Journal of Clinical and Translational Research
Peer review process file 05.201901.001

# Racial and ethnic disparities in surgical amputations following serious musculoskeletal infections in a diverse New Mexico cohort

Martha L. Carvour, Allyssa Chiu, Kimberly Page

*Corresponding author:*
*Martha L. Carvour*
*Division of Epidemiology, Biostatistics, and Preventive Medicine, Department of Internal Medicine, University of New Mexico, Albuquerque, NM, United States*

Handling editor:
Michal Heger
*Department of Experimental Surgery, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands*

Review timeline:

Received: 4 July, 2018
Editorial decision: 4 September, 2018
Revision received: 26 September, 2018
Editorial decision: 28 October, 2018
Revision received: 24 November, 2018
Editorial Decision: 17 December, 2018
Revision received: 30 December, 2018
Editorial decision: 9 January, 2018

1st Editorial decision

Date: 04-Sep-2018

Ref.: Ms. No. JCTRes-D-18-00016
Racial and Ethnic Disparities in Surgical Amputations Following Serious Musculoskeletal Infections in a Diverse New Mexico Cohort
Journal of Clinical and Translational Research

Dear Dr. Carvour,

Reviewers have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.

For your guidance, reviewers' comments are appended below.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript.Also, please ensure that the track changes function is switched on when implementing the revisions. This enables the reviewers to rapidly verify all changes made.

Your revision is due by Oct 04, 2018.

To submit a revision, go to https://jctres.editorialmanager.com/ and log in as an Author. You

will see a menu item call Submission Needing Revision. You will find your submission record there.

Yours sincerely

Michal Heger
Editor-in-Chief
Journal of Clinical and Translational Research

Reviewers' comments:

Reviewer #1:

Dear authors,

Firstly, I must apologize on behalf of the editorial board that it has taken so long to process your manuscript. A total of 16 reviewers were invited, of which zero were willing to accept the review request. Consequently, I took over as reviewer. This means that any anonymity is no longer the case as it would be with all other manuscripts. However, I do not have a competing position in the field so there is no real bias. Moreover, I have taken the liberty to implement corrections in an already well-written manuscript. The track changes function was engaged so that you can see where changes were made.

MAJOR / GENERAL

1. Since this is an epidemiological/sociodemographics study, please ensure that the statistics part in section 2.4 provides full detail on which statistical parameters were measured and, where applicable, how the data were processed for analysis. Also, why were 6 years of data chosen to permit stratification? Why were 10 covariates selected for the multivariable model, and did the authors pay attention to potential overfitting issues as laid out in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4178069/? Did you perform any calculations to arrive at this number, such as a power analysis? Also, how were the reference categories selected in the multivariate analyses? It may be useful to lay out the analyses in a schematic, such as clinical trial reports often do with patient inclusion and exclusion. I will present the statistical case to one of our editorial board members who is an expert is biostatistics and epidemiology, and what to do that with a complete case presentation.

2. Section 1.2: it is postulated that "such complex clinical processes could result in different disparities across the phases of prevention and treatment." Firstly, do the differential diagnosis and corollary treatment plan not eliminate the potential disparities well enough? Secondly, you make an assumption that the complexity of clinical presentation and management of musculoskeletal infections could result in alleged disparities. You refer to your previous work on diabetic foot pathologies as underpinning of these statements, but do not address the main findings of that study to serve as the underpinning. So it is warranted to present your argument ("such complex clinical processes could result in different disparities across the phases of prevention and treatment") using the key findings of the previous study (ref 3) as premises. Thirdly, this section should provide the rationale for stratifying the patient cohorts as you did (i.e., based on ethnicity).

3. The last paragraph of section 1.2 should end with the hypothesis to clarify the purpose of

this study.

MINOR / SPECIFIC

1. Section 1.1: please explain what the potential healthcare disparities are that need to be identified and mitigated (e.g., by defining what a health disparity is). The phrasing is hard to understand without appropriate contextualization. Based on what do you claim that there are disparities at all? I am not questioning the validity of the statement, but asking you to provide more background and description.

2. Figure 1: the figure caption states the frequency of "DM, surgical intervention, and amputation", while the figure legend states "amputation, any procedure, and DM." This is very confusing, also because an amputation is in fact a surgical intervention. Please define what is meant by "any procedure" in the caption (as you did in the Materials & Methods) and synchronize the nomenclature.

---

Authors' rebuttal

**Responses to Reviewers:**
**Authors' responses appear in bold.**

Reviewer #1:

Dear authors,

Firstly, I must apologize on behalf of the editorial board that it has taken so long to process your manuscript. A total of 16 reviewers were invited, of which zero were willing to accept the review request. Consequently, I took over as reviewer. This means that any anonymity is no longer the case as it would be with all other manuscripts. However, I do not have a competing position in the field so there is no real bias. Moreover, I have taken the liberty to implement corrections in an already well-written manuscript. The track changes function was engaged so that you can see where changes were made.

