# Response to reviewers

## Referee: 1

* *"This paper presents an investigation of the relationship between NEFAs measured at baseline and subsequent development of markers of insulin sensitivity. The analysis is very thorough and the conclusions, in particular the interpretation of results with biological considerations is excellent. I am not an expert on this subject area and although I am a statistician, GEE is not an approach that I am familiar with, so my comments are from the perspective of a technically-minded but non-expert reader. Therefore some of my comments may not be wholly relevant and are given more as possible considerations; or areas that may require clarification for other non-experts."*
* We thank the reviewer for these supportive comments and hope that our responses below adequately address the issues and suggestions raised.
* *"A key question for me is the choice of using the GEE approach, which is of course an established and appropriate method for many research questions, but I think could be better justified here. The main advantages I am aware of are: Richer correlation structures, easily handles missing data, robust standard errors. Here, the number of follow up visits is not large and the amount of missing data is relatively small. My first thought when approaching the analysis would therefore be to consider a standard linear regression analysis with robust/clustered standard errors to incorporate repeated measures, and potential multiple imputation of missing values and a comparison with complete case analysis. So it would be nice to briefly spell out the things that GEE offers here over other approaches to make this clear (in the ESM if need be so as not to add to the word count). Currently the main statement is the ease of handling missing data - which is handy but perhaps not essential - and some comments on the working correlation matrix, which are not terribly informative. So some statements about the improvement in estimates under a specific correlation structure that is not easily handled otherwise, for instance, might be helpful. Of note is that in the ESM it is stated that "other ... matrices had similar fit" - it may be worth noting that the model fit may be similar, but estimates may be more efficient under a particular correlation structure."*
* We agree entirely with the reviewer's suggestion to more fully describe the justification for using GEE over other methods. In the revised manuscript, we have expanded our rationale for using GEE in both the Methods and the ESM Methods sections. We chose the GEE technique for a variety of reasons. As the reviewer states, GEE is able to handle visit dates that differ between participants (e.g. some participants come exactly 3 years later for a follow-up while others come some months before or after 3 years) and for dealing with the inherent correlational structure underlying longitudinal data. Because of these features, among others, GEE can compute robust standard errors of the estimates that more accurately reflect the underlying data. The ability to specify a working correlation matrix makes this a powerful technique for longitudinal data compared to standard linear regression, as it removes the underlying assumption that observations are independent (which is not the case with longitudinal data). While choosing the specific correlation matrix is important, GEE still produces robust estimates and standard errors even with a misspecified correlation matrix. In the case of longitudinal data, an auto-regressive correlation matrix is the most appropriate, statistically and biologically, since measurements on an individual over time will change (e.g. people gain or lose weight over time) but still be highly correlated to their previous measurement (e.g. people don't gain or lose too much weight from their original weight).
* *"Another thing that might be commented on briefly is the different nature of estimates from GEE models, which are population averaged and have a slightly different interpretation. This difference might be seen as a disadvantage to GEE modelling - or do the authors feel this is more appropriate for this particular question? This is something I do not have a thorough understanding of, but I do note that the authors comment in the methods "...the interpretation of the GEE results is as an expected percent difference in the outcome variable for every standard deviation (SD) increase in the predictor variable given the covariates are held constant" does not appear to be consistent with what I have read in terms of keeping other factors constant, e.g.: "However, because the GEE approach does not contain explicit terms for the between-cluster variation, the resulting parameter estimates for the contrast of interest do not have the usual "keeping other factors constant" interpretation." (Hanley et al, Am J Epidemiol 2003;157:364–375)."*
* Thanks for this comment. We were only concerned with population level averages, and not individual-level estimates. GEE is a marginal model, which estimates the population average, and is one of the other reasons we chose GEE. However, even though GEE is a marginal model, for continuous outcomes (e.g. insulin sensitivity) the estimates derived from GEE's counter-part, namely mixed effects models, are more or less equivalent in both the value and in the interpretation. For binary outcomes (e.g. diabetes cases), the interpretation and estimates are substantially different from mixed effects models.
* Regarding the cited paper, the authors state the paper is an *"...orientation focused on correlated data arising from the relatedness of* ***several individuals*** *in the* ***same*** *cluster, rather than several "longitudinal" observations in the same individual"* (emphasis added). This is an important distinction and their statement about the "keeping other factors constant" and the lack of terms for the "between-cluster variation" is true in this case. However, this present analysis has the repeated measures at the individual level; the cluster is the individual and the within-cluster variation is the individual over time. In this context, interpreting the estimates derived from GEE is comparable to the classical regression equation:
