Author's response to reviews

Title: Analyzing repeated data collected by cell phones and frequent text messages. An example of Low Back Pain measured weekly for 18 weeks.

Authors:

Iben Axén (iben.axen@ki.se)
Lennart Bodin (lennart.bodin@ki.se)
Alice Kongsted (a.kongsted@nikkb.dk)
Niels Wedderkopp (nweedderkopp@health.sdu.dk)
Irene Jensen (irene.jensen@ki.se)
Gunnar Bergström (gunnar.bergstrom@ki.se)

Version: 3 Date: 2 May 2012

Author's response to reviews: see over
Response to Reviewer’s report

Abstract
I would add “and the between measurement occasion correlation” to the list of vital issues to consider. Interpreting complex outcome measures in a clinically meaningful way is, of course, most important, but has nothing in particular to do with repeated measures.

We have added the suggested sentence and moved the sentence about clinical importance to the aim. We have also added a sentence about the overarching idea for the article.

Is Hazard regression the same thing as proportional hazards regression (Cox 1972) – if so please use the familiar term.
Thank you, we have changed the wording as suggested.

I think you will find that GEE usually stands for ‘Generalised Estimating Equations’ not ‘Generalised Estimation Equations’. I know the gee software uses the second version, but it is very much in the minority.
We agree. This is now changed throughout the text.

What are ‘Frequent measures’?
This refers to the fact that the data used in the example are not only measured repeatedly, but also frequently (which is not always the same thing). For clarity, we have now used “repeated” throughout and added “frequent” where appropriate.

Background
What has EMA got to do with seasonal variation – I don’t see any obvious link. The possibility of collecting data in the natural environment makes it possible to capture the fluctuations of a condition in detail. Some conditions vary with time of day and, indeed, season. The second point we were making, is that people bring the mobile phone with them on holiday, thus not disrupting data collection. This is now clarified in the text.

‘The data are accessible by the researcher’ not ‘to the researcher’
The sentence refers to the fact that the responsible researcher has access to the data online. It is now re-worded.

You mention ‘previous studies have shown compliance to be high’ and then cite only one – ref 4. Are there more? We have corrected this sentence and added references.

‘The method has shown good user friendliness’ is awkward ‘The method is very acceptable to users’ might be better. We have re-written the wording, but the term “user-friendly” is used in the paper referred to, hence the use herein.

Which two of the 'referenced studies' produced the data – I suggest a reference!
We apologize for the uncertainty. Data from two studies were merged and used as a model data set. The references are now added.
Methods
It would be useful to say how big each study was. We did not add this information as the model data set was constructed with subsets of data from each study. However, this information may be found in the two original papers, and we have added information concerning the formation of the dataset.

'supervised' and 'unsupervised' – I've never come across this use of the two words before. There is a quite different, and common use in the machine learning and classification literature. In any event I still don't understand it after reading your definitions. Please explain a lot more fully. Supervised classification has been used in connection with discriminant analysis (groups known), unsupervised in connection with cluster analysis (groups unknown). In the present context with variabel versus person-oriented analysis these concepts don't fit so well, we agree. These concepts have now been deleted.

The phrasing of 'inherent covariance' is wrong – you mean either or both of within-subject or between-occasion covariance, which are the two sources of non-independence in this type of data. Any data set with more than one variable has a between-variable covariance. I would talk about lack of independence instead, as it is a clearer concept. We have made the corrections as suggested.

The location of a confidence interval is an unfamiliar idea to me – what do you mean? If a different estimation method is used, going from GEE to multilevel modeling for instance, the estimate of the parameter may change, and the width of the confidence interval may also change. Then the location of the confidence interval shifts as it is centered on the estimate as the midpoint. Examples of this kind are found in for instance our reference (20). We have clarified this in the text.

Research questions
1 How many days with pain do patients experience?
This is awfully elementary. Yes. However, this new method of gathering data presents a unique opportunity to accurately answer the question. We have condensed this section of the commentary somewhat.

2: What is the proportion of participants “recovered” at different time points?
‘To describe the variance of the population’ - do you mean the variability within person, or between persons? The variance is not either of these. We agree, and have clarified this as suggested.