**Thank you for your thorough and thoughtful review as well as your implementation of formatting corrections within the draft. We have accepted your tracked changes, and we have further revised the manuscript in response to each of the items below.**

MAJOR / GENERAL

1. Since this is an epidemiological/sociodemographics study, please ensure that the statistics part in section 2.4 provides full detail on which statistical parameters were measured and, where applicable, how the data were processed for analysis.
**Section 2.4 has been revised with further details about the modeling procedures.**
Also, why were 6 years of data chosen to permit stratification? Why were 10 covariates selected for the multivariable model, and did the authors pay attention to potential overfitting issues as laid out in https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4178069/? Did you perform any calculations to arrive at this number, such as a power analysis?
**In the revised version, we have clarified the rationale for including multiple years of data and have connected this rationale with the objectives of the study. The 10-covariate number is based on existing power analyses/power simulations published in**

**the reference cited and frequently applied for similar models. This is the power analysis underlying the study. [In fact, some have proposed that this criterion is conservative (Vittinghoff and McCulloch, 2006, DOI 10.1093/aje/kwk052).]**
**Importantly, this does not reflect a required selection of 10 covariates; rather, 10 is the maximum number of covariates statistically supported by 100 surgical outcomes ("up to 10 covariates"). In other words, if we aimed for at least 100 outcomes in our dataset based on the number of years of data captured, this would support a multivariable model with up to 10 covariates.**
**We did not force 10 (or more) covariates into either model. This itself helps to prevent overfitting. Variables were included or excluded based on the strength of associations with the outcomes, as defined using the p-value thresholds outlined in the text. We sought models that provided the best combination of prediction and parsimony, rather than models that tightly fit (or overfit) the data. In the Results section, we also provide cstatistics for each model, the values of which demonstrate predictive value but not tight or near-perfect fitting of the data.**
Also, how were the reference categories selected in the multivariate analyses?
**We have added more information about the selection of reference categories.**
It may be useful to lay out the analyses in a schematic, such as clinical trial reports often do with patient inclusion and exclusion. I will present the statistical case to one of our editorial board members who is an expert is biostatistics and epidemiology, and what to do that with a complete case presentation.
**Figure 2 summarizes eligibility for the two main predictive models featured in the text.**

2. Section 1.2: it is postulated that "such complex clinical processes could result in different disparities across the phases of prevention and treatment." Firstly, do the differential diagnosis and corollary treatment plan not eliminate the potential disparities well enough?
**We propose that healthcare disparities are often related to differential interactions with the healthcare system and are not fully eliminated by diagnosis and treatment. That question (at least in part) is under investigation in this study. Here, diagnosis and treatment—as measured by diabetes control and surgical access in this example—do not fully explain disparities in amputation. We have clarified our approach in Section 1.2.**
Secondly, you make an assumption that the complexity of clinical presentation and management of musculoskeletal infections could result in alleged disparities. You refer to your previous work on diabetic foot pathologies as underpinning of these statements, but do not address the main findings of that study to serve as the underpinning. So it is warranted to present your argument ("such complex clinical processes could result in different disparities across the phases of prevention and treatment") using the key findings of the previous study (ref 3) as premises.
**We propose that this is true, but this study is intended to evaluate this view. The conceptual paper referenced (ref 3) provides some epidemiological theory behind this view and a conceptual basis for the hypothesis of this study. We have revised the section to provide better context for this information.**
Thirdly, this section should provide the rationale for stratifying the patient cohorts as you did (i.e., based on ethnicity).
**We incorporated some text about this issue, including the importance of understanding racial/ethnic disparities in our clinical population. In the predictive modeling laid out in the text, multiple sociodemographic variables (e.g., age, sex/gender, and race/ethnicity) were eligible predictors, using traditional epidemiological strata.**

3. The last paragraph of section 1.2 should end with the hypothesis to clarify the purpose of this study.
**A clarifying hypothesis statement has been added.**

MINOR / SPECIFIC

1. Section 1.1: please explain what the potential healthcare disparities are that need to be identified and mitigated (e.g., by defining what a health disparity is). The phrasing is hard to understand without appropriate contextualization. Based on what do you claim that there are disparities at all? I am not questioning the validity of the statement, but asking you to provide more background and description.

**This section has been revised to provide a better working definition of disparity and to foreshadow our goal to investigate whether disparities exist, which is also better outlined alongside the hypothesis in section 1.2.**

2. Figure 1: the figure caption states the frequency of "DM, surgical intervention, and amputation", while the figure legend states "amputation, any procedure, and DM." This is very confusing, also because an amputation is in fact a surgical intervention. Please define what is meant by "any procedure" in the caption (as you did in the Materials & Methods) and synchronize the nomenclature.

**The caption has been modified to more closely parallel the verbiage in the text. The caption is now more synchronous with the terms in the legend (where the terms are shortened slightly for readability in the figure).**

---

2nd Editorial decision

Date: 28-Oct-2018

Ref.: Ms. No. JCTRes-D-18-00016R1
Racial and Ethnic Disparities in Surgical Amputations Following Serious Musculoskeletal Infections in a Diverse New Mexico Cohort
Journal of Clinical and Translational Research

Dear Dr. Carvour,

Reviewers specialized is statistics and epidemiology have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.

For your guidance, reviewers' comments are appended below. Please note that, following the original submission, the editorial board announced in the comments on the original submission that your manuscript must be subjected to further scrutiny by experts in statistics. The editorial board also proclaimed that submitting a revised manuscript in the first round was no guarantee for acceptance. We therefore kindly ask you to implement the reviewers' remarks to the fullest extent possible so as to arrive at a maximally sound textual representation of your work.

If you decide to revise the manuscript, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript. Also, please ensure that the track changes function is switched on when implementing the revisions. This enables the reviewers to rapidly verify all changes made.

Your revision is due by Nov 27, 2018.

To submit a revision, go to https://jctres.editorialmanager.com/ and log in as an Author. You will see a menu item call Submission Needing Revision. You will find your submission record there.

Yours sincerely, also on behalf of the JCTR editorial board,

Michal Heger
Editor-in-Chief
Journal of Clinical and Translational Research

Reviewers' comments:

Reviewer #2: The current manuscript presents the results of a retrospective study on predictors of various outcomes in a cohort of patients with serious musculoskeletal infections. The manuscript is in general well-written and results are presented with a balanced discussion of the strenghts and weaknesses of the study. However, I'm afraid that the presented results might be strongly biased due to the way that missing values were handled.

