Reviewer’s report

Title: Identification of early changes in specific symptoms that predict longer-term response to atypical antipsychotics in the treatment of patients with schizophrenia

Version: 1 Date: 21 July 2010

Reviewer: Eske M. Derks

Reviewer’s report:

This is a well written manuscript, in which it is investigated whether response to treatment (>= 30% reduction in PANSS total score) can be predicted by individual PANSS symptoms rated at one and two weeks after baseline. The learning set consisted of six randomized clinical trials with a duration of at least eight weeks. The prediction model was evaluated in two independent clinical trials. I have one major question with respect to this paper.

Major compulsory revisions

1) The authors show that a model with 2 predictors (a 2-point decrease in at least two out of five individual panss items, and a 2-point decrease in the PANSS excitement item) is the best model. However, this model was not compared to other models, for example including PANSS total scores or CGI. The authors do mention that it is advantageous if prediction of outcome can be based on only a subset of the PANSS items, but would they recommend collecting only these items in clinical or research settings instead of a complete PANSS assessment? And what about the CGI, maybe the same PPP and NPP would be feasible based only on CGI improvement and CGI severity?

Other comments (Minor essential revisions)

2) The authors state that the training and the learning set are very similar but the duration of illness and PANSS baseline scores appear to be quite different

3) How large was the dropout in each of the clinical trials?

4) What was the percentage of patients with >=30% improvement in the two independent studies in which the model was tested? Without baseline response rates it is difficult to evaluate PPP and NPP especially since specificity and sensitivity are not reported

5) I think that if the authors want to show the analysis of model performance by study and compound that they should also formally test whether performance is different (the statement “suggested that the creation of the model had not been driven by any one study or any one compound … but rather was a reasonable reflection of its component parts”. To me the numbers in Table 3 and Table 4 show that there is quite large variation between study and compound.
**Level of interest:** An article of importance in its field

**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