Given your comment about multiple testing, you do realise that testing for a defined outcome, after graphing the data, poses exactly the same problem? You must fully pre-define your outcome of interest We do not at all understand the basis for the reviewer’s comment as it is clearly stated in our text, section 2A, that testing at different time-points should only be done at a prespecified time-point, and no testing of the example data set is done in section 2A. In the next
section is says in the title 2B ‘...at a prespecified time-point’, and then a numerical example is given when the time-point of interest is given by previous experience concerning the fourth or fifth week. We have added a statement to the text to further illuminate this point.

B: Incidence of recovery at a prespecified time point, Table 1. Hazard rates are the ratio of instantaneous failure rates in survival analysis, and are conceptually distinct from both odds ratios and relative risks. Please say how you calculated the relative risk from the odds ratios. Using GLM and related methods you may choose different link functions when the outcome is a binomially distributed variable. With the logit link the outcome parameter is the odds ratio (OR), with the logarithmic link the outcome parameter is the relative risk (RR). Conversion formulas between OR and RR exist but there is no use for them here, a direct estimation which also gives CI is preferred.

3: What is the time to recovery? “to study incidence of an event over time” ought to be “to study the incidence of an event over time” Thank you!

For the spline regression – what is the evidence that there is a statistically significant difference between the locations of the two knots in the two groups? The knots are reported in the module for nonlinear functions in SPSS with standard errors (SE). From this information it is easy to perform a test of the equivalence of the knots (null hypothesis) versus a difference (alternative hypothesis).

There is a large and well developed literature on using survival analysis with fluctuating conditions, where the event of interest can occur more than once. As stated in the Background section, our data were not suitable for recurring events, as the subjects were expected to get better as time progressed.

“as included poor compliers only marginally “ ought to read “as including poor compliers only marginally” Thank you

4: How is the repeatedly measured data associated with baseline (predictor) variables? Please explain, much more fully, ‘fixed effect’ this is meant to be an elementary introduction. We have added text that is it beyond the scope of this article to explain the specifics of the more advanced methods. Rather, we encourage readers to use the references that we suggest. Should we begin to try to explain features in general linear models including multilevel models we will expand outside the limits set for an article such as this.

The explanation here is very unsatisfactory – I'm quite familiar with these methods, and I don't understand what you are trying to say. Going from a long description of a t-test to this highly condensed description of very tough methods is not a runner. GEEs estimate marginal models – again you need to explain what these are. You don't discuss random effects at all, that I can see. Please see the comment above. We feel it is better to say directly that features of the more advanced models are best explained in regular textbooks in statistics or
quantitative methodology. Our intention is to show examples of where advanced models can be found and be applied and what the results may be. What does IRR stand for? We apologize that the explanation of this term was placed in the second paragraph of which it appeared. This is now corrected.

How, exactly, do you model a covariance structure? In the different softwares used for repeated measurements (longitudinal data) there are options for defining the correlation structures of the residuals. Some of these options were tested. Best fit of models was evaluated by considering Akaike’s index, as well as similar indices. This is now explained in the text.

You speak several times about ‘standard statistical software’ What do you mean? Both R and Stata (I think) implement all of these methods. You are probably right. However, this commentary is meant as a guide to clinical researchers, not to statisticians, and the “standard” software referred to are the software most used by clinical researchers, in our experience. We have clarified this further.

5: Are there subgroups of patients with similar courses of pain within the studied population?
In my experience all of this makes more sense if you show the cluster graphs, dendrograms etc...We agree. We have added a dendrogram for clarity.

Please use one term consistently – I suggest ‘repeated measures data’ as it is the common jargon. Thank you. We have changed this throughout.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Needs some language corrections before being published
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:
I am the lead author on a manuscript which seeks to introduce people working in the field of biomechanics to multi-level modelling. This has not yet been submitted for publication, but it does cover some of the same ground as this paper.
Response to Reviewer II

I understand the concept of using an example to go over the different types of analysis; however, I think that in this case it got a bit confusing. Although “the overall objective of the study was to explore and describe different methods of analyzing repeated data”, I think the focus was lost throughout the study and shifted to back pain and SMS measures. For instance, the results of your abstract do not match the purpose of exploring different methods because it primarily reports the results found for number of days with pain. In my opinion the comments part of the study were very interesting and probably most relevant in many cases. Perhaps the authors should consider focusing the study on the analysis and statistics; guiding the readers why they would choose one stats over the other and use number of days with pain mainly as an example.