MAJOR COMMENTS
1. Figure 2: According to the legend "The final multivariable model included patients with available data for all variables included in the model." That is, only complete cases were analysed. However, it is well-known that such an analysis only gives unbiased estimates if the data is missing completely at random (MCAR). If complete cases are systematically different from the sample as a whole, analysing only the complete cases will give biased estimates. The data presented actually seems to provide strong evidence that the data is not MCAR. Let us consider Figure 2. Here, of the 1694 patients (308 amputations) eligible for the predictive model only 1294 (289 amputations) were included in the final model. The reduction in the number of patients is mainly driven by the 'comorbidity' variables such as rheumatoid arthritis for which there are only 1325 complete cases (Table 1). However, note that while the number of patients is reduced by 1694-1294=400 that is 23.6%, the number of amputations is only reduced by 308-289=19 that is 6.1%. This implies that of the patients for whom information about comorbidities is missing far fewer than expected if the data was MCAR underwent an amputation. This is just one example, but probably something similar is the case for the other variables with missing data.The standard soluton is to use multiple imputation and this is what I would urge the authors to do.
Another problem with the analysis is that when performing forward or backward selection the number of complete cases and therefore the size of the cohort was allowed to vary with the variables included in the model. The authors state for example in Section 3.3 that "However, the model with rheumatoid arthritis had a larger available cohort size (N = 1294 compared to N = 914 with the immunosuppressants variable)." This is wrong and means that successive models in the selection procedure cannot be properly compared. In a complete case scenario one can only include those patients with data available for all variables that are eligible for

inclusion in the multivariable model. Again, this problem is avoided when using multiple imputation.

MINOR COMMENTS
1. Section 1.2: Please clarify what is meant by "... for instance, amputation rates by race/ethnicity that are not proportionately demonstrated for surgical treatment or diabetes control."
2. Section 2.3: The auhtors report HbA1c in percentages. However, in many European countries the current standard is to report mmol/mol. Please also provide these values.
3. Section 2.4: Did the authors evaluate whether it is reasonable to include the variable 'age' as a continous predictor instead of a categorical predictor based on age intervals?
4. Table 1: Please also include the odds-ratios for the univariable analyses.
5. Figure 1: According to the legend "P-values are from a Chi-square test". Since all other analyses were done using logistic regression,I suggest to just present logistic regression results here for reasons of consistency.
6. Sections 3.2 and 3.3: For the variable 'infection type' the reference category chosen was 'infective/septic myositis'. This was done based on the criterion specified in Section 2.4: "... the category with the lowest odds served as the reference group". However, according to Table 1 there are only 19 patients with infective/septic myositis. It is well-known that a small reference category might lead to instable estimates of the odds-ratios and the associated confidence intervals. I suggest to use one of the other two categories (with 964 and 711 patients) as reference category in this case.
7. Table 2: Diabetes mellitus (p=0.0979) was included in the multivariable model. This is in agreement with the statement in Section 2.4 that "Variables with a p-value of < 0.10 in the multivariable model were eligible to remain in the model," However, usually only variables with a p-value <0.05 are included in a multivariable model. Why did the authors decide to deviate from standard practice?
8. The statements on the statistical power of the study are confusing. On p.10 it is stated that "Based on existing guidance about statistical power in multivariable regression models, an estimated minimum of 100 procedures was expected to provide adequate power for a multivariable model with up to 10 covariates". This is indeed a commomly used rule-of-thumb. However, it would be helpful if the authors would relate this to the actual number of parameters and the actual number of procedures. For example with 308 amputations according to this rule 30 parameters could be included. According to Table 1 the largest model in this case involved 17 parameters and the final model involved 12 parameters (Table 3). Note that I used the word 'parameter' since a categorical variable with 5 categories, for example race/ethnicity (Table 1), leads to 4 parameters in the logistic regression model.
9. Could the authors comment upon the possible bias introduced due to the fact that "Procedures and amputations recorded within the first three months after the infection diagnosis were included." (Section 2.3). This means that procedures and amputations after the first months are not taken into account. Why?

Reviewer #3: The authors claim their sample size is large enough because there are ten events per variable used in their final model. But this is not computed correctly. The ratio should be based on the total number of variables considered, which in this case is 20, not the number in the final model. There are 5 events per variable not 10.

"The number of EPV is the number of events divided by the number of predictor variables considered in developing the prediction modell."

Austin, P.C., and Steyerberg, E.W. (2014). Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. Stat Methods Med Res 26: 796-808.

Since the EPV is only 5, I don't think the results are really meaningful.

Beyond that, a key variable is diabetes but the analysis seems to combine people with early diabetes and long-standing diabetes

---

Authors' rebuttal

**Authors' Responses to Reviewers:**

**The authors would like to thank the expert biostatistical reviewers for their thoughtful impressions and suggestions on our manuscript. We have revised our manuscript, and we offer our direct responses to each comment below. Our responses appear in bold print.**

Reviewer #2: The current manuscript presents the results of a retrospective study on predictors of various outcomes in a cohort of patients with serious musculoskeletal infections. The manuscript is in general well-written and results are presented with a balanced discussion of the strenghts and weaknesses of the study. However, I'm afraid that the presented results might be strongly biased due to the way that missing values were handled.

