**Author's response to reviews**

**Title:** Shape information from glucose curves: Functional data analysis compared with traditional summary measures.

**Authors:**

Kathrine F Frøslie (k.f.froslie@medisin.uio.no)
Jo Røislien (jo.roislien@medisin.uio.no)
Elisabeth Qvigstad (elisabeth.qvigstad@ous-hf.no)
Kristin Godang (kristin.godang@ous-hf.no)
Jens Bollerslev (jens.bollerslev@ous-hf.no)
Nanna Voldner (nanna.voldner@ous-hf.no)
Tore Henriksen (tore.henriksen@ous-hf.no)
Marit B Veierød (m.b.veierod@medisin.uio.no)

**Version:** 3  **Date:** 27 November 2012

**Author's response to reviews:** see over
Dear Editor,

Please find attached a revised version of our manuscript *Shape information from glucose curves: Functional data analysis compared with traditional summary measures*, manuscript number MS: 1063001331805422

We are grateful to the reviewers for their positive and thorough evaluations. The comments were to the point and very helpful.

We have used underlining (text added) and strikethrough (text removed) to indicate revisions in the manuscript.

All sentences and paragraphs that have been rewritten are described in the point-by-point response letter following on the next page. The new text is cited in quotation marks as underlined text.

As requested by the editor, we have specified the information about the ethics board who approved the study. The text “The study was approved by the Regional Ethics Committee …” is now replaced with “The study was approved by the Regional Committee for Medical Research Ethics, Southern Norway, Oslo, Norway (reference number S-01191),…”

We hope that you find our revisions satisfactory.

Yours sincerely,

Kathrine Frey Frøslie
Statistician, MSc
Point-by-point response to reviewer #1, Robert West

Unusually the FDA approach is taken despite the very limited (5) time measurements for each participant. It appears that the method still works well, and the authors have discussed taking measurements with a continuous monitor. There are other circumstances in which a limited number of measurements are available but the FDA approach might be of interest. Further comments would be useful for other researchers.

Reply: We agree with the reviewer. Our work shows the usefulness of FDA even when the temporal measurements are few, and this opens for application of FDA in other areas. We have added the following text to page 21, to emphasise the potential for FDA in other clinical projects:

“Our work shows that the FDA approach worked well, despite the very limited number of measurements for each participant. Dynamic, physiological processes will often be represented by scarcely sampled measurements, especially when repeated blood samples are required. In addition to glucose regulation, other examples where an FDA approach can be valuable include diurnal measurements of hormone regulation, metabolic changes during or after meals, or after physical exercise.”

Time warping might be of interest but this approach is essentially ruled out due to the very few time points with which to identify landmarks/registration points. Some comment might be useful.

Reply: We agree. In our project, time warping was not an issue, as the glucose intake was controlled and regulated. For analysis of data from continuous glucose monitoring (CGM) devices, in contrast, warping may be both important and non-trivial. People’s dietary habits will be subject to variation, regarding both timing and composition of meals and size of intake. Hence, to assess daily life glucose dynamics, warping will be essential. To forestall these challenges in the analysis of CGM, we have added the following text on page 20:

“Time warping can be a challenging task in FDA [17, 18]. In the present analyses, this was not an issue, as the glucose challenge test was controlled by health personnel and all measurements were done in the morning, immediately before a regulated glucose intake. For analysis of data from CGM devices, warping may be important.

OGTT data from a CGM device will not need time warping, but such adjustment may be essential in assessment of daily life glucose dynamics, as people’s dietary habits will vary largely regarding timing and composition of meals, and size of intake.”

The arguments are considerably undermined by the very few participants who are identified as having GDM - only 3 of the 974. A modification of the outcome might be able to overcome this objection.

Reply: We regret that we have not been clear enough about the GDM outcome. The first part of the analyses in the paper (including the curve smoothing and FPCA) involves only the glucose measurements at gestational weeks 14-16. At this point of time, only 3 women in the sample were diagnosed with GDM. However, in the regression analyses in the last part of the paper, the outcome is the 2-h value at gestational weeks 30-32, where 51 women were diagnosed with GDM. To clarify this, we have added information about this in Table 1 and in
the first paragraph in Results: “The number of women with a GDM diagnosis increased from 3 (0.3%) at gestational weeks 14-16 to 51 (5.5%) at gestational weeks 30-32 (Table 1).”

