Tina Memo No. 2000-010
Internal Report.

Strategies for Identification and Mark-up of Common Brain Structures Visible on Transverse Sections.

Marietta Scott and N.A.Thacker.

Last updated
20 / 12 / 2000

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Strategies for Identification and Mark-up of Common Brain Structures Visible on Transverse Sections.

Marietta Scott, Neil Thacker

1 Introduction

The aim of this document is to outline the general problems of measuring the volume and cross-sectional area of any 3D structure, and then more specifically, to provide a standard method of measuring brain structures. This work is to support the marking-up of structures in the Research into Ageing project. The images used in this project are obtained using MRI, and are T1 inverse-recovery images, taken in the transverse plane. For this reason, this paper mainly covers those structures which can be best visualised using these slices.

1.1 Problems in measuring structures

The measurement of anatomical regions for morphometric analysis of brain structures in disease is common (with many papers published making use of "medieval scribe" techniques) and naively considered to be easy. Typically, subjective approaches are taken, where clinicians mark up (by hand) neuro-anatomical features on the basis of land-marks in specific image types (eg: T1 weighted MR scans). However, doing this in a way which is to prove more clinically informative than anecdotal correlations between groups is likely to be difficult.

Inaccuracies in these techniques, which lead to poor repeatability, arise for three main reasons:

- there is no general agreement (even by experts) about the anatomical boundaries of many structures.
- the structures often have indistinct boundaries, and land-marks which delineate regions reliably are scarce.
- In addition, the software tools used to identify structures often sacrifice accuracy in preference to user friendliness (eg: using manually selected thresholds in preference to data driven approaches).

The frequently emphasised issue of providing the researcher with repeatable viewing conditions does little to address these problems.

When undertaking the mark up of datasets it is also worth accepting from the outset the potential limitations of such an analysis, so as to avoid expending large amounts of effort which result in inconclusive or uninformative data. For example, marking up structures and looking for group level correlation with disease may give some indication of the areas affected but will rarely result in data which can effectively be used as an aid in diagnosis. Mainly for one of these two reasons:

- correlations are often only sought between two groups of subjects (generally one being the control) on several individual volumetric variables. Any such correlation will have little diagnostic value outside of that one, forced choice scenario. A patient who is not already known to fall in either of the two groups cannot be diagnosed.
- a structure which atrophies in all atrophic diseases (for example the hippocampus) has very little diagnostic specificity, so clinically its measurement is unimportant.

In order to overcome these limitations and obtain more informative data we require a strategy. We clearly require the observation of several significant variables in order to gain any real understanding of spatially correlated morphologcal change from multiple potential causes. These measurements need to have good reproducibility so that small differences can be monitored.

We have said that the difficulty is one of reproducibility and not absolute measurement and this is an important distinction. From the point of view of quantitative data analysis we would be quite happy

with a measurement which was reproducible even if it gave systematically different values from the anatomically accepted (but less reproducible) definition. Given the difficulties listed above, in order to develop quantitative techniques which make use of the data available we may expect to have to generate our own delineations for structures, which may not correspond to any precise anatomical definitions. To do otherwise, and to attempt to extract information which is clearly not reliably present in the data can only invite subjectivity and bias. In other work we have taken this observation to its extreme and defined simple rectangular regions which have no anatomical relevance whatsoever. Even so we have demonstrated that useful, diagnostic data can be easily extracted reliably and automatically from MR scans by learning the associations between reproducible measurements and disease. Such automation would currently be very difficult based upon anatomical structures.

In order to make specific evaluations on the longitudinal change of a structure in an individual we may well need to make measurements with percent level repeatability. This involves minimising the great degree of variability between both the size and shape of brains of even normal subjects. Given that most analysis techniques are based on the null hypothesis of equivalent distributions, this is not a good situation [1]. Therefore before any analysis, standardisation to a normalised brain size must take place [2], although obviously this will not correct for the slight variations in the positions of the structures. To overcome difficulties in repeatability we must also make concerted efforts to remove likely unwanted variability from the **measurement process** wherever we can. The first strategy we can use is to select regions and define measurement variables which can be obtained from stable configurations and easily identifiable structures. Following this we can also attempt to automate location and measurement wherever possible in order to eliminate subjective effects and produce a transferable methodology. This document deals with the first of these issues. It is intended that automated approaches for regional measurement (based on pixel level segmentation and structure location) will be applied as our experience of this problem develops. There is clearly little point in attempting to automate structure extraction until we have some idea of what structures can be reliably delineated and are of practical value, as we will require manual gold standard examples to build such a system.

While discussing the general problems in this area it is probably also important to mention the method of "grey-level morphology" which has gained popularity recently. As this technique attempts to unambigoiusly match the movement of tissue between two MR scans it is best used in the context of a longitudional study. However, the approach has so many flawed assumptions that we believe unambiguous interpretation of any results is difficult. We can expect the popularity of this technique to wane as these problems become understood by researchers in this area. In the meantime we suggest that the technique would best be avoided.

The points covered here may seem obvious in some cases but have been specifically selected as issues which are frequently understated in the current literature.