**We appreciate this positive reception to several aspects of our manuscript. We have addressed the concerns about bias below.**

MAJOR COMMENTS
1. Figure 2: According to the legend "The final multivariable model included patients with available data for all variables included in the model." That is, only complete cases were analysed. However, it is well-known that such an analysis only gives unbiased estimates if the data is missing completely at random (MCAR). If complete cases are systematically different from the sample as a whole, analysing only the complete cases will give biased estimates. The data presented actually seems to provide strong evidence that the data is not MCAR. Let us consider Figure 2. Here, of the 1694 patients (308 amputations) eligible for the predictive model only 1294 (289 amputations) were included in the final model. The reduction in the number of patients is mainly driven by the 'comorbidity' variables such as rheumatoid arthritis for which there are only 1325 complete cases (Table 1). However, note that while the number of patients is reduced by 1694-1294=400 that is 23.6%, the number of amputations is only reduced by 308289=19 that is 6.1%. This implies that of the patients for whom information about comorbidities is missing far fewer than expected if the data was MCAR underwent an amputation. This is just one example, but probably something similar is the case for the other variables with missing data.The standard soluton is to use multiple imputation and this is what I would urge the authors to do.
**We appreciate this reviewer's important insight into the potential impact of the difference in sample size between the total cohort and the modeled cohort where comorbidities were included. In response to this reviewer's input, we pursued a thorough evaluation of the reasons for this difference.**

**In summary, we concluded that absent values for comorbidities or medications should best be coded as "0", representing the absence of the comorbidity or medication. We have revised our paper to reflect inclusion of these full records for the cohort. Because of this, we did not pursue multiple imputation. Although we agree that multiple imputation would have been a valuable approach had these been genuinely missing values, this approach would likely have imputed inaccurate values in this case (that is, in the case where we know values should be "0"). Ultimately, we believe that the revisions made as a result of this reviewer's comments have strengthened our paper and increased our confidence in the validity of our core findings.**

**If further detail about the steps we used to reach these conclusions is desired, we have provided the following supplemental information:**

**First, we retraced the methods for the data query and data linkage that yielded this difference in sample size. As described in our Methods, we obtained data from our local clinical data warehouse that serves our hospital and clinic system. Using the case definition for musculoskeletal infection (as defined in the paper), a study cohort was selected from the clinical records in the warehouse. Then, all associated diagnoses, procedures, medications, and laboratory values corresponding to our specific data request (e.g., specific comorbidities) were pulled for all members of the study cohort. This resulted in analytic subfiles (e.g., one or more comorbidities files) that could be linked to the original cohort file using the unique identifier assigned by the data warehouse. The analytic subfiles provided to us contained records for members of the study cohort, wherever a match with one or more of our pre-specified covariates occurred. We have confirmed with our data source that the absence of a cohort member in one of those subfiles represents the absence of specific matching codes (e.g., the absence of a specific comorbidity) and not missing data.**

**Second, we considered whether some of these ICD-based records could still be subject to significant misclassification by underdiagnosis. We assessed this using the diabetes variable, where we had already allowed for a more complex definition (as described in the Methods) that used ICD codes +/- medications or laboratory values. Upon comparison of ICD-based coding +/- medication-based or laboratorybased criteria, we found that ICD coding was quite accurate (only about 3.3% of the cohort with diabetes would have been missed with ICD coding alone). We've added a note about this to our revised Discussion.**

**Third, we assessed the difference between the original results and those with "0" coding for these previously excluded records. As we have shown in our revised paper, this does not produce markedly different findings for our study. There are subtle differences in some estimates, as expected, and in the preliminary eligibility of a few variables for the multivariable procedure. The sepsis variable is also excluded from the amputation model. (Clinically, this is actually more plausible. Moreover, sepsis—for various billing reasons—is very unlikely to be under-coded; so most of these "0" values were probably true zeroes.)**

**As a further step, we obtained a consultation with a senior biostatistician (Dr. Clifford Qualls, now cited in our acknowledgements) at our institution to discuss our findings and our thought process regarding the best way to handle this issue. Our revised findings and results reflect both our own review and Dr. Qualls' recommendations.**

**As a secondary matter, we note that the HbA1c data in the cohort are not complete. This likely represents truly missing data—that is, where a measurable value could always be knowable but where the blood test was not collected. However, we only use**

**this measure as a subanalysis, and we have indicated in both the original and current versions of the manuscript that this variable must be viewed with caution.**

Another problem with the analysis is that when performing forward or backward selection the number of complete cases and therefore the size of the cohort was allowed to vary with the variables included in the model. The authors state for example in Section 3.3 that "However, the model with rheumatoid arthritis had a larger available cohort size (N = 1294 compared to N = 914 with the immunosuppressants variable)." This is wrong and means that successive models in the selection procedure cannot be properly compared. In a complete case scenario one can only include those patients with data available for all variables that are eligible for inclusion in the multivariable model. Again, this problem is avoided when using multiple imputation.
**This particular issue was resolved by the changes described above, as the cohort size was stable across these variables and as the backward and forward selection procedures for both models produced the same results.**

MINOR COMMENTS

1. Section 1.2: Please clarify what is meant by "... for instance, amputation rates by race/ethnicity that are not proportionately demonstrated for surgical treatment or diabetes control." **This statement in the manuscript is revised for clarity.**

2. Section 2.3: The auhtors report HbA1c in percentages. However, in many European countries the current standard is to report mmol/mol. Please also provide these values.

**In the revised section 2.3 (and later in the manuscript where HbA1c values are used), we have included this conversion.**

3. Section 2.4: Did the authors evaluate whether it is reasonable to include the variable 'age' as a continous predictor instead of a categorical predictor based on age intervals?

**The continuous variable was the predictor of clinical interest. Categorization by age intervals would require selection of somewhat arbitrary cutoffs for this clinical scenario, and this may limit clinical interpretation. However, we did attempt natural log transformation of this variable to assess for any differences between findings; there were no significant differences in the outcomes of our models with log transformation.**

4. Table 1: Please also include the odds-ratios for the univariable analyses.

**These have been added. Note that reference categories for this table were selected based on those used in Tables 2 and 3 to permit more direct comparisons.**

5. Figure 1: According to the legend "P-values are from a Chi-square test". Since all other analyses were done using logistic regression,I suggest to just present logistic regression results here for reasons of consistency.

**We've adjusted the legend to reflect the use of logistic regression, which gives the same pvalues as those shown in the original figure. We also updated the figure using the larger available cohort size to be consistent with other changes (re: comorbidities) above. We prefer to graphically represent percentages (rather than odds ratios) here, since we expect this will be a useful complementary figure for some clinical readers who may be less familiar with regression methods.**

6. Sections 3.2 and 3.3: For the variable 'infection type' the reference category chosen was 'infective/septic myositis'. This was done based on the criterion specified in Section 2.4: "... the category with the lowest odds served as the reference group". However, according to Table 1 there are only 19 patients with infective/septic myositis. It is well-known that a small reference category might lead to instable estimates of the odds-ratios and the associated confidence intervals. I suggest to use one of the other two categories (with 964 and 711 patients) as reference category in this case.

**We have changed the reference category to osteomyelitis.**

7. Table 2: Diabetes mellitus (p=0.0979) was included in the multivariable model. This is in agreement with the statement in Section 2.4 that "Variables with a p-value of < 0.10 in the multivariable model were eligible to remain in the model," However, usually only variables with a p-value <0.05 are included in a multivariable model. Why did the authors decide to deviate from standard practice?

**In the revised analysis, the diabetes variable is no longer affected by this issue; the only variable with a p-value between 0.05 and 0.10 is rheumatoid arthritis (in Table 3). Clinically, the authors believe this variable adds some valuable information to the model, because of the different mechanisms and management issues involved in rheumatoid arthropathies.**

**We had opted to use the higher p-value cutoff of 0.10 in order to permit more predictors into the model, which can then be evaluated or validated in the future using larger databases. We've added a statement about this rationale in section 2.4 of the manuscript.**

8. The statements on the statistical power of the study are confusing. On p.10 it is stated that "Based on existing guidance about statistical power in multivariable regression models, an estimated minimum of 100 procedures was expected to provide adequate power for a multivariable model with up to 10 covariates". This is indeed a commomly used rule-of-thumb. However, it would be helpful if the authors would relate this to the actual number of parameters and the actual number of procedures. For example with 308 amputations according to this rule 30 parameters could be included. According to Table 1 the largest model in this case involved 17 parameters and the final model involved 12 parameters (Table 3). Note that I used the word 'parameter' since a categorical variable with 5 categories, for example race/ethnicity (Table 1), leads to 4 parameters in the logistic regression model.

**Thank you for this recommendation. We have revised the section to directly tie the number of outcomes to the number of parameters. We agree that up to about 30 parameters could be used for a model with about 300 amputations. In this case, the number of parameters (now 16, from 17 in the prior version, where the sepsis variable was eligible) is supported by the sample size.**

9. Could the authors comment upon the possible bias introduced due to the fact that "Procedures and amputations recorded within the first three months after the infection diagnosis were included." (Section 2.3). This means that procedures and amputations after the first months are not taken into account. Why?

**As we've noted in the revised Discussion section, this three-month cutoff primarily limits the interpretation of our findings to short-term outcomes and/or more acute infections. (Chronic or relapsing cases are clinically likely to differ in important ways.) We believe that inclusion of cases over longer periods could introduce several significant sources of bias that would be difficult to overcome with a retrospective study—chiefly, intervening clinical events, treatments, or courses of care, including those which may have occurred outside our single center, that could be difficult to accurately capture or define.**

Reviewer #3: The authors claim their sample size is large enough because there are ten events per variable used in their final model. But this is not computed correctly. The ratio should be based on the total number of variables considered, which in this case is 20, not the number in the final model. There are 5 events per variable not 10.

"The number of EPV is the number of events divided by the number of predictor variables considered in developing the prediction modell."

Austin, P.C., and Steyerberg, E.W. (2014). Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. Stat Methods Med Res 26: 796-808.
Since the EPV is only 5, I don't think the results are really meaningful.

**Using the Austin and Steyerberg definitions for events and events per variable, we calculate the following:**
**For the amputations model: The event number is approximately 300, which would support up to 30 parameters for an EPV of 10. The total number of eligible parameters is approximately 20. Thus, the actual EPV is approximately 300/20 or approximately 15. Reviewer #2 also indicated that such a ratio was adequate.**
**For the procedures model: The event number is approximately 500, if we take the view that the absence of a procedure is the modeled event (that is, the less frequent event). This would support up to 50 parameters for an EPV of 10.**
**For either model, adequate sample size was present to support approximately 20 predictive parameters, while maintaining an EPV of 10 or higher. We did provide some clarifying language about this issue in section 2.4.**

Beyond that, a key variable is diabetes but the analysis seems to combine people with early diabetes and long-standing diabetes.
**We agree that the timing and progression of diabetes (including progression from prediabetes to diabetes) is of interest in this clinical population. However, we believe this question can be more effectively addressed using a prospective, population-based study design, ideally with built-in screening protocols.**
**In a retrospective cohort, attempts to differentiate between early and late diabetes using diagnostic, laboratory, and/or pharmacy data can be difficult and bias-prone. For instance, a serious musculoskeletal infection could be the presenting syndrome of diabetes in some cases—particularly in those with poor access to healthcare or potential disparities in primary healthcare. In such cases, the actual onset of diabetes could have been much earlier than the clinical diagnosis was documented. In the limitations section of our revised Discussion, we have added a caveat about the unknown timing of DM.**

3rd Editorial decision

Date: 17-Dec-2018

Ref.: Ms. No. JCTRes-D-18-00016R2
Racial and ethnic disparities in surgical amputations following serious musculoskeletal infections in a diverse New Mexico cohort
Journal of Clinical and Translational Research

Dear Dr. Carvour,

Reviewers have commented on your paper. You will see that they still have concerns. Reviewer 3 has recommended a reject because the person feels that the conclusions are not supported by the data, and believes that 'statistical juggling' was performed to fit the data into a preconceived framework. This reviewer is a living legend in statistics and recommended a reject. Reviewer 2 is also a highly qualified expert in statistics and epidemiology, and recommended minor revisions.