* y = B0 + B1 X1 + ... + Bn Xn
* Such that the interpretation of e.g. total NEFA with ISSI-2 is that for every one SD increase in total NEFA there is an average 5% lower ISSI-2 for individuals at the same time point and of the same age, waist, ethnicity, etc.
* *"Another area that is unclear is the exact model specification. The unadjusted results I can assume are the outcome variables regressed via GEE against each NEFA measure. Then adjusted results include age, gender, WC and so on. Now, for the proportion of NEFA results, the approach seems to be that adjusted models were fitted with each NEFA as a predictor in turn while adjusting for the anthropomorphic and other covariates, as a PLS approach was used to examine all the NEFA components together as proportions. Was this approach the same for the analysis of NEFA as a concentration, or in this analysis were all NEFA components included in a single multivariable model? I think that this should be spelled out: if each NEFA component was considered in turn, and the variables adjusted for, for the concentration/proportion analysis. At present the assumption might be that multivariable models include all of the NEFA components in one model." ... "If so, I do not fully understand why there is not a model that includes the joint effect of all NEFA components together. Could this not be done for the concentration analysis? Also, it could perhaps be done for the proportion too, if one of the components is set as the baseline - similar to how categorical variables are handled in regression analyses, but allowing percentages for each indicator rather than the usual 0/1 dichotomy, vs. a "baseline" component. Such an analysis could also adjust for total NEFA, with the interpretation that proportion of each NEFA component is assessed while accounting for other NEFA components, and also the total NEFA. This is also something I struggled to understand a little in general - why should proportions of NEFA components be important? There may be a relatively high proportion of a particular NEFA but if the patient has a generally low overall NEFA, is there any reason to think a relatively high proportion of something will increase insulin sensitivity? These ideas may be wholly implausible of course and I may have missed the point - as I said my background knowledge of the biology is not ideal."*
* The reviewer brings up excellent points. The reason NEFA were included as a proportion in the PLS analysis rather than as a concentration was due to the inherent interdependence between NEFA values as a proportion. They all must equal to 100. It is this dependency structure in the data that allows PLS to more accurately identify clusters within the data. Concentration data do not have this inherent dependency (all NEFA values do *not* equal to 100). If one NEFA increases in concentration, other NEFA values do not necessarily decrease, as would be the case with the proportion. We have clarified this in the methods.
* The reason we did not include all NEFA variables (as proportion or as concentration) into a single GEE model is because GEE cannot handle this level of complexity particular due to the collinearity between NEFA leading to biased estimates and model instability. How the reviewer described setting the variables in the regression equation may be useful for some research questions and with fewer variables. However, we had 22 different fatty acids (44 if including mol% and nmol/ml) and 4 different outcomes. The high dimensionality of this analysis makes interpretation using standard linear regression approaches extremely challenging, not to mention introducing potential bias due to correlation among the NEFA (especially as a proportion). These reasons are why we chose to run multiple GEE models for each NEFA-outcome pairing in addition to running a PLS analysis with all NEFA in a single model (since PLS is designed for this particular analysis).
* The reviewer is correct that proportions *alone* are not enough to draw a meaningful conclusion, as per how the reviewer described it. This, we believe, is a major strength of our study since previous literature using other fatty acid fractions (e.g. phospholipid) generally only investigate proportions but do not report on or measure the concentration. Concentration data are more technically challenging to quantify in the lab and so it is understandable why large cohort studies do not measure these data. Using both concentration and proportion, we can identify what specific role the absolute size and specific proportion of individual NEFA have on the risk for diabetes.
* *"I do also wonder somewhat whether there may be over-adjustment with some of the covariates included, and agree with another reviewer who pointed out the importance of the unadjusted results. A potential issue is confounders that may be on the causal pathway, which is a judgement that must be based on expert knowledge. The authors note for instance that adjusting for WC improved the fit of the model, and was indicated for inclusion by the DAG. Neither in my mind is reason alone for its inclusion, and there is no guarantee on this basis that adjustment for this is necessarily appropriate - this must be based on subject-specific knowledge of whether adjustment is appropriate to determine the independent effect of the variable of interest."*
* Given that the underlying physiology of NEFA's role in diabetes pathogenesis has not been fully elucidated, it is possible that there may be some over-adjustment in our models. However, we have taken extensive steps to understand the biological and causal pathways underlying our associations and to empirically determine which covariates to include in the model.
