Reviewer's report

Title: Family-based clusters of cognitive test performance in familial schizophrenia

Version: 2 Date: 10 May 2004

Reviewer: William Horan

Reviewer's report:

General

-----------------------------------------------------------------------------------------------

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

This paper addresses two related but distinct goals: 1) The identification of meaningful subgroups of families containing an individual affected by schizophrenia using neurocognitive measures and 2) The introduction of a new clustering algorithm and visualization procedure for family-based data. While these two goals are interesting and important, neither is fully addressed in the manuscript as it is written.

Comments regarding goal 1: Identification of neurocognitively based subgroups of families

The authors provide a reasonable rationale for why the identification of more homogeneous subgroups of families would benefit genetic studies of schizophrenia. However, a compelling case that three valid, meaningful subgroups that are not merely arbitrary demarcations on a continuum of severity of cognitive impairment is not provided. The fact that the cluster algorithm identified three clusters only partly supports the validity of these putative subgroups. The clusters demonstrate hardly any differences on the clinical variables examined. The authors do not discuss whether the clusters demonstrate any meaningful differences in their patterns of cognitive impairment, aside from overall severity of cognitive impairment. If I understand the results correctly, the 3 clusters account for only 27% of the total variation in neuropsychological test performance.

The approach of using families, rather than individuals, in a cluster analysis is novel in schizophrenia research and seems potentially informative. However, might there be potential downsides to lumping affected and unaffected relatives into a single cluster analysis? For example, might the inclusion of unaffected siblings, who presumably have minimal cognitive deficits, merely add “noise” in the identification of patient subgroups that differ in pattern as well as level of cognitive impairment? If the three identified subgroups are indeed meaningful, would one also expect a similar structure to emerge in cluster analyses conducted separately for affected and unaffected subjects?

The description of the affected and unaffected cases on p. 4 is quite cursory. What were the overall exclusion criteria (e.g., neurological conditions, substance use)? What was the patients’ symptom status at the time of testing? It is stated that the final sample includes 165 subjects with a psychiatric diagnosis – I don’t get the same total – 82 schizophrenia + 13 schizoaffective + 12 bipolar + 48 non-psychotic. Were there more family members with psychiatric diagnoses? P.10 refers to subjects with “affective psychotic disorders” – does this refer to the bipolar subjects?

Were the unaffected siblings screened for schizophrenia spectrum disorders (e.g., paranoid, schizotypal, schizoid)? There seems to be a fairly high prevalence of other psychiatric disorders in the unaffected relatives. Might any cognitive deficits in these “unaffected” relatives reflect
diagnosable spectrum disorders or other Axis 1 disorders rather than a latent vulnerability to schizophrenia?

The characteristics of this particular sample (i.e., multiple affected siblings) are unique. While the authors mention this as a potential limitation, the generalizability of these findings seems highly questionable.

In the Discussion section (p 10, paragraph 1), the authors suggest that the severely impaired subgroup would be the most informative for genetic analyses. Further explanation of the logic here would be helpful. If the clusters each identify meaningful schizophrenia subtypes, why would one particular cluster be more appropriate or promising? The final sentence of this paragraph (“In our analysis….”) is difficult to follow.

Comments on topic 2: Introduction of clustering algorithm for family based data

My main comment on this section is that the description of the procedure that determines the degree of similarity/dissimilarity within each family is far too brief (p. 6) - pairwise test performance differences between the family members are only briefly mentioned on p. 7. This within-family index of similarity seems to be the main contribution of the new procedure – a significantly more detailed explanation would increase the potential usefulness of this procedure for other researchers.

Much of the description of the visualization procedure appears in the figure captions. I would suggest providing a more comprehensive, sequential description of the cluster and visualization procedures within the methods section that guides the reader through the process. I wonder whether a psychiatric journal is the most appropriate medium for a full description of the cluster analysis and visualization procedure.

A third figure (scatter diagram) is included but is not discussed in the text.

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

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: Not suitable for publication unless extensively edited

Statistical review: Yes

Declaration of competing interests: