief sentence to clarify.

**The sensitivity analysis provided the fraction of perturbed parameter sets that produced greater outcome variables in the Limb Loss group. This information was added (Line 208).**

**If the reviewer was referring to how the interpretation of these fractions was made (i.e. the statements in the Discussion on Line where 98% = “high degree of confidence” and 73% = moderate degree of confidence”), these interpretations are purely subjective. We are open to other interpretations if the reviewer disagrees. The amount of parameter variation tested in this analysis (CV = 10%) is a reasonable assumption for these types of parameters (e.g. Hasson & Caldwell, 2012; Line 203), given that this is a fairly homogenous population (young male military service members).**

Results:

L219: I do not follow the interpretation of transtibial subjects walking with less knee flexion (fig 6). It appears from figure 6 (although the flexion/extension direction is not labelled) that they have more flexion than the control group during early stance.

**We mis-read this figure in the original submission, thank you for pointing this out. The inaccurate statement was removed.**

Conclusions:

L291: In my opinion to suggest that the increased loads observed may be a risk factor for OA would require more substantial insight into the magnitude of the load increase that you have observed (also see my second comment in the methods section). Whilst you have observed “relatively high loads” can you link this to a meaningful change/adaptation in the cartilage to qualify this as a risk factor?

**We have expanded this commentary as noted above (referring to the comment on adding confidence intervals; Line 306).**

Reviewer 2 (Anonymous)

**Preamble: With all due respect, we struggled in responding to Reviewer 2’s comments because the technical basis for their concerns/objections were unclear to the authors in several places. We believe the reviewer may have misunderstood what was done in the sensitivity analysis and/or what our purpose of performing it was. We have endeavored to clarify these points in the revision.**

Basic reporting

The text is clearly written. Units of %BW are not typical and I think should be changed to just the ratio BW (not multiplied by 100).

**This change was made.**

The introduction adequately describes and references the background and motivation of the work. Structure is fine except that there is no statement of funding which I think is required and is standard:

"Funding Disclosure

1 Separately from declaring Competing Interests, PeerJ also requires that authors disclose the financing which made their work possible.

2 The Funding statement is published in the final article. This disclosure provides added transparency."

**A statement of funding was included in the original submission. The authors do not know if reviewers are able to see it at the reviewing stage.**

Figures are good, except units of %BW instead of BW makes numbers much larger, should be changed. Submission seems to be sufficiently 'self-contained'. Is the raw data and code available? I only see the processed/final measures/outcomes (Average? medial contact force throughout stance for each subject, average? peaks, impulses, for each subject) I believe raw data (which you have not included) is required as per Peerj guidelines:

Obviously the raw data would be a lot with all the markers and ground reactions forces for all used trials for all subjects, but there are repositories available for sharing. Some of the code and models may not be yours to freely distribute or you may not have access to the code, but I believe the intent behind this guideline is to make as much data freely available as possible so that others could similarly analyse the results.

**We provided the data for the outcome variables that a reader would need to verify the statistics presented, which we believe meets the standards of transparency for this journal. Additional data in “rawer” stages of the processing can be included if required by the journal. The data in its “rawest” form (marker positions and force platform output) is a very large amount of data that will be of little use to readers seeking to replicate the present results without the commercial software (Vicon Nexus, Visual3D) used to process it.**

Experimental design

Seems to be mostly original work, well defined question, relevant to the issues presented, and seemingly meaningful. Knowledge gap is presented and explored. Investigation seems to be conducted not very rigorously and is to a very basic/low, not high, technical standard regarding the estimation of muscle and joint reaction forces. Methods are sufficiently described. The research seems ethical.

**Thank you for your comments.**

Validity of the findings

I do not feel comfortable with the validity of these findings due to the uncertain nature of the perturbations analyses you have shown and poorly described.