* In the case of WC, we feel there is substantial evidence to support adjusting for this variable when studying NEFA's role in metabolic capacity: 1) biologically NEFA comes from adipose tissue; 2) intra-abdominal fat tissue is more metabolically active; 3) adipose tissue is an active endocrine organ that contributes to insulin sensitivity and metabolism in general; and 4) WC is a direct measure of abdominal (and indirectly intra-abdominal) fat mass. It is for these reasons that adjusting for WC is, we believe, strongly warranted.
* Having said that, because of comments from previous reviewers about this very issue, we included an analyses to address this concern. We ran models adjusted for all covariates *except for WC* and reported on the results in the ESM material. We had also included a discussion about this issue in the Conclusions of the original manuscript, which we hope will inform the interested reader about this particular issue and allow them to draw their own interpretations of the results.
* *"Which brings me to the DAGs. "Directed acyclic graphs" just means a collection of boxes representing different bits of information with arrows between them that do not form loops; DAGs are used in Bayesian analysis for instance to represent how data and parameters in a model are linked. A DAG therefore is a representation for how information in linked, not a process to in itself to determine appropriate model building or make a "recommendation". I think that if the approach employed is to be used as justification of the model building process, it needs to be better described: presumably there is some algorithm being used that determines interrelationships of variables in the data, of which I would like to see at least a brief outline – a link to a website is not really sufficient here. Further, the graphs provided give no indication of what the different arrows and colours are."*
* We agree with the reviewer that the wording around DAG use in model building was not entirely clear and we have worked to make it clearer in the ESM and methods of the revised manuscript. DAGs *are* in fact a process to identify potential covariates to include in a model when used algorithmically. The mathematical construct underlying DAGs and model building is the reason the DAGitty software was developed, so that this mathematical process could be automated. Much greater detail on how DAGs are used to build models can be seen in the two references listed below. The DAG figures in the revised manuscript will be described in more detail, including what each colour and arrow indicates.
  + Greenland S, Pearl J, Robins JM (1999) Causal diagrams for epidemiologic research. Epidemiology 10:37–48.
  + Textor J, Hardt J, Knüppel S (2011) DAGitty: A graphical tool for analyzing causal diagrams. Epidemiology 22:745.
* *"Last of all, the analysis excluded diabetes cases, which is justified. However, could incident cases be included with follow up time included to the last visit prior to developing diabetes? This may not be so straightforward in practice of course. It’s not really clear if any of any pre-incident follow up time is included at present, or if all data on incident cases is discarded "incident diabetes cases were excluded from further GEE analysis" – does the "further" bit mean post-incidence here?"*
* Yes, all follow-up data are included *until* the visit with the incident diabetes case (e.g. a participant is diagnosed with diabetes at the 6 year visit, but the baseline and 3rd year visit data are still included in the analysis). This reason, among others, is why GEE and similar techniques are so powerful. The wording has been fixed in the revised manuscript to make this point clearer.
* *"MINOR COMMENTS:"*
  + *"Abstract and elsewhere: OGTT not spelled out, even if this is obvious to most readers."*
  + This suggestion has been added to the abstract.
  + *"List of abbreviations: if this is to be an exhaustive list many are not given, OGTT, HOMA, NEFA"*
  + These have been added as suggested.
  + *"QIC - presumably this is quite like the AIC and other information criteria - maybe mention this, and how the calculation differs? Is the QIC specifically for GEE models? Also in the ESM: "This identifies a specific model from a number of models that is the 'best' fit and least complex of all the models" - it can’t be the best fit AND least complex, as mentioned in the previous sentence it’s a trade-off - so maybe omit this sentence."*
  + The reviewer is correct that QIC is similar to AIC. We have expanded on the explanation about QIC vs other information criterion. The reviewer is correct that QIC is specific to GEE-type models. Because GEE uses quasi-likelihood rather than maximum likelihood estimation (as other techniques use such as mixed effects models), information criteria such as AIC cannot be used since they rely on maximum likelihood. We agree that the wording surrounding 'fit' and 'complex' could be clearer. However, we disagree that it should be omitted since techniques such as QIC *do* select models that have both a good fit and are less complex. While techniques such as AIC are powerful model selection methods, they have an inherent limitation in that models with more parameters will have a better fit. So models with more parameters (more 'complex') will be more likely to be selected by default when using AIC. QIC and other similar information criteria techniques on the other hand try to balance this limitation by more strongly penalizing a more 'complex' model (more parameters) while still emphasizing a good fit. We have clarified this in the manuscript with the wording: "... that has both a good model fit while keeping model complexity low (less variables included)." See the reference below for more information on QIC:
  + Pan W. (2001) Akaike's Information Criterion in Generalized Estimating Equations. Biometrics 57: 120-125
  + *"ESM - collider variables. As mentioned in main comments this is all a bit vague - "arrows appearing to collide" and so on doesn’t really convey what is going on here. This is a technical description so please be a little more precise!"*
  + We have added this to the ESM Methods section.

## Referee: 2

* *"In this manuscript, the authors investigated the link between total and several fractions of serum NEFA and the clinical indexes of insulin secretion and sensitivity by using GEE model in a diabetes risk group over 6-years longitudinal follow-up."* ... *"Overall, this study is interesting and novel to show the long-term exposures to some specific NEFA fractions are negatively associated with mainly beta-cell function in human, which convincingly confirms and advances the previous basic and clinical studies in this field."*
* We thank the reviewer for these supportive comments and the suggested edits below. We have incorporated the vast majority of the suggestions and have provided responses that hopefully clarify these concerns.

1. *"Serum insulin and/or NEFA levels could be affected by several factors: I would suggest excluding the participants who have endocrine, hepatic or renal disorders. Also, I suggest mentioning the detail of medications, which could affect insulin/glucose/NEFA measurement; page 24, line 7, "Medication use did not influence NEFA composition, the outcomes, or the GEE models (data not shown)" is not clear enough."*

* At the baseline visit, individuals were excluded from participating in the study if they CKD, liver disease, or other chronic conditions. We have expanded on the details about medication use (i.e. general types taken).

1. *"Glucose clamp test is the standard to evaluate insulin sensitivity instead of the indexes used in this study. I know it is very hard to perform it in a large number of participants, but it may be one of the reasons why the effect of NEFAs on insulin sensitivity was limited. I would suggest discussing the limitation of these indexes."*

* We agree, this suggestion is pertinent and has been included in the limitations section of the conclusion in the revised manuscript.

1. *"Most of the indexes in this study are based on glucose and insulin levels. I think the reader would know the data of baseline and follow-up glucose and insulin levels as well as indexes (in Table 1)."*

* Both glucose and insulin have been added to Table 1.

1. *"I understand several NEFA fractions are statistically related to indexes of beta-cell function; however, does each individual show the same trend? For example, the graph which shows NEFA concentrations and the individual decline ratio of beta-cell function over 6-years (value of slope, if compatible) may convince the reader more, at least for the representative NEFA fractions."*

* This is an interesting question. GEE does not model on the individual level, only on the population as a whole. As such, we cannot answer research questions specific to individuals. There are statistical techniques that can extract individual level associations, however, our objective was (at least initially) to examine population level associations. We therefore chose statistical techniques suitable to the study design and our research questions, which was in this case GEE modelling.

1. *"Referring previous longitudinal studies showing the link between NEFA level and beta cell deterioration, even which were investigated in the diabetic patients, could confirm the results of this study."*

* We have added some relevant references and associated discussions to the conclusions.