We have also added “at gestational weeks 14-16” in the second paragraph in Results and in the figure legend for Figure 1, to clarify further. See also the text added on page 20, as a response to reviewer 2, comment 2.

Page 6 known known - word repeated

Reply: Thanks. The redundant word has been deleted.
**Point-by-point response to reviewer #2, Helle Sørensen**

I find the functional approach appropriate for the data, but also want to stress there could be alternative approaches (see comments 1 and 2 below). I certainly don’t think that all these different approaches should be incorporated, but I believe that it would be fair to mention that there are alternatives to the proposed analyses.

1. Broadly speaking, the paper studies two kind of analyses: functional data analysis and analyses based on summary measures. There are of course alternatives to these analyses, for example:

   − The relation between BMI and glucose values could be examined with a classical longitudinal data analysis (five measurements per woman), with random effect of woman and careful modeling of the covariance structure. This is common in studies on nutrition where similar glucose tests are used.
   − As input to the regression analysis of GDM, you could use the scores from an ordinary PCA on the five measurements (rather than the scores from the functional PCA).

Both analyses should be pretty straightforward because there are only five measurements per curve, and because the measurements are taken at the same time points for each woman. Someone might even be hard and argue that FDA is not really necessary for these data, because you could probably extract similar amount of information from the data with various kinds of “ordinary” multivariate methods.

Let me stress that I actually like the functional approach, and my point is not that you should carry out and report all these analyses in the paper. However, I think it would be fair to mention that there are other “standard” options than the analyses based on summary measures. At the same time, you could explain the extra benefits from the functional approach: It would not be a problem if the curves were samples of different time points, and missing values are not really problematic either (at least from a theoretical point of view).

**Reply:** Yes, we agree. The FDA is not the only approach to these data, and it is even less obvious due to the scarcely sampled glucose measurements. As pointed out by the reviewer, other multivariate methods would be adequate, and would by many be considered more straightforward than the FDA, and give similar results as the FDA. But as the reviewer also pointed out, FDA has advantages. We have added the following text on page 19 to elaborate on these issues:

“Although FDA or parametric modelling are the most natural approaches to glucose data for the study of glucose curves as single entities, there are alternatives to these analyses for the data presented in this article. For instance, the relation between BMI and glucose values could have been examined with a classical longitudinal data analysis with five repeated measurements per woman, with random effect of woman and modelling of the covariance structure. Also, instead of scores from FPCA, ordinary PCA scores based on the five glucose variables could be used as input to the regression analysis of glucose tolerance later in pregnancy. With only five measurements per curve, and measurements taken at the same time points for each woman, such traditional multivariate methods would be expected to extract similar information as the FDA. However, FDA is easier to apply in situations with more frequent sampling, sampling at unequal time points and missing data. In addition, FDA
emphasizes the basic assumption about continuity of the underlying process and its
derivatives, and opens for analysis of the derivatives of the curves.”

2. There are two glucose curves for each woman, one taken in early pregnancy and one in
late pregnancy. Only the first curve is considered as functional, whereas the second curve
enters into the analyses only via the final value. This may very well be appropriate for the
GDM analysis (see page 9), but it could also be interesting to see the relation between curves
from early pregnancy and curves from late pregnancy. At least, you should emphasize that
only one of the glucose curves is used as functional.

Reply: Yes, we do have two OGTT curves in the cohort, and in this manuscript only one is
considered functional. We are currently working on a project where we study the two glucose
curves against birth outcomes. However, in the present manuscript, we focused on a maternal
outcome within the pregnancy, and we wanted a clinically relevant outcome to illustrate the
usefulness of information from FDA to clinicians. Therefore, we chose to focus on the 2-h
value in third trimester (which has a direct clinical application) instead of the entire curve in
third trimester. We do agree with the reviewer that the relation between the early and late
pregnancy glucose curves is very interesting, but defer to our next project to exploit this in
detail. To clarify our choices, we have added the following text on page 20:

“The women in the cohort underwent two OGTTs, but only one was considered functional in
the present work. We chose the 2-h value in third trimester as the main outcome instead of
the entire curve in third trimester, due to the clinical relevance of this value in pregnancy
care. As glucose curves are not commonly used, inference about the 2-h value would better
illustrate the usefulness of information from FDA for a maternal pregnancy outcome in
clinical practice.”

Moreover, you choose to discretize BMI and the 2h value from the late glucose test. This is of
course not wrong, but I believe that using the variables as continuous would be more
appropriate (see comments 3 and 4 below).

3. Comments on the regression analyses:
   A. You carry out a multinomial regression with seven groups constructed from a
discretization of the 2h glucose values in late pregnancy (page 9, 13). One of the
limits corresponds to the WHO definition of GDM, but the other limits are somewhat
ad hoc (based on quantiles). Why did you choose to make seven groups in the first
place? And why did you not use the 2h glucose level as a continuous variable,
perhaps as a supplement to a logistic regression with limit equal to the GDM?
Perhaps the analysis with continous response could be mentioned as a supplement to
the multinomial analysis.

Reply: See joint comment following 4B.

   B. You carry out the regression analysis with the principal component scores as
explanatory variables. Another type of regression model assumes that a linear
predictor takes the form … see for example Goldsmith et al. (2011) and Muller and
Stadtmuller (2005). Did you try to carry out such analyses? This is doable for
continuous and binary response (use for example the gam function from the R-
package mgcv, but is far from standard for a multinomial response.)
Reply: We thank the reviewer for bringing this to our attention. No, we did not try this, and were also not aware of this option, but we will keep this in mind in future FDA projects.

4. Comments to the FANOVA and ANOVA analyses:

A. I guess that estimates $\beta^\text{ref}_*(t)$ and $\beta^g_*(t)$, as well as the p-values $p(t)$ are computed $t$ by $t$ (pointwise)? Is this true, or has the estimates and p-values been smoothed? This is not clear from the paper (neither from Appendix C). Moreover, the permutation test giving the overall p-value could be slightly better explained than in Appendix C.

Reply: Yes, this is correct. We agree with the reviewer that this was too briefly summarised in Appendix C. The permutation test implemented in the fda package only allows for comparison of two groups. The pairwise implementation makes the notation of the F-statistic in Appendix C complicated, and we have therefore added a general explanation of the test rather than a strict mathematical notation, and refer to the publications for computational details. The following text has been added to the last paragraph of Appendix C (we have not used underlining for this text in the manuscript, as it made the mathematical notation less readable):

“The presented permutation tests are based on 1000 permutations of the fitted curves in two different categories. The CIs and $p(t)$ curves are calculated point-wise over the $t$ range, using the estimated F-ratio $FR(t) = \frac{MRS(t)}{MSE(t)}$, calculated as the ratio of residual variance, $MRS(t)$, to predicted variance, $MSE(t)$. The permutation distribution is found for the point-wise F-statistic, giving CIs and $p(t)$ curves over the $t$ range, and for the maximal value of the point-wise F-statistic, giving an overall $p$ value.”

Remark: Due to the implementation in the fda package, we have done four pairwise comparisons. The p values were either non-significant or highly significant. Hence, (Bonferroni) correction would not have changed the conclusions, and we have therefore not corrected for multiple testing.

B. Wouldn’t it be more obvious to use the BMI variable as a continuous rather than a categorical one? Then the analysis would not be an ANOVA type (but rather a regression with functional response), but the analysis could still illustrate the difference between ordinary and functional models and analyses, p-value curves, overall p-values, etc.

Reply: We do understand the referee’s concerns (3A and 4B) and these issues have also been subject to discussion prior to submission. Contrary to general statistical advice, we have categorised two continuous variables in the analyses. We considered it important to introduce the FDA techniques to a clinical audience, and we believed that a categorisation of BMI and the 2-h glucose at gestational weeks 30-32, based on the use of these variables in clinical practice, would ease the presentation of FDA. Different BMI categories are assumed to represent different risk groups [27], and BMI categories are frequently reported in clinical literature. The analyses were consequently done with the categorised BMI variable, although functional regression with BMI as a continuous variable would be preferable from a
statistical point of view [33], especially when there is no obvious signs of nonlinearity (Figure 4a).

Further, to show clearly how the early pregnancy glucose curve differed for women with low or high 2-h values in late pregnancy, we chose a large reference group in the middle range, and groups with lower and higher values. As the clinical focus is on the 7.8 mmol/l value, but not on other values, we chose percentiles as our criterion for categorisation. Nevertheless, although the 7.8 mmol/l limit is used to classify GDM, there is no evidence for an abrupt change in risk for adverse health outcomes for values just below or just above 7.8 mmol/l, and there is an on-going discussion about such limits. We therefore chose to split the GDM cases into two categories, and to specify two categories between the reference category and the GDM cases, to emphasize the differences around this value. We could of course have chosen other percentiles. But the ones we chose worked well to illustrate our point visually, and gave an acceptable number of women in each category. If we had chosen to use the 2-h glucose level as a continuous variable, we would have had to include nonlinear terms in a regression model, and this would be difficult to communicate. Binary logistic regression with the women split into GDM and non-GDM cases could also have been done, but as we mentioned above, there is no evidence for an abrupt change in risk for adverse health outcomes for values just below or just above 7.8 mmol/l, and our results show the power of FDA to discriminate also between sub-groups with or without a GDM diagnosis.

