Reviewer's report

Title: Valid and efficient manual estimates of intracranial volume from magnetic resonance images

Version: 3
Date: 15 July 2014
Reviewer: Alexander Hammers

Reviewer's report:

Intracranial volume (ICV) is an important and widely used reference against which regional brain volumes are compared to correct for premorbid brain size. Various methods exist for manually or automatically estimating ICV.

The authors assessed the influence of various methodological choices when manually determining ICV. Their reference standard were manually determined ICVs based on magnetic resonance (MR) head images in 62 participants of unknown age. About half were controls, 8/62 had subjective or mild cognitive impairment, and the remainder were demented. Intracranial volume was manually determined on three-dimensional T1 weighted MR images of the head, acquired with high resolution at 1.5T and downsampled to 1 mm isotropic resolution using an unknown interpolation technique. Image size, brightness and contrast were adjusted.

ICVs were then manually outlined on the millimetric slices following the usual protocol by Eritaia et al. (Magn Reson Med 2000 44:973-977). Inter-rater as well as intra-rater reliability of the measurements was assessed on every 10th or 40th slice in a subgroup.

ICV estimates were then derived from the manually obtained data by changing orientations (coronal / sagittal / transverse), spacing of slices (2 – 50 mm), starting slice, and method for obtaining the volume. The latter was either piecewise constant interpolation (Cavalieri 1635), piecewise linear interpolation, or cubic spline interpolation.

Assessment was via Pearson linear correlation coefficients and intraclass correlation coefficients, but not via overlaps.

With the caveat that coronal and transverse slices were not actually manually traced, the results confirm and extend previous studies: as long as slice spacings below 10-15 mm are used, all orientations and interpolation methods perform very well.

This study extends previous results to extreme conditions, e.g. sampling of slices every 30-50 mm only, where differences between the methodological choices emerge.

BMC requires reviewers to consider a number of set points. The question posed by the authors is well defined. The methods are appropriate and, bar some details pointed out below, well described. The data are sound. In my opinion, the
manuscript adheres to the relevant standards for reporting and data deposition. The discussion and conclusions are well balanced and adequately supported by the data, bar some details pointed out below. Some minor limitations of the work should be more clearly stated, see comments below. The authors acknowledge much of the published work upon which they are building; logically I am not in a position to assess whether they acknowledge unpublished work! Some thorough early studies on the importance of standardisation are not cited (e.g. Free SL et al. Am J Neuroradiol 1995 16:637-645). The title and abstract largely accurately convey what has been found; see below for some minor suggestions. The writing is entirely acceptable, with very few mistakes.

On the whole, this is an extremely thorough study considering a large number of methodological details in estimating ICV. Overall it confirms current practice (i.e. the use of slice spacing of no more than 10-15 mm). It also reminds of the importance of stating such details in publications; replicates and extends standard studies in the field; and will be a useful resource for researchers deciding on their methodology for ICV determination. Given that it takes days to identify, screen, include, and scan subjects in research studies, the importance of reducing analysis times by some minutes (e.g. section 2.3) at the risk of more unreliable estimates can be disputed.

Protocols for manual outlining are very important (e.g. Bergin PS et al. Neurology 1994 44:1770-1771, Hammers A et al. Hum Brain Mapping 2003 19:224-247). As the authors correctly state in the Discussion, it has not formally been shown here that outlining on coronal or transverse sections is practically possible; researchers wanting to use those orientations should therefore conduct additional assessments of feasibility and reliability.

“Major compulsory revisions”

This is the compulsory heading provided by BMC Medical Imaging.

1. Section 2.1: Age is important but not given.

2. Section 2.2: The sequence is given as 3D with a slice thickness of 1mm but also a spacing between slices of 1mm which would be unusual. Presumably this is simply a tautology, and the spacing can be left out?

3. Section 2.3: The MR images were reformatted, changing voxel sizes in two out of three orientations – but the interpolation method used for achieving this is not given. Interpolation can have a major influence on more detailed measurements (e.g. Ahsan RL et al. Neuroimage 2007 38:261-270). While it is highly unlikely this would make much of a difference for large structures like intracranial volume, given the level of detail in the remainder of the paper, the method should be added.

4. Section 2.4: Please explain how a spacing of n slices / millimetres in x datasets leads to n^x combinations. In my opinion, as the original millimetric data are unchanged, there are only n ways of starting the sampling for each x, resulting in a much more manageable n times x different estimates per direction etc.
5. Section 2.5: Why were no overlaps used for assessing quality? Jaccard indices would presumably have shown much smaller overlap with ground truth for the piecewise constant interpolation method for larger sampling spacing and would have allowed to highlight this difference between the methods which is not captured by the correlations used.