## Referee: 3

* *"The authors addressed all major concerns related to the manuscript. I feel that the paper improved by the changes and became more focused by omitting the PLS-DA analysis. By adding a detailed explanation to the discussion on the potential importance of potential confounders and mediators, the interpretation of the findings became much clearer."*
* We thank the reviewer for these comments and have made the suggested changes as requested below.
* *"Some minor comments regarding the wording of the conclusions and missing information on the DAGs remain but I only have the following comments:"*
* *"Minor comments"*
  1. *"Introduction: "Therefore, our objective was to examine the association of serum NEFA composition on changes over time in insulin sensitivity and beta-cell function in a longitudinal cohort. We hypothesized that higher palmitic acid and lower PUFA such as eicosapentaenoic acid in the NEFA fraction would associate with declining insulin sensitivity and beta-cell function over 6 years." I would prefer to use the word differences instead of changes, as the main focus of the paper is not on the time interaction but on the main effects."*
  + We agree with this wording. It has been edited in the revised manuscript.
  1. *"Conclusions: "In a Canadian population of adults who were at-risk for diabetes, we found that higher total NEFA concentrations independently predicted lower beta-cell function after 6 years." I would prefer to use over 6 years instead of after 6 years, as the difference was present from the measurement point to the end of follow-up."*
  + We agree with this wording. It has been edited in the revised manuscript.
  1. *"Please add the number of observations for each clinic visit in Table 1."*
  + As requested, this has been added lto Table 1.
  1. *"Please add the number of incident and prevalent diabetes cases to the respective exclusions in ESM Figure 1. For example: n=736 had baseline data collected, n=XX diabetes cases were excluded; n=619 had 3-year data, n=XX incident diabetes cases were excluded..."*
  + The exact number of excluded cases of diabetes is dependent on this analysis and doesn't reflect the main PROMISE cohort. Since the CONSORT diagram represents the main cohort, we included the number of excluded diabetes cases in the text of the Results section since the text reflects the present analysis.
  1. *"Please add the colour codes for DAGs (i.e.: exposure, outcome, ancestor of exposure, ancestor of outcome, ancestor of both, causal path, biasing path)."*
  + This has been added to the DAG ESM Figures, as requested.

## Referee: 4

* *"This is an interesting analysis on NEFA ,insulin sensitivity and beta cell function."* ... *"My major concern is that with the current approach the complexity of metabolic regulation may be lost to some extent. This makes it hard to interpret these findings with respect to biological significance:"*
  + *"With respect to peripheral insulin sensitivity (referred to in the discussion) the effects of saturated fatty acids may strongly depend on metabolic phenotype, ie more pronounced in insulin resistant subjects, was initial metabolic phenotype taken into account in the present analysis?"*
  + The initial metabolic phenotype and subsequent influence of saturated fatty acids is taken into account in the time-by-fatty acid interaction. In the present analysis, there were no significant time-by-NEFA interactions.
  + *"it is a pity that the analysis of fatty acids composition in phospholipids and serum cholesteryl esters have been published in a separate paper, altogether this would add very much to the total interpretation."*
  + While we agree with the reviewer that inclusion of the phospholipid and cholesteryl ester (as well as the triacylglycerol) fractions would have added to the interpretation, the volume and complexity of the analyses would not have been publishable given current journal word and figure count limits. This was the reason we decided to publish the fractions in separate manuscripts, to maintain a clear and focused narrative within the constraints of the journals limits.
  + *"some discussion on what determines plasma fatty acid profile may be useful, beside spillover also (selective) release by endogenous lipolysis. To what extent does it reflect dietary composition, was habitual fat intake taken into account?"*
  + Fasting NEFA derives from adipose tissue, which contains fat stores from both diet and de novo lipogenesis. Besides potential dysfunction in the adipose tissue for the release of NEFA, other factors that could determine the NEFA profile include genetics and long term dietary fat intake (adipose tissue fat stores have a half-life of ~1.5 years). These other factors were not considered and discussed in the present paper due to the difficulty in measuring long term diet accurately enough on the specific type of fat consumed, the lack of genetic information contained in the cohort, and the fact that disentangling the specific contribution of de novo lipogenesis from dietary fat on adipose tissue stores is well outside the scope of this paper.
* *"Altogether, this paper would gain strength by addressing the limitations in interpretation in more detail."*
* Based on comments from multiple reviewers, the limitation section has been expanded.

## Associate Editor

* *"The reviewers provide numerous minor and major comments, for example regarding motivation of the statistical approach, which require revision or careful response."*
* We hope that the revisions made and the responses to the reviewers adequately address all of their concerns or comments.