Reviewer's report

Title: Modeling early recovery of physical function following hip and knee arthroplasty

Version: 1 Date: 7 July 2006

Reviewer: Jos W Twisk

Reviewer's report:

General
This is a nice paper describing the modelling of early recovery of physical function following hip and knee arthroplasty. However, there are a few point the authors need to consider.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

After creating a model for development over time, the authors address the influence of several predictors on the development over time. I think the authors should distinguish between the effects of the predictors on the outcome (which they probably did and which is not the influence on the rate of change; it is only the influence on the average value over time) and the effects of the predictors on the rate of change. For the latter the authors should look at the changes in either the regression coefficients for the linear or quadratic component of time, or at the changes in random coefficients, which indicates that the individual development over time is influenced by the several predictors. I don’t think the authors should change their analysis, but they have to change the interpretation of their results. Furthermore, table 3, in which the results of these analyses are summarised has a terrible lay-out. Please provide regression coefficients, 95% confidence intervals (and perhaps also p-values). Besides this, I suggest the authors to do three steps in the analyses. First, the development over time. Second, the influence of additional predictor variables and third potential interaction terms. The analysis with interaction terms should be reported separately from the other analyses, because otherwise they can’t be interpreted in a straight way anymore. By the way, the interactions with time indicate that the development over time is different for different subgroups.

A major concern is the fact that more than 50% of the subjects only have two measurements. If one models a quadratic function over time, it means that for more than half of the subjects in the analyses this can not be done. Besides this, the number of subjects is rather low and when also for more than half of the subjects the development can not be modelled, this will give definitely problems. At least the authors should say something about this in their discussion.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

The way the authors mention the linear and quadratic component of time is not very clear. Why not stick to the linear and quadratic component of time instead of the rate of change at 1 week and the change in change rate.

In the introduction the authors explain HLM. This part should be taken to the methods section.

The way the authors describe the HLM is not very clear. I think they can do better.

Line 5 of the results: p<0,60?????

In the results section, the authors state that a significant interaction means that the development lines cross. This is not really true, because the lines can also go further apart. It depends on the sign of the regression coefficient of the interaction term. In the results section, the authors should provide this additional information of the observed interactions.

At the end of the results section the authors state that the second degree polynomial provided a reasonable fit for the data. I wonder how the authors investigated this reasonable fit and what they mean by this statement.
Table 1: There is a mistake in the N of the LEFS score. Furthermore, in the last column the T-value and the degrees of freedom are not informative; they are only used to obtain the p-value. I suggest to leave that information out. The same holds (of course) for table 2.

When I look at the figures, I have doubts about the normality of the TUG scores. In the methods section the authors state that they have investigated the normality of the residuals. However, first of all, it is further not mentioned in the paper what they have found, and secondly, again looking at the predicted TUG scores, I hardly believe the residuals are normally distributed. Also the Stair test looks suspicious.

As mentioned before, more than 50% of the patients have only two measurements. I wonder whether the patients with only two measurements differ from the patients with more than two measurements. I can imagine that the subjects with two measurements recover faster, so they are the â€œbetterâ€™ patients. When that is true, the development at the third and fourth measurement is based on the patients who did not recover fast, so the development over time for the whole population is highly biased.

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

I declare that I have no competing interests