For your guidance, reviewers' comments are appended below.

I am therefore tied between 2 horses going into opposite directions. I would like to give you another chance to process the remarks and make sure that all conclusions are unequivocally supported by the data. Both reviewers have been very elaborate in their thinking, which you should use to redirect the paper into the right direction. For us it is not so important what the conclusions are, but the conclusions have to be predicated on sound and valid premises. So the paper must in all instances exude this requested soundness and validity; we do not want anyone questioning the outcomes of your study because of technical inadequacies. Naturally, I as well as other editors are always looking for common ground between authors and reviewers, so we like to stay accommodative to you and find a proper solution to the raised issues.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript.Also, please ensure that the track changes function is switched on when implementing the revisions. This enables the reviewers to rapidly verify all changes made.

Your revision is due by Dec 31, 2018.

To submit a revision, go to https://jctres.editorialmanager.com/ and log in as an Author. You will see a menu item call Submission Needing Revision. You will find your submission record there.

Yours sincerely

Michal Heger
Editor-in-Chief
Journal of Clinical and Translational Research

Reviewers' comments:

Reviewer #2: I would like to thank the authors for the care taken into replying to my comments on the previous version of the manuscript and their efforts in reanalyzing the data and revising the text accordingly.

The decision taken by the authors to code missing values for comorbidities and medications as "0" seems justified given that "We have confirmed with our data source that the absence of a cohort member in one of those subfiles represents the absence of specific matching codes (e.g., the absence of a specific comorbidity) and not missing data." However, related to the issue of missing data I would like the authors to address the following points in the text:
1. In the Patients and Methods section elaborate on the statement on p.9 that "Diabetic peripheral neuropathy (or other neurological complications of diabetes), PVD, osteoarthritis, rheumatoid arthritis, renal disease (including chronic kidney disease and end stage renal disease), cirrhosis, obesity, and sepsis were defined as present if any ICD code corresponding to each respective diagnosis was recorded." Specifically:
a) Clearly state that this decision was taken based on the fact that absence of a matching ICD code was assumed to correspond to the absence of a comorbidity.
b) State that the same reasoning was applied to the type of medication (immunosuppressants/antibiotics).
c) Explicitly mention that for race/ethnicity 46 (=1694-1648) missing values were not imputed and that a complete case analysis was performed for the models that included this variable.
2. On p.20 it is stated that "Clinical data from other referring sites, including those affiliated with Indian Health Services facilities within the state, were not available for this study.Although we recognize that this may have influenced the completeness of some diagnostic data for AI/AN patients in the cohort, the effect on observed, UNM-based amputation rates for patients who were referred, diagnosed, or treated for musculoskeletal infections at our center may have been less significant than for other kinds of data (e.g., comorbidity diagnoses)." This seems at odds with the statement above regarding absence of a comorbidity. Please address this point in the Discussion.

OTHER COMMENTS
1. Figure 1: "P-values are from unadjusted logistic regression models for diabetes mellitus, any procedure, or any amputation," However,
- In the figure only the P-value for DM is mentioned and the legend with color key disappeared. Please correct.
2. Please check whether the following statements are correct, since they were unchanged compared to previous version despite the changes in the analysis:
- p.15: "... was associated with an aOR of 1.16 (95% CI: 1.08, 1.24) for undergoing a procedure."
- p.16: "... was associated with an aOR of 1.20 (95% CI: 1.13, 1.28) for undergoing an amputation."

Reviewer #3: The title and discussion make it clear what the main question of the study was: Were people of different races equally likely to get surgery for joint infections? You conclude: No

Table 1 shows that an unadjusted analysis shows that there is no systematic effect of race on the rate of any surgery or amputation.

But other variables (covariates) may confound the results.

Next you do multiple regression with variable selection to get rid of the variables that don't contribute much to the model. This kind of variable selection is frowned upon by many. It leads to models that fit too well, P values that are too small, parameter values that are too small (actually their absolute value is too small), and confidence intervals that are too narrow. References below.

But this analysis of all surgeries, Table 2, did not find that race was a significant factor. The table doesn't show race because it doesn't show "nonsignificant" variables.

So then you drill down on only amputations (Table 3) . The effect of race is barely "significant" with a P value of 0.04. The baseline group is defined to be native americans. Blacks were three times more likely to have an amputation, with a 95% confidence interval ranging from a 30% increase to a 6.7 fold increase. Note the very wide confidence interval. And note that Blacks only make up 14% of the study population, so the results aren't expected to be very precise. Hispanics had an odds ratio 1.65 with confidence interval ranging from 1.10 to 2.48. A large fraction of the study population is hispanic, so these are the key findings of the paper. But they result from a second analysis (after the first analysis didn't give the hoped-for results) using a variable selection procedure which is known to create too-small P values and too-high odds ratios, and has a P value of only 0.04, barely under the usual definition of "significance".

The paper purports to show that blacks and hispanics got more amputations after infection than whites or native americans (after correcting for diabetes and a bunch of other variables). I don't think the data supports the conclusion. The manuscript doesn't say that the analysis plan was preplanned or preregistered, so I imagine that they tried a few things. This means readers should be wary.

The conclusion only was found with variable selection multiple regression and only after the planned (it seems) regression failed and they searched for another regression (amputations only) that gave the results they want. And even then, the P value was barely below 0.05.