We understand your concern. One could speak in general terms and perhaps use different data set (studying different conditions) to illustrate the examples. However, there is also a point in using the same data throughout to illustrate the differences and similarities between the methods, as we have done. Our idea was to illustrate the possibilities inherent in repeated data, and as such, we think it is simpler to use the example of low back pain throughout.

Following the same argument, I think that the introduction of the study is primarily focusing on the appropriateness of using SMS for data collection and how diaries may not be as good. However, the methods presented on the study, can be used to analyze data collected using both methods. Therefore, because the focus of the study is not to discuss SMS, the authors should consider focusing the introduction on statistical methods. We agree, we do emphasize the SMS method in the introduction, and repeated data collected with other tools may use this commentary as a reference as well. We have changed the wording somewhat in the Background section and throughout the manuscript to stress this point.

Minor Essential Revisions:

1- “Throughout this commentary, we will use “previous duration” as an example baseline characteristic. I suggest changing this to: Throughout this commentary, we will use “previous duration” as the baseline characteristic of interest. This has been changed as suggested.

2- As part of your methods, I was curious about: What were the instructions given to the patients at the beginning of the study regarding text message? What did the message say, word by word? If measures were collected once a week, how did patients record days of pain? Did that include recall as well? The answers to these questions are found in the original studies, but a short summary is now added to the text.

3- Which statistical software(s) was/were used in this study? Different softwares were used. We have added a list of examples of what we call “standard statistical software”, and we have included references to all of these. These are purposefully used as we believe that these are readily available to most clinical researchers.

4- You have mentioned on your methods that log and square root transformations were used. However, there were not descriptions of this use
during the study. When did you use this? **This is a general point to be aware of. In our dataset, the variables did not need transforming, and this is now removed in the methods section.**

5- Throughout the study the authors mention high compliers. However, there is no description of who is considered high compliers on the methods section. Despite of that, perhaps the authors should consider removing all of the analysis of compliers and non-compliers from the study as in my opinion this classification does not add to the main purpose of the study to look at different analyzes. **Thank you for noticing. This was stated under “Missing answers”, which was really far too late in the text, as the term is used consistently in the examples. We have now moved this section so that the definition (> 80 % response rate) and purpose (to illustrate the effect of missing answers) becomes clear. We think it may serve as a nice illustration of excluding subject with missing data, which is definitely an important issue in dealing with repeated measures.**

6- B: What is the average number of pain days? - I think that the “per week” should be added to the question to clarify the time frame evaluated. Also, the results (numbers) are better read if per week is added after the numbers. **We agree, this has been changed as suggested.**

8- It is not clear to me why you choose to “control” or separate groups based on chronicity on some analysis but not others. For example you use it for questions 1.b but not for 1.a. **You are correct, this is now corrected.**

9- Questions 2.b: The proportions of patients that recovered were compared to the proportion of those that did not recover. How did you calculate OR and CI? I think the readers would like to know. The same for risk ratio. Also, you mention NNT but don’t explain or show the results. **These are logistic regressions, and the result from comparing two groups is the OR with a corresponding CI. This has now been clarified in the text.**

10- I think that the authors should consider increasing the discussion on why someone would choose one methods over another and get more in depth about specific methods. For instance I think that the explanation of Survival analysis and cox regression were very superficial. **We have aimed to keep this commentary very broad and therefore not to explain the different methods in detail. We now state clearer in the text that we encourage readers to use the references provided when needed.**

11- “Throughout, regardless assumptions of hazards, data (continuous) count and level of compliance, the group with short previous duration of LBP had significantly higher “risks” (chance) of recovery. This suggests that for our model data set, the methods are robust.” I did not understand how the authors came to the conclusion stated on the last phrase. **We have added some text to further illustrate this point.**

12- I think that a lot of times throughout the manuscript the author assumed knowledge of the reader. For example, dendogram of cluster analysis. I think this
should be explained clearer so readers can understand the analysis. Thank you, we have added an example dendrogram and a brief explanation. However, you are right: this text does assume some knowledge on behalf of the reader. We have included relevant references for those interested in deeper knowledge.

13- Methods: Please add standard deviation to mean age. The change has been made as suggested.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:
I declare that I have no competing interests