We have added the following text on pages 19-20 to elaborate on these issues:

“Contrary to general statistical advice [33], we have categorised two continuous variables in the analyses. An important aim of the present work was to introduce FDA and its benefits to a clinical audience. To ease the presentation of FDA, we chose to categorise BMI and the 2-h glucose at gestational weeks 30-32, based on the use of these variables in clinical practice. Different BMI categories are assumed to represent different risk groups [27], and BMI categories are frequently reported in clinical literature. The categorised BMI variable was therefore used in the analyses, although functional regression with BMI as a continuous variable would be preferable from a statistical point of view [33], especially as there were no obvious signs of nonlinearity (Figure 4a). The categorisation of the 2-h glucose value at gestational weeks 30-32, in contrast, revealed important non-linear relations (Figure 6). Hence, as an alternative to the multinomial logistic regression model, a regression model with the 2-h value as a continuous response variable could have been used.”

5. I’m not always a great fan of functional PCA since the results can be difficult to interpret. At least, that is my experience. However, in your case the first two components have very nice interpretations. The interpretation of the third one is not quite as clear (in my opinion you are on the edge of overinterpreting it), and you could just have stopped after two components.

Reply: We do agree with the reviewer that the proportion of the variance explained by FPC3 is so small that we could have stopped after two components. But the shape indexes that we have referred were constructed to differentiate between curves with one or two/three peaks, assuming a priori that this was essential information, and important to extract. Therefore, we needed the third component to show that the “one-versus-two-peaks” feature was not essential shape information in the curves.
We also understand the reviewer’s concerns about potential over-interpretation of FPC3. However, the other publications, in particular the individual curves displayed in Freckmann, 2007 (reference 14) and the physiological theories about feedback mechanisms in the glucose regulation, for instance described in Trujillo-Arriaga, 2008 (reference 35) support our term “oscillating”. It is also possible that the third component would explain a larger part of the total variation if the sampling was more frequent and over a longer time period.

We have added the following text to page 17 to relax our interpretation of FPC3:

“This proportion was so small that FPC3 could have been left out of the analyses. We chose to include FPC3 for the comparison of FDA with the shape index.

(…)
The shape feature of FPC3 was however less clear than for the first two components, and although it is possible that the third component might explain a larger part of the total variation if the sampling was more frequent and over a longer time period, this component should be used and interpreted with caution.”

6. I find the graphs very illustrative, except perhaps Figure 4b which is a bit difficult to access with overlapping confidence intervals of the same colour.

Reply: Ok. Figure 4b is now edited so that the confidence intervals have different colours (the same colours as the corresponding means), and the stripes/hatches are not at the exact same position, which makes it easier to tell them apart.

7. And then some minor details:
   - Page 6, line 5: The word “known” is repeated.
     Thanks. The redundant word has been deleted.
   - Page 8, line 1: I would write “FPC scores for each curve” rather than “FPC scores for each women” since there are actually two curves per woman.
     We agree. This is now changed.
   - Page 9, lines 5–6: Define the intervals as (18.5–25) and (25–30) rather than as (18.5–24.99) and (25–29.99).
     We agree. This is now changed.
   - I think Table 3 would be slightly easier to read if you put parentheses around the CI’s for the OR’s, such that, for example, the top left “OR 95% CI” part would read 1.08 (1.04,1.13)
     We agree. This is now changed (with no underlining of changes in the table in manuscript).
   - The layout of Figure 5 is different from the other figures where each subgraph is quadratic. This would be better for this figure, too.
     We agree. Figure 5 has been edited and each subgraph is now quadratic.

Other minor changes

We have introduced the abbreviation for continuous glucose monitoring (CGM) on page 4. The abbreviation is also added in the abbreviation list on page 22.
We have made minor adjustments to the \( n \) column in Table 3, due to some missing values for BMI in these analyses (no underlining is done for these changes in the table).

The words \textit{group mean} have been inserted at page 25.