Reviewer's report

Title: Brain size and brain/intracranial volume ratio in major mental illness

Version: 1 Date: 19 January 2010

Reviewer: David Tate

Reviewer's report:

Abstract
The authors refer to “hand cutting” methods as being the initial analytic tool. I am not sure that is an accurate description of the methods used in this study though the structure of the sentence is difficult to follow. If indeed it was meant to be a description of the methods used in this paper, I think it is a bit misleading as the methods are based on MRI scans and not actual brains. I think you should eliminate these descriptions in the abstract to make sure there is no confusion.

Introduction
The introduction could be shortened significantly. The authors spend a great deal of time documenting the history of imaging studies spending a disproportionate amount of time on pneumoencephalography which is by far the most removed method to the methods utilized in this study.

The hypotheses of this particular study could be spelled out more directly. What findings did you expect to get in this study? ICV is usually considered invariant in late adolescents. It is unclear from the intro why you might expect ICV to be altered other than a brief reference (uncited reference) to schizophrenia being a developmental disorder.

Overall, I think the intro needs to be a bit more focused with regards to the hypotheses.

Methods
Though the authors claim that schizoaffective disorder is more prevalent in the Denver Medical system, yet the schizoaffective group is the smallest sample (<30 subjects). Any explanation for that?

With regards to variability in quantitative measures, it is interesting that you have varying numbers of males in each group (47% to 69% males). Each of these measures is highly dependent on head size and though you control for gender in some of the analyses, it is unclear to me

Though TBV, ICV, VV, and CSF volumes are typically invariant to variations in scanners, it would be helpful to know the distribution of images over time as well. It is amazing to me that the resolution of the scans has been kept that consistent over such a long period of time.
I am not that familiar with the IDL software though as far as I understand it, it is a general purpose mathematical analysis and programming language much like MatLab. Are you using it to run a specific program to perform you analyses or are you using some sort of homegrown software developed in IDL? In other words, can you provide more information in this part of the manuscript?

Maybe you could take a look at several other ratio measures using the data that you have as the ratio data are more likely to reveal any difference due to the fact that the ratio uses the person’s head size as a way to make the variables more directly comparable. For example, you could also take a look at VBR (ventricle to brain ratio) which is a widely used and reported measure in the literature.

Results
Are you simple testing for gender differences without regard to the groups in the first paragraph of the results section (page 13)?

The authors suggest (in several places) that there is a significant amount of variability in the quantitative MRI measures though it is difficult to tell unless you go to several tables that are provided as supplemental information. Unless these tables will be readily available in the published text, I would suggest that the authors include the means and SD in the text so that readers can more readily access this information or include it in one of the tables provided with the text.

Since the group main effect for TBV/ICV ratio had a p-value of 0.05 I would suggest you reword and say only marginally significant rather than significant.

You only report the post hoc test for the TBV/ICV ratio, were there no other post hoc tests that were significant for the other measures especially total brain volume?

The authors do not provide any data regarding rater reliability in obtaining the measures. Certainly, if the quantitative MRI post-processed data were obtained as the images were acquired (since 1992), there could be some changes in the way a rater produced the data. Could you please provide clarification and/or data that speaks to rater reliability when producing these quantitative measures?

Discussion
The discussion seems to be rather long. I appreciate the various subheading but it appears that the data could be discussed in a more succinct fashion especially since there were limited significant findings.

Some of the points for reasons behind the history of equivocal results are well taken.

The significant finding for TBV/TICV ratio is interesting and maybe underestimated by the authors in some respects. In this age range, TICV is known to be invariant. It was long ago (late adolescence) determined by dynamic forces of brain growth and then remains stable the rest of the patient’s life. So, the variable that is actually changing in this ratio would be brain volume. Since
each patient’s maximal head size (ICV) was used in the ratio, this makes the TBV more directly comparable between the groups. You already were seeing some reduction in TBV that was “trending” toward significance so when you examined the group differences while there were no such trends in ICV. The overall head size variability may have been obscuring the difference in raw TBV even though you controlled for gender, though this would not have been as good a control as the actual head size. I think these relatively small atrophic observations are difficult to interpret in clinically meaningful ways but are nonetheless observable differences.

Intra and inter rater reliability is certainly a problem within this study and may also be contributing to your variability. It is true that your SD are similar in range to other studies but it would be important for the authors to acknowledge and document possible sources of variability within their own study.

The whole methodological issues discussion section seems to lack relevance to the particularly study and the methods employed.

The discussion of altitude is interesting but a bit esoteric. There should be other potential sources of variance that the authors could look to before altitude. For example, the fact that the paper is looking at gross measures of brain volume to explain mental illness my be an oversimplification of the problem and it may be combinations of interrelated brain areas that might better explain the differences between patient groups.

Tables
Seems to me that Tables 1 and 2 could be one table.

**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