Links to information on the problems of variable selection:

https://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/26/lecture-26.pdf
https://onlinelibrary.wiley.com/doi/epdf/10.1111/tri.12895

---

Authors' rebuttal

**Authors' Responses to Reviewers:**

**We would again like to thank the reviewers for their time and input about the manuscript. We have revised the text in several locations, and our direct responses to the reviewers' input appear below in bold print.**

Reviewer #2: I would like to thank the authors for the care taken into replying to my comments on the previous version of the manuscript and their efforts in reanalyzing the data and revising the text accordingly.

The decision taken by the authors to code missing values for comorbidities and medications as "0" seems justified given that "We have confirmed with our data source that the absence of a cohort member in one of those subfiles represents the absence of specific matching codes (e.g., the absence of a specific comorbidity) and not missing data." However, related to the issue of missing data I would like the authors to address the following points in the text:

1. In the Patients and Methods section elaborate on the statement on p.9 that "Diabetic peripheral neuropathy (or other neurological complications of diabetes), PVD, osteoarthritis, rheumatoid arthritis, renal disease (including chronic kidney disease and end stage renal disease), cirrhosis, obesity, and sepsis were defined as present if any ICD code corresponding to each respective diagnosis was recorded." Specifically:

a) Clearly state that this decision was taken based on the fact that absence of a matching ICD code was assumed to correspond to the absence of a comorbidity.

b) State that the same reasoning was applied to the type of medication (immunosuppressants/antibiotics).

c) Explicitly mention that for race/ethnicity 46 (=1694-1648) missing values were not imputed and that a complete case analysis was performed for the models that included this variable.

**Thank you for your thorough review of these revisions, and thank you again for your insightful input during the first round, which helped us to look further into this issue. We have revised the Patients and Methods section (in 2.3 and 2.4) to reflect our classification process for ICD codes (part a above), medications (part b above), and procedures. We have also added information about the missing data for race/ethnicity (part c above).**

2. On p.20 it is stated that "Clinical data from other referring sites, including those affiliated with Indian Health Services facilities within the state, were not available for this study.Although we recognize that this may have influenced the completeness of some diagnostic data for AI/AN patients in the cohort, the effect on observed, UNM-based amputation rates for patients who were referred, diagnosed, or treated for musculoskeletal infections at our center may have been less significant than for other kinds of data (e.g., comorbidity diagnoses)." This seems at odds with the statement above regarding absence of a comorbidity. Please address this point in the Discussion.

**Thank you for this important perspective. We have adjusted the Discussion section to reflect some persisting uncertainty about these data, and we have reinforced our call for multicenter work to address this question.**

OTHER COMMENTS

1. Figure 1: "P-values are from unadjusted logistic regression models for diabetes mellitus, any procedure, or any amputation," However,

- In the figure only the P-value for DM is mentioned and the legend with color key disappeared. Please correct.

**It appears that this is related to a file conversion issue. We will re-upload our version with the full legend at the time of revision, and we would be happy to work with the editor/journal to ensure the file conversion is successful or to provide an alternative format.**

2. Please check whether the following statements are correct, since they were unchanged compared to previous version despite the changes in the analysis:

- p.15: "... was associated with an aOR of 1.16 (95% CI: 1.08, 1.24) for undergoing a procedure."
- p.16: "... was associated with an aOR of 1.20 (95% CI: 1.13, 1.28) for undergoing an amputation."

**Yes, thank you for your attention to this detail. Both the aORs/CIs for HbA1c were confirmed in the revised models and were unchanged (or at least, rounded to the same values).**

Reviewer #3: The title and discussion make it clear what the main question of the study was: Were people of different races equally likely to get surgery for joint infections? You conclude:
No
**Respectfully, this is incorrect on multiple fronts. We regret that these concerns were not raised during the first round of review; we would have welcomed the opportunity to clarify then as we do now.**

1. **As we outline extensively in both the Introduction and Discussion, we are interested in whether access to care (or patients' experience with care) differs in important ways at different steps in the care process. This is why we sought to compare "the predictors of surgical intervention - a general proxy of treatment access at the time of infection - to the predictors of surgical amputation - a serious complication of infection."**
    a. **Our group has a demonstrated interest in the issue of limb salvage or limb preservation, not simply in surgery overall. These two entities differ in important way. For instance, surgical debridement after many musculoskeletal infections is a clinically desirable outcome, whereas amputation is not.**
    b. **Although prior research has suggested that disparities in amputation rates may exist, we understood that evidence of disparities in amputation rates in our population, if these were found, might raise an important and valid question: Are these differences simply a reflection of differences in access to surgical care overall? This question is the basis of our comparative models.**
2. **The statement that our primary focus was on "joint infections" is also inaccurate. We clearly indicate that we are studying bone, joint, and muscle infections. This, too, is a deeply important distinction and one specifically related to the risk of amputation, as our findings about infection type across the two models confirm— findings which, in fact, lend strong clinical validity to the two comparative models.**
3. **Race/ethnicity was among the sociodemographic variables we considered as a potential source of disparity. However, as we describe, apparent racial/ethnic disparities in amputation rates, if these were found, might raise important questions about other preceding steps in the care process. For instance: Is the difference**

**simply a reflection of disparities in diabetes, peripheral vascular disease, etc.? As we explain in our paper, this is the basis for our approach, in which these other factors are specifically considered.**

Table 1 shows that an unadjusted analysis shows that there is no systematic effect of race on the rate of any surgery or amputation.

**From Table 1, in the unadjusted model, that is correct; there is no observed effect of race/ethnicity on the composite surgical outcome. Also from Table 1, however, there is an apparent effect of race/ethnicity on the amputation outcome.**

But other variables (covariates) may confound the results.