6. No actual intracranial volumes are shown in the paper. This is an oversight that has happened before in the writing of papers on ICV estimation (e.g. Keihaninejad S et al. Neuroimage 2010 50:1427-1437), but will be very useful for sanity checking for subsequent authors using the methods described. It would also be useful for the authors to check whether the ICVs they obtain “behave” as expected – i.e. are insensitive to age, but sensitive to gender (as a proxy to body height).

“Minor essential revisions”

1. Section 2.1: It probably does not matter for the purposes of this study, but if the study is called the “Gothenburg MCI study”, one can reasonably expect participants to actually have MCI. Yet, 25/62 were demented, 29/62 were controls, and only 8/62 were somewhere in between. It is probably impossible to retrospectively rename the study but “MCI study” gives the wrong impression.

2. Section 2.1: “who were remitted”: “who were referred”

3. Section 2.2: The use of capitals for the first word of each of the acquisition parameters is superfluous.

4. Section 2.3, 2nd paragraph, “gamma correction”: This should be explained.

5. Section 2.3: “were sagittally oriented”: I.e. in the acquisition orientation? Or was the inclination of the head also standardised, leading to further interpolation?

6. Section 2.3: “undersurface”: inferior surface

7. Section 2.3, 3rd paragraph: “135.9”: Artificial precision. I suggest “136”, complemented with some measure of spread, e.g. a standard deviation.

8. Section 2.3, last sentence: “Both raters were blinded to participant age, gender and cognitive status during all ICV segmentations”: This is very thorough. However, when this level of detail is given, logically even the obvious should be stated: presumably they were also blinded to the most important aspect that can skew manual measurements, i.e. the previous tracing?

9. Section 2.5, 2nd paragraph: According to the text, errors were calculated for two of the three interpolations only, but Figure 1 shows all three. The text could therefore be simplified to “using all three interpolation methods”.

10. Discussion, 3rd paragraph, last line: “established use <…> were”: “established use <…> was”

11. Discussion, 4th paragraph: “possible difficulties connected with actually segmenting in coronal or transversal orientation are not covered”: I believe this is an important point that deserves more prominence and should be mentioned in the abstract or Conclusions.
12. Discussion, 6th paragraph: “a correlation of 0.937”: Pearson?

13. Discussion, 6th paragraph: The authors benchmark their results against one study on automatically deriving ICVs only. In addition, correlations (Pearson’s?) are used which are suboptimal as they do not assess potentially large differences in slope and offset which could render comparisons of different studies meaningless. ICCs or measures of overlap (e.g. Jaccard, Dice) should be used instead for comparisons.

14. Discussion, 6th paragraph: I am aware of at least one study of automatically derived ICVs that achieved ICCs of automatic versus manual measurements in the 0.98 – 0.99 range (Keihaninejad S et al. Neuroimage 2010 50:1427-1437, Tables 1-3), i.e. close to the accuracy achieved here with manual test-retest data, and likely much better than the Pearson correlations cited for other automatic methods here. The section could be slightly expanded to put the importance of manual sampling in context.

15. Section 6: MRI stands for image in the singular in the present paper.

16. References: Reference 1 includes an abbreviation that I believe is unofficial. I have noted similar problems for reference 7 (unofficial abbreviation), reference 10 and 15 (unabbreviated), and there may be others.

17. Figure 4: Units missing on y axis

“Discretionary revisions”

18. Section 2.5: The nomenclature for the various ICC variants can be confusing. It would be least ambiguous if the formula used could be given.


20. Discussion, 2nd paragraph: “A similar relationship”: As ICCs were identical to the third digit after the decimal point, “A relationship that was identical for practical purposes” would, in my opinion, better convey the message of this paper.

21. Figure 1: Purple and blue are difficult to distinguish here.

22. Section 8: This is a discretionary comment which will not be published. It is always fraught with difficulty commenting on such issues from the outside, but I would like to make sure that the authors have reflected whether learning the technique and performing inter-rater measurements is a smaller contribution -to this particular manuscript- (only resulting in acknowledgement in section 10) than conceiving and designing the original Gothenburg MCI study (resulting in co-authorship), or whether co-authorship would also be warranted for the student.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being
Published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

Potential non-financial competing interest:
I am senior author on Keihaninejad S et al. Neuroimage 2010. In that paper, we described an automatic method to determine intracranial volume which was insensitive to field strength differences (1.5T and 3T) and equivalent to manual estimation (intraclass correlation coefficient 0.99) and could hence be reasonably seen as competing with the methods shown in the current BMC Medical Imaging paper.