Reviewer's report

Title: Subgrouping patients on the basis of their individual course of low back pain over a six month period

Version: 3 Date: 19 April 2011

Reviewer: Ivan Steenstra

Reviewer's report:

First, I have to apologize for mixing up two Danish sub grouping studies sent in to Biomed Central that I was asked to review. The other was on data from a RCT.

The authors didn't make it easy on the reviewers however, by sending in a manuscript (s) that still needed considerable work.

The description of the process are still a bit fuzzy in my opinion. You refer to other publications, but a paper should be interpretable without them.

About the data collected: You refer to a submitted paper. So we need the info in this paper as well. For instance, did you use the questions from the EQ-5D and what is the possible range in scoring again (0-5?) or the health thermometer only. which valuation did you used. The paper you refer to describes several.

Here's the reference you couldn't find and I think it is relevant.


The recovery patterns of back pain among workers with compensated occupational back injuries.

Chen C, Hogg-Johnson S, Smith P.
Institute for Work & Health, Toronto, Ontario, Canada. cchen@iwh.on.ca

OBJECTIVES: To investigate the longitudinal patterns of recovery among workers with compensated occupational back injuries.

METHODS: A longitudinal cohort study, with one-year follow-up via structured telephone interviews, among respondents off work because of "new" back injuries.

Self-reported pain intensity was recorded at baseline and at four follow-up time points over the course of one year. Workers who answered the questionnaire on at least three occasions (n = 678) were classified into clusters according to their
patterns of pain intensity over time using a two-step cluster analysis.

RESULTS: Four pain recovery patterns were identified: workers with high levels of
pain intensity showing no improvement over time (43%); those experiencing
recovery in the first four months with no further improvement or possibly even
some deterioration, in the second half year (33%); those experiencing a slow
consistent recovery but still with considerable back pain at the end of the
follow-up (12%); and those quickly progressing to low level of pain or resolution
(12%). Trajectories of average Roland-Morris Disability scores and SF-36 Role of
Physical scores for above clusters mapped consistently with the corresponding
patterns in pain. However, individuals with fluctuating, recurrent pain patterns
showed the shortest cumulative duration on 100% benefit and the earliest
return-to-work among other clusters.

CONCLUSIONS: Four clinically sensible patterns were identified in this cohort of
injured workers, suggesting inter-individual differences in back pain recovery.
The results confirm that recurrent or chronic back pain is a typical condition in
respondents with new back injuries. Pain intensity and disability scores are good
measures of recovery of back pain at the individual level. After initial
return-to-work, or cessation of benefits, administrative measures of percentage
of respondents back at work, or no longer on benefits, may not accurately reflect
an individual's condition of back pain.

PMCID: PMC2078491
PMID: 17387134 [PubMed - indexed for MEDLINE]

I am still not convinced however of the relevance of the study. maybe I have to
state my bias: I don't like data mining approaches. And since you have an
abundance of clinicians in your group of authors it only makes sense to think of
some hypotheses first.

Methods

I have trouble reading the methods section. A lot of assumptions are made. First
the practical impossible to use 26 time points in a cluster analysis, is not what I
have experienced. Is this a limitation of Sleipner? Could you mention which
software package you use for each analysis, a reference number will do.

How do you decide where the turning point in recovery is and why did you decide
on two slopes and not more?

Furthermore, you mention standardizing the parameters. It seem to me that
which each step you take, you make decisions and lose information in each
step. I showed this paragraph to some statisticians and asked them if they would be able to replicate your analysis if we would collect the data. The answer was: "NO" and "Yes, but I would have to make a lot of assumptions (and would prefer to use another method like latent trajectory / class analysis)". The possibility to replicate a study is essential. Therefore, I think this part needs a rewrite, although I am not sure it would convince me in the end.

"Visual inspection of the individual curves and of the aggregated curve for all respondents was the starting point of the exploratory cluster analysis." So based on the inspection you made some decisions I guess. Can you elaborate on that process?

What does: "We hypothesized that the pain course over time would be similar in groups of individuals, and different from the course of other groups." mean?

And how does this relate to this in the discussion: "This study is explorative, i.e. we had no hypothesis regarding the number or size of clusters in our population."

From what I know is that clusters can never be based on statistical information alone. So their must have been some kind of process why you decided to go for this solution.

Findings:
"The use of text messages to collect data opens up possibilities for accurate descriptions of fluctuating conditions since it is possible to obtain detailed information on changes over time." but you loose the information by summarizing the data into the 4 parameters you use.

Table 1: Pain score of the excluded patients is 4.4.

Table 2 reporting on the slopes and intercepts. Are those the standardized values? Standardized values loose their meaning: number of days bothersomeness. The slope of regression line 2 is rather small, so the parameters used in the analysis (slope 1 & difference slope 1& 2) seem of the same magnitude. Doesn't this provide you with a problem? I think it would be helpful to give information on the variance in the parameters for each cluster.

How can group 3 be both fluctuating and moderate persistent? I think because for the methods and reductions in data you have applied you are not able to detect fluctuations.

At first glance (and second) Cluster 2 & 3 seem quite similar. Why did you decide to keep the two separate?

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published
**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

I declare that I have no competing interests