Next you do multiple regression with variable selection to get rid of the variables that don't contribute much to the model. This kind of variable selection is frowned upon by many. It leads to models that fit too well, P values that are too small, parameter values that are too small (actually their absolute value is too small), and confidence intervals that are too narrow. References below.

**As described above, one goal of our multivariable modeling procedure was to address the question: If a racial/ethnic disparity appears in the unadjusted model, could this simply reflect disparities in other predictors—such as diabetes, peripheral vascular disease, etc.? (Again, understanding whether disparities differ at different steps in the clinical process is a foundational premise of the paper.)**

**If the apparent predictive value of race/ethnicity from an unadjusted model disappeared completely in a multivariable model with other significant predictors, we would have evidence that disparities in amputation rates, if these exist, might be driven by (or at least correlated with) disparities in other, potentially intervenable predisposing conditions. This would be a clinically and epidemiologically valuable observation—and one more amenable to immediate intervention. As we explain, we observe that some, but not all, of the racial/ethnic disparity is related to these other outcomes. This is an important finding.**

**We certainly recognize that limitations are inherent in any study design or statistical analysis. In this study, we have used established methods for predictive model development, but we clearly state that limitations exist and that the purpose of these procedures was to permit ongoing evaluation and refinement in future work and larger datasets.**

But this analysis of all surgeries, Table 2, did not find that race was a significant factor. The table doesn't show race because it doesn't show "nonsignificant" variables.

**By convention, Table 2 included all variables for the final model for the surgical outcome. We followed the procedures described in the Patients and Methods section and did not force other variables (e.g., race/ethnicity) into that model.**

So then you drill down on only amputations (Table 3) . The effect of race is barely "significant" with a P value of 0.04. The baseline group is defined to be native americans. Blacks were three times more likely to have an amputation, with a 95% confidence interval ranging from a 30% increase to a 6.7 fold increase. Note the very wide confidence interval. And note that Blacks only make up 14% of the study population, so the results aren't expected to be very precise. Hispanics had an odds ratio 1.65 with confidence interval ranging from

1.10 to 2.48. A large fraction of the study population is hispanic, so these are the key findings of the paper. But they result from a second analysis (after the first analysis didn't give the hoped-for results) using a variable selection procedure which is known to create too-small P values and too-high odds ratios, and has a P value of only 0.04, barely under the usual definition of "significance".

**As stated above, we present this as a preliminary model which should be evaluated and refined in future work and larger datasets. We concur that the p-value is just below the significance cutoff. It is our duty to report this value accurately in any case, as we have done here. As reflected by our comments in the manuscript and above, we view this as a residual disparity; some, but not all, of the apparent (unadjusted) racial/ethnic disparity is related to other predisposing factors such as diabetes.**
**It is important to note (as we indicate in our Discussion) that this pattern of disparity for Hispanic and Black non-Hispanic patients is also generally consistent with work at other centers. For researchers and clinicians in this field, our findings are an important contribution to the available literature on this subject, especially given the majorityminority population we serve and the higher proportion of American Indian/Alaskan Native patients in the cohort.**

The paper purports to show that blacks and hispanics got more amputations after infection than whites or native americans (after correcting for diabetes and a bunch of other variables). I don't think the data supports the conclusion. The manuscript doesn't say that the analysis plan was preplanned or preregistered, so I imagine that they tried a few things. This means readers should be wary.

The conclusion only was found with variable selection multiple regression and only after the planned (it seems) regression failed and they searched for another regression (amputations only) that gave the results they want. And even then, the P value was barely below 0.05.

**Although we welcome scientifically rigorous criticism from both reviewers and readers, the speculations above are incorrect. As we explain in the paper, we are interested in whether the patterns differed for procedures overall and for amputations specifically. We obtained data for a multi-year cohort of patients with musculoskeletal infections, and we requested procedural data for that cohort in order to ask clinically relevant questions. As we also explain, we made determinations about the necessary sample size for the amputation model before deciding whether we could set out to answer that question.**
**We deeply understand the need to test and refine all results, as we have proposed, because our aim is ultimately to develop accurate, actionable information. We regularly see patients at risk of losing life and limb from these infections, and we want to see those outcomes in our clinical population improve over time. Respectfully, "the results [we] want"—the "hoped-for" results—are any results that begin to fill in the gaps, to understand who loses limbs and who doesn't, and to help us get better at improving these outcomes for our patients.**

Links to information on the problems of variable selection:

https://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/26/lecture-26.pdf
https://onlinelibrary.wiley.com/doi/epdf/10.1111/tri.12895

4th Editorial decision

Date: 09-Jan-2018

Ref.: Ms. No. JCTRes-D-18-00016R3
Racial and ethnic disparities in surgical amputations following serious musculoskeletal infections in a diverse New Mexico cohort
Journal of Clinical and Translational Research

Dear authors,

I am pleased to inform you that your manuscript has been accepted for publication in the Journal of Clinical and Translational Research.

We hold the opinion that you have adequately improved the manuscript after 3 rounds of revisions and that you properly rebutted there where necessary. Although it was not possible to fully satisfy both reviewers, the critical comments raised by reviewer #3 have been extensively addressed in the Discussion section, where readers are cautioned about the possible pitfalls of the statistical model used and the data that were obtained. Moreover, your conclusions are supported by other literature and your clinical experience. Finally, all reviewer comments and author responses will be made available online as metadata, so interested readers will get a chance to study the argumentation of both sides.

You will receive the proofs of your article shortly, which we kindly ask you to thoroughly review for any errors.

Thank you for submitting your work to JCTR.

Kindest regards,

Michal Heger
Editor-in-Chief
Journal of Clinical and Translational Research

Comments from the editors and reviewers: