

**Biostatistics Tore Wentzel-Larsen**

(1): The t-tests and chi square tests seem adequate. In the revised article Fisher's exact has been replaced with the more appropriate exact chi square test, and in mixed effects models the assumption of no baseline differences has appropriately been discarded since the groups compared are not randomized groups (2): See (1).

(3): "Only homogeneous data can be averaged.": I don't understand this requirement, exactly homogeneous data are probably rare. Should averaging almost never be used? Anyway, no problems with this matter detected.

" Standard deviations are preferred to standard errors.": Standard deviations, not standard errors, have been reported used in descriptive contexts. For model based estimates there is no such thing a standard deviation, in these cases the article mostly reports confidence intervals that are superior to reporting standard errors (except in the abstract presumably due to lack of space).

" The number of observations and subjects ( $n$ ) is given.": No problems detected.

" Losses in observations, such as drop-outs from the study, are reported;": It might have been possible to investigate closer the relationships between baseline characteristics and dropout at the different time points.

(4): Confidence intervals, based on the models used, are computed. There are no ED50, LD50, IC50 or probit models in the article.

(5): This is a reasonable requirement in Results, while p-values should normally not be stated in Discussion. Statements of e. g. 'no differences' based on high p-values alone have been replaced throughout the article by more appropriate statements.

Oslo 1/9 - 2016, Tore Wentzel-Larsen

Befym, 09.09.2016  
Erik Johnsen  
ERIK JOHNSEN