**This comment is difficult to address without specifics on what the reviewer felt was uncertain and poorly described (above, the reviewer commented that the text was “clearly written”). The purpose of the sensitivity analysis here in this study, and in scientific modeling in general, is to determine how sensitive the original results and conclusions are to the model’s parameter values in situations where there is uncertainty in the original values, e.g. are there are large numbers of parameter value sets that would suggest different conclusions than the original results? The sensitivity analysis indicated that, when the analysis was performed with several thousand sets of feasible parameter values, loading rate was greater in the limb loss group 98% of the time, peak was greater 73% of the time, and impulse was greater 25% of the time (Line 235). From this result, we concluded that loading rate was likely greater in the limb loss group, peak may have been greater, and impulse was likely not greater (Line 249). We are open to other interpretations of this result as mentioned in the response to Reviewer #1, but this interpretation seems reasonable.**

It appears that these perturbations of model parameters were required to have results with statistically significant differences for both the peak and impulse measures since at iteration 0 for both of these, you show 0% limb loss>controls.

**This interpretation is incorrect. The original (unperturbed) parameters produced the results described in the text (greater loading rate and greater peak in the limb loss group). “Iteration 0” (which is actually iteration 1; there are no data at x=0 in Fig. 5) has no particular special relevance. It is not the statistical result from Fig. 4 with the original parameters, it is just the results with a perturbed set of parameter, one of several thousand. We have modified the figure caption in the revision to make this abundantly clear (Line 578). We were unsure why the reviewer had the impression that we presented an unmentioned perturbed-parameter result as the baseline result, but if the passage of text that lead to this conclusion can be pointed out, we can endeavor to clarify it.**

How do you know that the perturbations are realistic and physiological.

**Coefficient of variation of 10% is a reasonable general estimate for these types of parameters (e.g. Hasson & Caldwell) given that this is a fairly homogenous population (young male military service members). The text was revised with this information (Line 203).**

Many other studies do not perform perturbations to the model and would thus not show any statistical difference for these measures.

**Then that is a strength of this study that this statistical analysis was included.**

It seems suspect that the standard scaled model does not support your conclusions regarding impulse and peak.

**Please see the response three comments above. We think the reviewer has misread or misinterpreted the data shown in Fig. 5.**

I think a detailed description of which parameters were required to be perturbed in order to obtain the significant differences between your groups for the peak and impulse measures and an analysis of whether this perturbation is physiological/realistic should be reported, especially if it is physiological and is thus something that other researchers should also replicate to ensure their studies perturb the same measures.

**We perturbed all parameters relevant to the conclusions of the study, i.e. those that affect the contact force model output. Sensitivity of the model to individual parameter values within this set was not a purpose of this study. Modelers should always make parameter choices with thought and care and assess the sensitivity of their results to these choices. We added a comment to the Conclusions paragraph that sensitivity to individual particular parameters may be relevant for future work (Line 340)**

Comments for the Author

This is a good study that requires a little more work for me to suggest it as being methodologically sound:

"PeerJ evaluates articles based only on an objective determination of scientific and methodological soundness, not on subjective determinations of 'impact,' 'novelty' or 'interest'"

I have attempted to follow the PeerJ guidelines regarding review diligently and effectively and have provided comments based on their suggested points. I have classified this paper as major revision primarily due to the uncertain appropriateness and completeness of some of the methods and minor technicalities regarding the lack of raw data and missing funding disclosure.

This is a good initial attempt to model the behaviors of limb loss subjects relative to controls in order to better understand and highlight the possible bio-mechanisms for which limb-loss subjects experience higher rates of knee OA. The authors clearly demonstrate knowledge of the issues surrounding the phenomena. However I believe the modelling efforts are not solid enough to say with any certainty that the results are accurate, primarily due to the inherent uncertainty of muscle co-contraction and activation during gait and the sensitivity of medial/lateral contact force forces and ratios to the muscle force estimations. Obviously you point this out and perhaps that is all you can do. I think at the very least however that a static optimization method should be employed to estimate muscle forces( not mentioned in methods).

**We did not only point out the issue of co-contraction, we described why it is not a major issue in this study (Line 166). If there is an argument for why our description is erroneous, misleading, or shortsighted, please present it and we will endeavor to address it.**

**Static optimization (SO) would not address the above issue of co-contraction. Systematically underestimating co-contraction is a well-known criticism of SO (e.g. Gottlieb, 2000; Ait-Haddou et al., 2000). Requesting that the entire study be repeated with SO instead of the present inverse dynamics reduction approach requires a supporting technical basis (i.e. why would static optimization be expected to produce results that would lead to a fundamentally different conclusion?). There is no obvious reason why this would be expected. For example, a recent study on the ankle/Achilles load showed that SO and inverse dynamics results are similar (Kernozek et al., 2016). Discussion on this topic was added (Line 266).**

Also, I would assume these are soldiers of relatively large muscle mass and I would think that some level of muscle scaling may be warranted based on EMG/dynamometer testing or other means if possible, e.g. the standard scaled model muscle maximums may not be appropriate for these subjects.

**This is why we did the sensitivity analysis; it assesses how sensitive the results are to the assumptions made about parameter values. A specific statement on this purpose was added (Line 200).**

**Uniform scaling of the PCSA values does not affect results in a model like this (or in static optimization), they can all be multiplied or dividing by any particular scaling factor and the results will be identical. The ratio of PCSAs within a muscle group (e.g. biceps femoris vs. semimembranosus vs. semitendinosus) does make a difference; this is again why the PCSAs were included in the sensitivity analysis.**

Perhaps it would be much easier and less modelling-focused and parameter-sensitive if you could simply show the inverse dynamic flexion and adduction moments/angles and compare the peaks/impulse/rate of these. Increased flexion moment seems to indicate increased muscle activation based on yours and other peoples findings and perhaps this is enough to show with a statement regarding increase muscle activation about the knee is likely to compress the joint more and thus increase the joint reaction force. As of now you are assuming the muscle activation to be dependent solely on flexion angle and flexion moment, and then calculating medial force using those muscle activation estimates, the assumed relative proportioning of muscle maximums, and the adduction moment. Are you confident that the muscles do not depend on the adduction moment as well? It seems possible given the broad range of muscles around the knee that could contribute in varying degrees to either adduction or abduction.

**Muscle force (not activation) is the output variable in this model. The quadriceps provide the great majority of the load, and their moment arm is much larger in the sagittal plane than in the frontal plane, so yes, we are confident that the results do not depend on the use of the flexion moment alone to calculate muscle forces. If this assumption was faulty, we would expect muscle forces that do not agree with EMG timing, and contact forces in the control subjects that do not resemble the shape and magnitude of in vivo forces (e.g. Kutzner et al., 2013), neither of which was the case in these results. These comparisons were presented in the original manuscript (Line 218).**

**The reviewer’s suggestion above of using SO would not address these concerns. SO as it is typically done does not include the non-sagittal moments at the knee in its constraints; doing so leads to either infeasible solutions or solutions with unrealistically large muscle and contact forces because these moments are not predominantly from muscle forces (Glitsch & Baumann, 1997). This is one of the main reasons why the adduction moment is such a widely used indirect marker of medial knee loading (e.g. Foroughi et al., 2009).**

**Our goal was to assess loading of the medial knee, not the net moments about the knee. This analysis indeed requires some assumptions e.g. on moment arms, co-contraction, muscle parameters etc., but such assumptions are necessary in some fashion in any study on internal musculoskeletal loading, including those that use only joint moments. Using joint moments instead of contact forces would replace these quantitative modeling qualitative speculation over how moments relate to medial loading, which requires assumptions about co-contraction, moment arms, etc. (i.e. the same assumptions we had to make anyway).**

Also, are you confident that the distribution of PCSA given in the standard model based on a limited cadaveric study is reflective of your special population group? It seems unlikely given the increased musculature demand and training required for basic training relative to the general population.

**We again emphasize that the sensitivity of the results to PCSA values was included in the sensitivity analysis (Line 195).**