1.2 The need for Standardised Alignment

In order to make quantitative measurements of volume and cross-sectional area we would prefer the brain to be at a standard orientation. The MRI radiographer is very experienced at aligning the axes of scan with the patients brain in the same way every time. This is done as standard proceedure on any acquisition.

In the transverse-axial view, the scan slices are taken parallel to a line drawn between the inferior borders of the corpus callosum. For this study, the top slice of the scan is at the level of the skin on the top of the skull, so all the CSF at the top of the brain is included in the scans. If the patient's head is tilted or rotated, this is corrected for. If the brain is tilted, it is aligned perpendicularly to the medial sagittal plane. The same offsets are used for every scan taken in each sitting, so that there is coherence between scans. If the same patient is rescanned, that patient's brain can be aligned to within 1mm of the original alignment. Therefore, there is also high repeatability between patients, although the position of the corpus callosum will affect orientation somewhat.

Before assuming that a complex deformable co-registration of data is needed in order to analyse the data it is well worth considering how much can be achieved without it. Segmentation and measurement strategies can be selected which minimise alignment problems and experience to date indicates that

[1] The extra variability associated with normal variation reduces the statistical power of any test.

[2] Analysis of the data including age or other factors as a confounding variable will always be less effective than direct correction for the expected normal dependancy.

normal variation may generally dominate the measurement process. Alignemnt of the entire brain may still leave orientaton differences between structures in different individuals. Re-alignment to a standard brain orientation via co-registration will only be used if it is found that structures cannot be marked accurately enough in comparison to normal variation. We are aware that this criteria is not considered by many researchers working in this area, but this approach is entirely justified when faced with practical limitations on resources and an absence of software or techniques which can be adequately trusted to solve this problem automatically.

1.3 Processing the obtained data

Before considering the specific structures, it is worth noting that when comparing structures between patient groups, normalisation of the data to head size has to take place, so that as much intra-group variability as possible is accounted for. If the structures are not normalised, any large differences in head size in only a few patients are highly likely to produce a significant difference between the patient groups in any standard null hypothesis test with small (ie: tens) quantities of statistics due to non conformance of the sample distribution to the assumed Gaussian.

Volumes will be normalised to a total brain volume defined as the minimum bounding box of the cranial cavity (with a robust anatomical landmark for the lowest point), and cross-sectional areas should be normalised to the corresponding cross-sectional area of the brain. This approach is justified on the assuption that all normal brains can be aligned by linear deformation (up to a scaled homology). Equivalent brain volumes can therefore be obtained via a mechanism analogous to "similar triangles" in trigonometry, but here it might be better termed "similar brains". Specifically, ratio changes in the bounding box of the brain should transfer to ratio changes in the volumes of anatomical structures.

The specifics of the particular statistical test used for individual hypotheses will vary, but it should also be stressed that the test hypothesis needs to be decided before the data is analysed, and that no analysis should be made on the raw data in the hope of finding chance correlations. Indeed, any correlation found in raw data and not found in normalised data would necessarily need to be treated with scepticism.

1.4 Stability of cross-sectional area (CSA) and volume measurements

As a slice is orientation dependent, inaccuracies in brain alignment will affect the accuracy of the variable obtained depending upon the shape of each structure.

The notion of stability incorporates two ideas. The first is that the CSA should be reasonably independent of any slight rotation in the orientation of the slice. This ensures that the area sampled is most representative of the actual area of the structure you are measuring.

The second is that of measurement accuracy. If a particular structure is only a few slices long, and it is ambiguous as to whether to include an extra slice, the possible error in the volume of the structure is very large. For example, an extra slice in a structure across six slices gives an extra 18 % in the measurment. In such a case the volume is unstable. If you are looking at possible percent level differences, such a change in volume due to lack of repeatability is unacceptable.

As a logical step, the structure studied in a particular orientation must persist for at least several slices, and it would be most sensible to move through the slices along the longest axis of the structure. If a stable orientation cannot be found then a CSA shold be selected in preference. Such cases are easy to identify, and there is little point in persisting in measuring the volume across all slices only to find the data to be of no quantitative value.

1.5 Indistinct boundaries and land-marks

A major problem when studying brain structures (and possibly in other tissues as well) is that the neurones in the brain do not form nice neat discrete structures, but flow from one region to another. If it is possible to obtain a closed volume of the structure then this would be the preferred measurement, as it is a measurement which is independent of the orientation of the structure. In cases where structures cannot be accurately delineated however, once again the cross-sectional area (CSA) may be preferred.

Conventional delineation arises from the anatomists desire to label structures and has often been aided

by microscopic evaluation which gives data which is unavailable in a conventional medical image. This leads to two problems. First, it may not be possible to determine the beginning and/or end of a structure. Secondly, you may know where the structures begin and end, as defined by anatomical landmarks, but they may be open ended.

In the first case, there are two possible solutions: use the maximum CSA of a structure, and compare that between groups, or, plot the CSA against slice number, and obtain a particular signature for that structure, which can then be used to compare between groups.

In the second case, use of landmarks may cause a problem. It is possible that a landmark used is not close to the structure being considered, so that if the slice is rotated slightly, the landmark is now associated with a different part of the structure, which may influence the inclusion/exclusion of a slice (leading to stability problems as described above). Also, it is possible that the structures (and landmarks) are not going to be in exactly the same place in all subjects. If this occurs in a manner which correlates with disease then not only is the measurement potentially unstable but apparent changes in volume may actually arise due to systematic movement of the land-mark. Landmarks which are not in close proximity to the structure being deliniated should therefore be avoided.

It is common to divide up cortical grey and white matter using the sulci (the infoldings of the cortex). However, this technique is highly subjective; no two persons could place the division in the same place, not every subject has the same number of cortical folds, and they are almost certainly positioned in slightly different places in everyone. Add to this the fact that the functions of the different parts of the cortex cannot be easily placed, and it is easy to see why for this study we are not considering cortical structures. We would also advise other to do the same [3].

1.6 True edges

The resolution of the image determines how well you can see the edge of a structure. A low resolution image may over-estimate the area of a structure, but there is a trade-off at higher resolutions between the gain in accuracy achieved by painstakingly drawing around all the nooks and crannies and the time taken. We recommend that researchers are realistic about the time they take for structure mark up with respect to the accuracy improvements gained.

With structures which taper to a point or edge, it is unnecessary to be overly precise, as the number of pixels involved may be so few that when considering volumes (of cubed order), a few pixels will contribute negligibly to measuring the volume. For longitudinal studies of patients we would ideally like to be able to monitor percentage level changes in structure, this should be borne in mind when developing measurements so that overlong manual segmentation can be avoided where possible.

The rest of this document focuses on those structures which can be viewed on transverse slices, and analyses them in terms of stability, and CSA/volume measurements.

2 Structures visible on transverse slices

2.1 Caudate Nucleus

The caudate nucleus has been shown to be smaller in depressed patients, so its measurement is very important in this study.

The bottom border of the caudate is at the level of the anterior commisure (see fig. 1) (ie appears on the same slice). Below this point, the caudate joins the putamen, and the region is called the substantia nominata.

Finding the top of the caudate nucleus is more difficult. The caudate has a head and a tail, as seen in fig. 2. When taking transverse sections of the brain, it is not immediately obvious what is head and tail. The top of the caudate is shown on the slice where the chordothalamic notches appear, and in this slice, anything anterior to the chordothalamic notch is part of the head of the caudate, the rest is tail.

For the purposes of this study, only the head is required. As the head is roughly spherical, the volume can be obtained, and the cross-sectional area is stable in this orientation. In addition, the structure is

[3] If your analysis really needs volumetric analysis of such structures then perhaps you should not be using MR data.

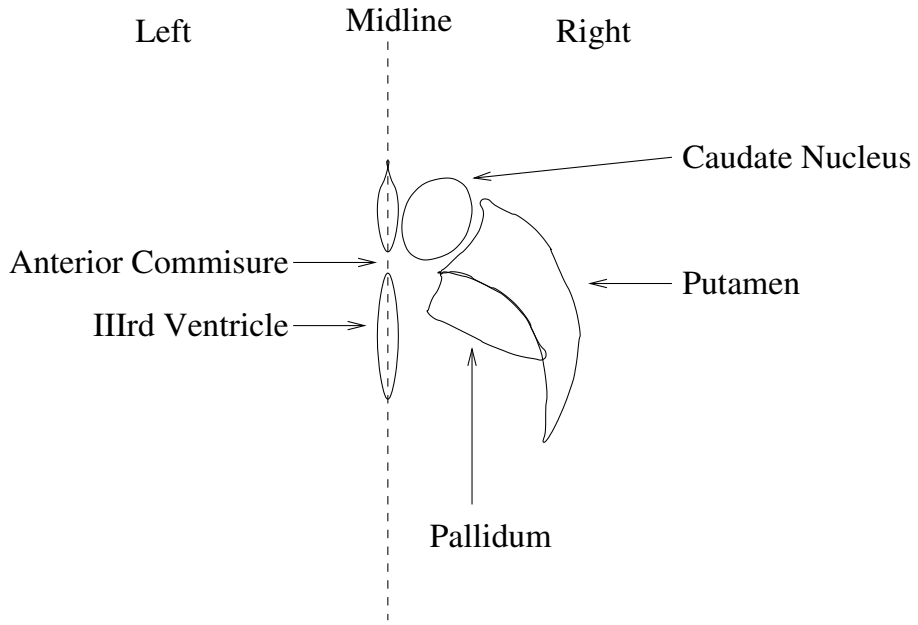


Figure 1: Transverse view of the right-hand side positioning of the Caudate Nucleus, Putamen and Pallidum at the level of the anterior commisure

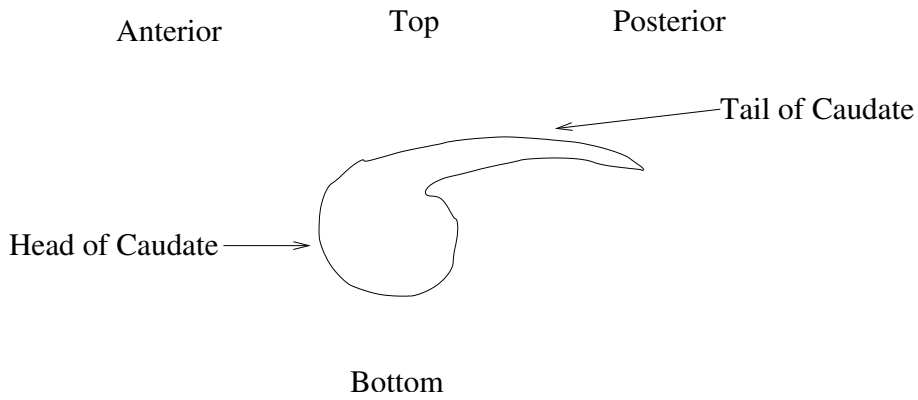


Figure 2: Sagittal view of the shape of the caudate nucleus

closed, so there is stability in terms of the number of slices.

The caudate nucleus covers approximately seven slices of our MRI images, and is shown on slices 4-10 at the end of this document.

2.2 Putamen

As with the caudate, the bottom of the putamen is at the level of the anterior commisure. It is important to differentiate between the putamen and the pallidum, which is a slightly paler grey.

The putamen persists for approximately six slices, until it is no longer evident. The putamen is shown on slices 6-10. Here the markers must trust their anatomical knowledge and instincts to determine the top of the head.

In terms of orientation, the CSA is stable, but in terms of number of slices there is some instability, due to observer confidence in choosing the right slice. The potentisl error in the volume if an extra slice is included or one is excluded is 18%.

2.3 Pallidum

Again, the structure starts at the anterior commisure. It is lateral to the internal capsule, and next to the putamen, as mentioned earlier. As you get older, the pallidum builds up iron depositions, which will affect its appearance on the scans of elderly patients, but until the scans are examined, it is not possible to say if it will make the marking-up clearer.

Given the difficulty of separating the pallidum from the putamen, there may be large inter-marker errors, so it will be necessary to perform reliability tests (and not just for the pallidum).

As with the putamen, the end of the pallidum is open, and as it persists for only two to three slices (slices 8-10), there is enormous instability in the volume, with the error possibly being between 33 and 50%.

2.4 Thalamus

The base of the thalamus sits on the medial lemniscus (which is white matter), and is lateral to the internal capsule. Any grey matter seen here is due to the bottom of the thalamus, but this is not always distinct. The thalamus persists for approximately four-five slices (slices 4-9 at rear of document). Its head sharply rounds off, so the thalamus 'unexpectedly' disappears. The pulvina should be excluded.

The thalamus has a definite start and end in this orientation and is fairly spherical, so the CSA measurements are stable and the volume can be calculated.

2.5 Brain stem

The main problem with the brain stem is that it is impossible to determine the top. The total volume therefore cannot be calculated, and the 'best' slice to calculate the cross-sectional area is unknown, and could not be compared between patients.

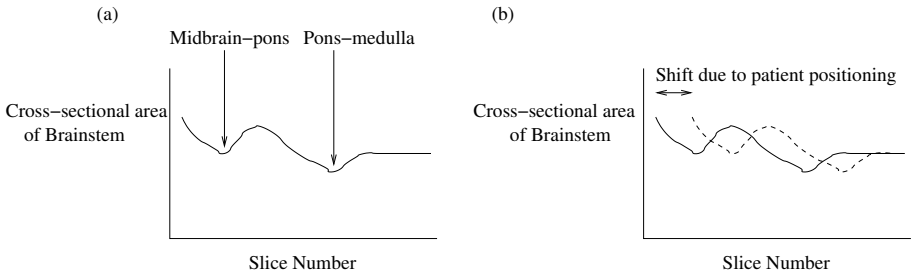


Figure 3: Graphs showing the general form of CSA/slice number for the brainstem, for one patient (a) and two patients (b) with variation due to positioning

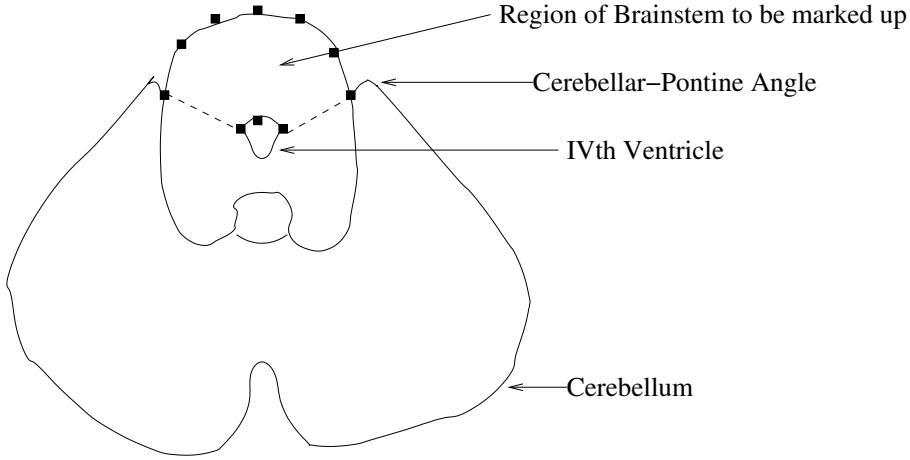


Figure 4: Transverse view showing region of Brainstem to be marked up at the level of the pons

Instead, the whole of the brainstem will be drawn around, and the area against slice number will be plotted, to give the characteristic plot of the brainstem, as shown in fig. 3. This graph shows two dips, which correspond to the midbrain-pons and pons-medulla junctions. The plots for all the patients can be fit to a model, or co-registered with each other using spatial corrections to correct for the fact that the slices may have sampled the brain in slightly different positions in different patients.

It is assumed that there is not enough variation between the normalised brains to cause errors in the co-registration which are wrongly attributed to patient positioning.

The start slice for the brainstem is chosen at the point where it is isolated, where the brainstem is anterior to the supatectal system.

About half-way down the brainstem, there is a small circle of CFS, which is not to be cut out, as the cutting procedure will result in greater inaccuracies than leaving it in.

When the pons is reached, and the 4th ventricle becomes apparent, the white matter posterior to the 4th ventricle should not be included as brainstem. This is achieved by drawing a straight line from the corner of the floor (obex) of the fourth ventricle to the cerebellar pontine angle, as shown in fig. 4.

Marking up should extend to the medulla.

The brainstem is a long tube, and has clearly defined boundaries with the surrounding tissue. This orientation is therefore stable. The brainstem is shown on slices 12-23 at the rear of this document.

2.6 Cerebellum and Vermis

The cerebellum has a foliate edge, so marking around the apparent edge on the images will lead to wildly inaccurate measurements of the cross-sectional area. It has therefore been suggested that the cerebellum CSA should be found from the grey matter probability maps. The left and right cerebellum will be measured seperately in order to obtain a measure of symmetry, and the vermis will be measured seperately from the cerebellum. The reason for this is that the vermis is developmentally different and vascularly seperated from the cerebellum, and in the patients should contain a lot of grey matter whereas the cerebellar structure will be very poor.

The cerebellum is given on slices 11-23.

2.7 3rd Ventricle

In sagittal section, the 3rd ventricle is straightforward to see (see fig. 5). However, as the third ventricle is very narrow, the view is unstable. The lower border of the 3rd ventricle is the optic chiasm, where the ventricle takes on a cross shape. Moving up through the slices, past the anterior commisure so that the posterior commisure is visible, the anatomy of the ventricle and surrounding tissues is given in fig. 6. In the next slice up, the 3rd ventricle becomes a small slit from the foramen of Munro to the inter-forneal recess. The top of the ventricle is when the foramen of Munro is lost.

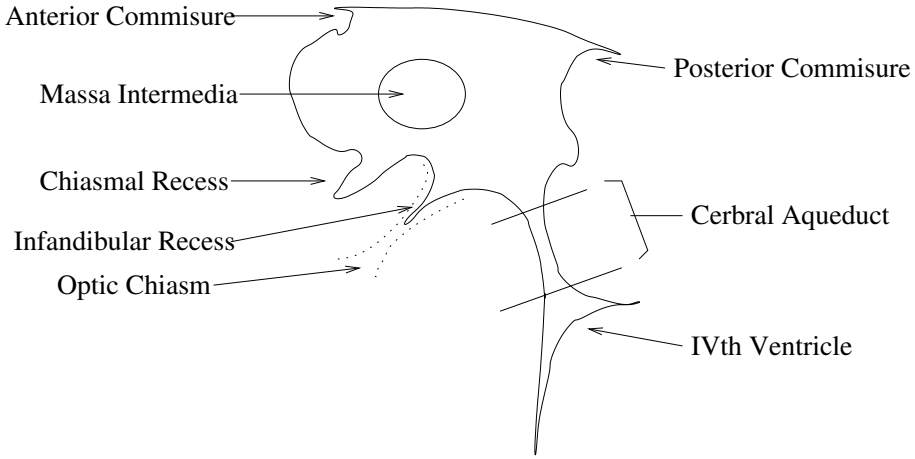


Figure 5: Sagittal view of IIIrd Ventricle and surrounding tissues

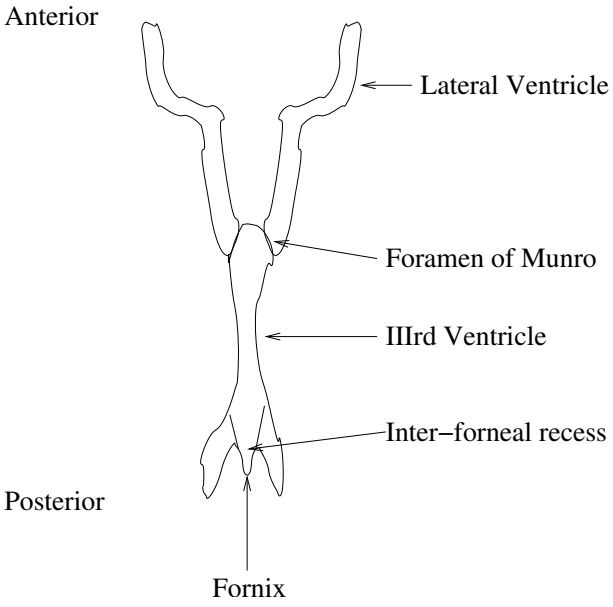


Figure 6: Transverse view of IIIrd Ventricle near the top of the structure

Rank	Structure	Vol/CSA?	CSA stable?
1	Caudate Nucleus	Vol	Yes
2	Third Ventricle	Vol	Yes
3	Midbrain and Pons	CSA	Yes
4	Thalamus	Vol	Yes
5	Hippocampus		?
6	Cortex		?
-	Putamen		No (18%)
-	Pallidum		No (33-55%)
-	Cerebellum		?
-	Corpus Callosum		No (orientation)

Table 1: Table giving structures, their importance to this project, and method of measurement

The 3rd ventricle persists for approximately 7 slices and is shown on slices 7-13. The narrowness of the ventricle suggests that it may be difficult to accurately mark it particularly in the upper slices. However, as explained earlier, the number of pixels involved are negligible in terms of volume. This orientation is the most stable.

3 Structures best viewed on coronal slices

3.1 Hippocampus

The hippocampus is theoretically another important structure to study. In chronic depression, there is over-production of cortisol, which leads to wasting of the hippocampus. However, for reasons outlined earlier, knowledge of such a process may be of less value when considered in the context of other diseases. Measurement of the hippocampus should therefore be regarded as something we may like to do for completeness rather than something which will provide a new clinical finding or insight.

It is very difficult to identify the hippocampus on transverse sections, and so coronal slices will be used. The resolution of the reconstructed coronal images is too coarse to be used for such detailed anatomical analysis.

The following structures are also best viewed in coronal section: temporal lobes, pararrhynal and entorrhynal cortex, amygdala.

Whether these are the most stable views remains to be seen.

4 Structures best viewed on sagittal sections

4.1 Corpus Callosum

The corpus callosum is best seen on midline sagittal images. These images need to be aligned precisely and accurately to a standard model, as a small error in the angle will lead to a large measurement error. This should not be a problem, as TINA has reliable registration software. If the corpus callosum is very thin, then this orientation is very unstable.

5 Conclusions

Table 1 ranks (1=most important) the six most relevet structures to this study, and which overall measurement should be taken. It is taken as read that if the volume is used, the measurement is stable. Note that the hippocampus and cortex have no entry at the moment, owing to the lack of images in the most appropriate view.

Note that for the four most important structures (which are best viewed in the orientation we have available) the CSA is stable, and for three of the four we are able to measure the volume, due to the

VoI	Marker1(NiAC)	Marker2(Laptop)	Error	Max % Error
L Caudate	1094.97	1320.55	225.58	20.6
R Caudate	784.74	687.62	97.12	14.1
L Thalamus	1187.00	1269.43	82.43	6.9
R Thalamus	1202.99	1425.68	222.69	18.5
Brainstem	7513.53	7592.24	77.71	1.0
	Marker1(NiAC)	Marker2(NiAC)		
L Caudate	873.88	900.20	26.32	3.0
R Caudate	899.94	989.94	90.00	10.0
L Thalamus	2575.84	2304.21	271.63	11.8
R Thalamus	2398.98	2844.59	445.61	18.6

Table 2: Table giving volumes of left and right caudates and thalami, and brainstem, for two patients; mark-ups performed by 2 markers, using NiAC Suns and Laptop

VoI	Marker1(NiAC)	Marker2(NiAC)	Error	Max % Error
L Caudate	1167.9	1213.09	45.3	3.8
R Caudate	1173.17	1113.83	59.34	5.3

Table 3: Results using auto method

structures being well delineated and well represented in the transverse views.

Although the volume of the brainstem (midbrain and pons) cannot be measured, the cross-sectional area along its length can be, such that either the greatest CSA can be used, or a signature based on changes in CSA along its length.

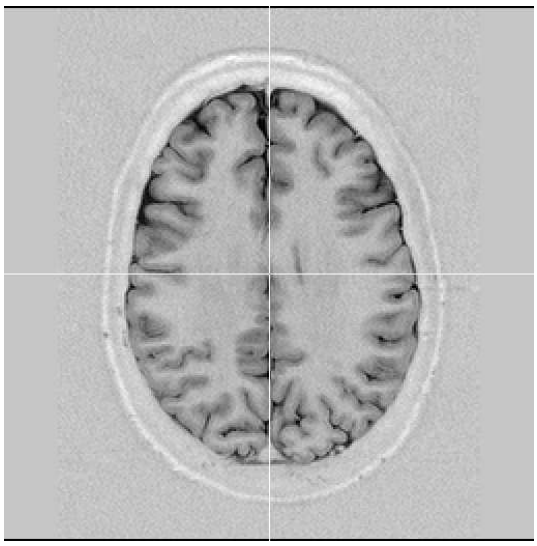
It is not possible to view the Hippocampus or cortical structures using the transverse view. but are included in the table due to their relevence.

Structures which have been discussed in this paper, but which are least relevent to this study, and are also unstable in the transverse view are given.

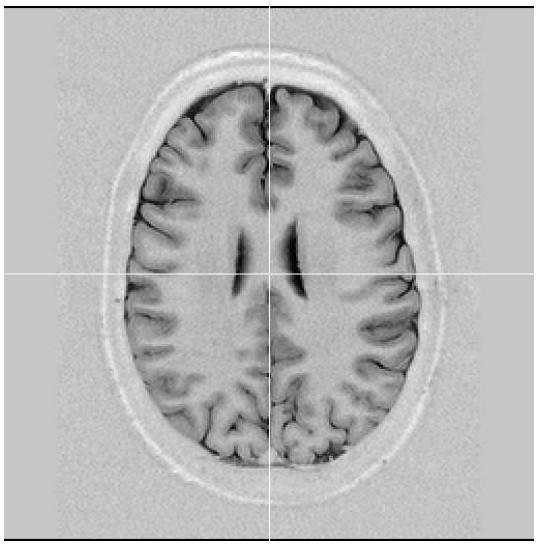
6 Results of Manual Markup

As can be seen from table 2, reproducibility between markers is poor, particularly when different platforms are used. There is no consistent trend between size of structure and error obtained.

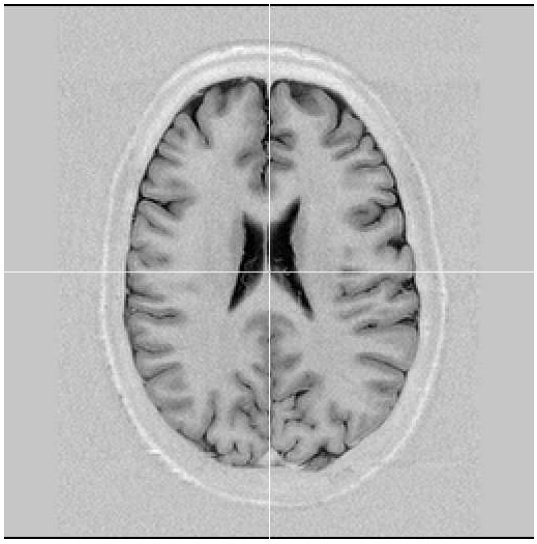
Table 3 shows volumes obtained for left and right caudates after application of the automatic isosurface process, based on the volumes produced by the markers. The automatic process gives greater volumes than both of the markers. In the case of the left caudate, the error over the volume is approximately the same as the Markers error. However, the error over the right caudate is reduced by half. The errors are still expected to be quite large, as the automatic method is based on the mark-ps produced by hand, so is still dependent on the error introduced by the markers.



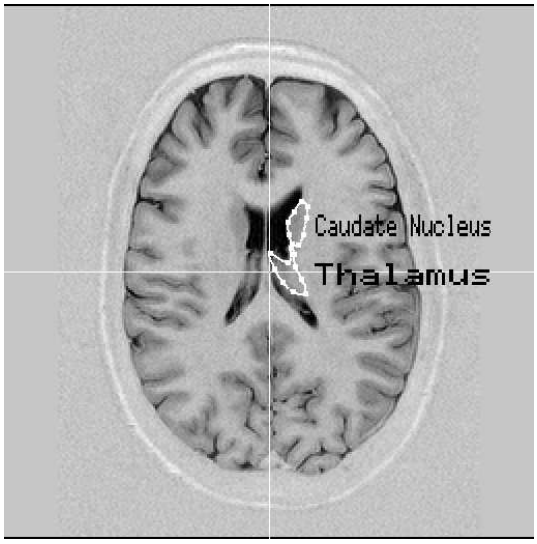
Slice 1



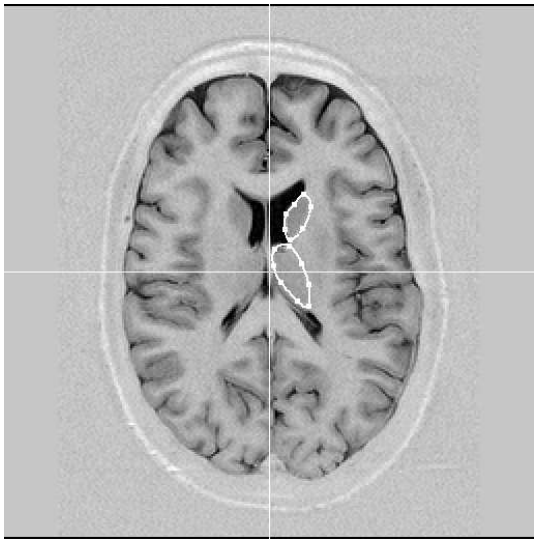
Slice 2



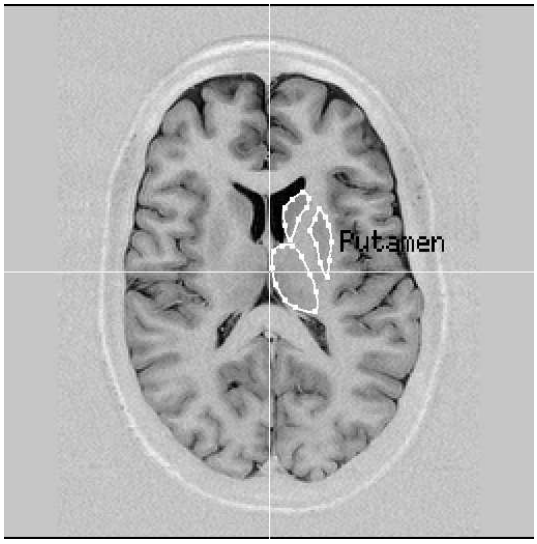
Slice 3



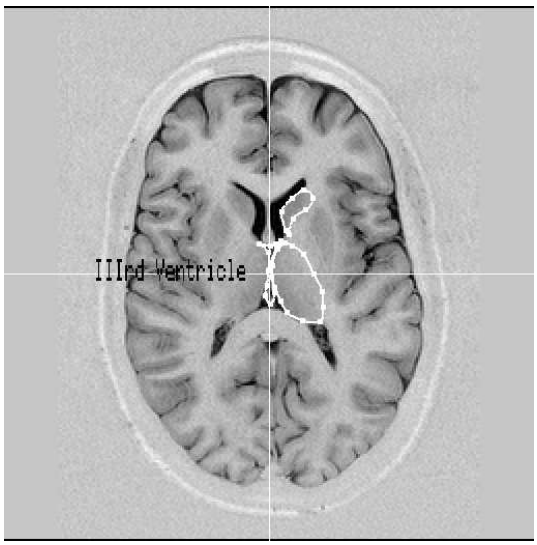
Slice 4



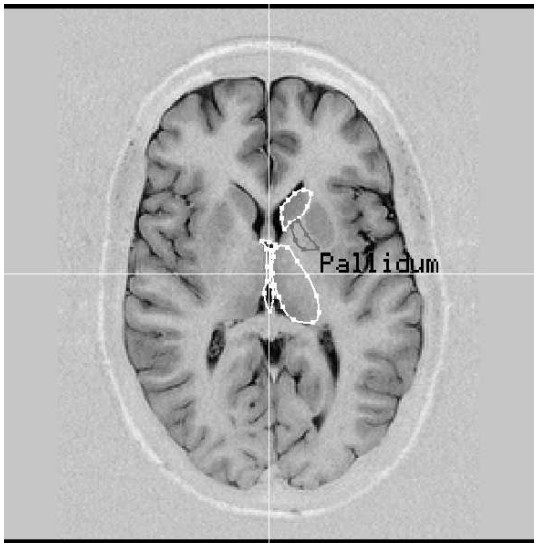
Slice 5



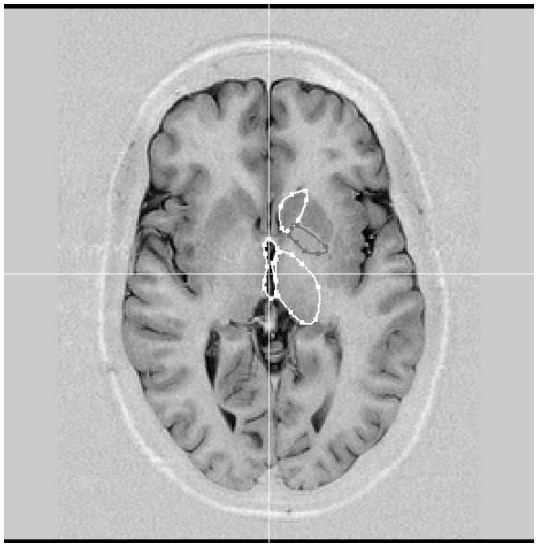
Slice 6



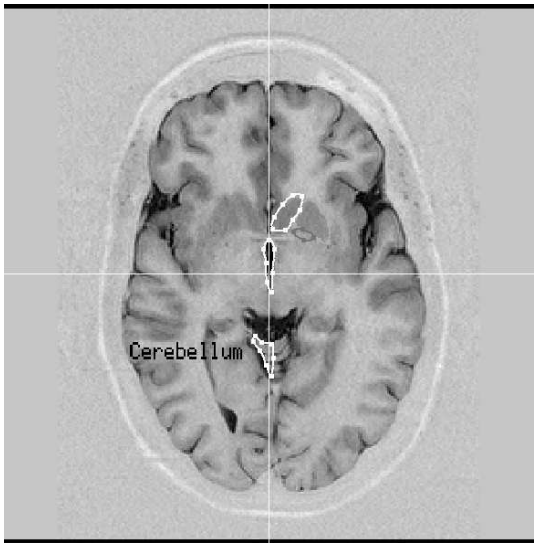
Slice 7



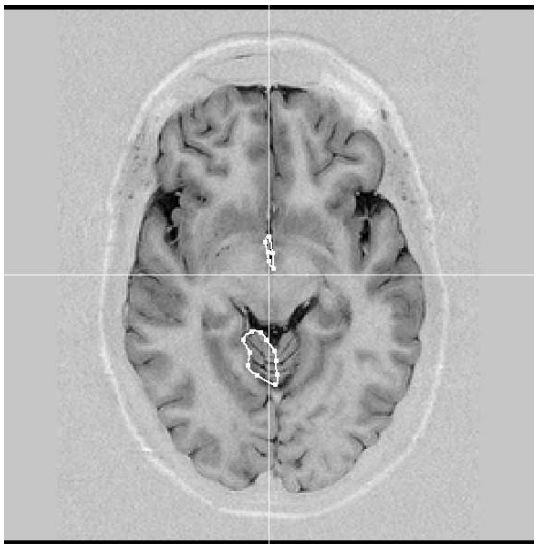
Slice 8



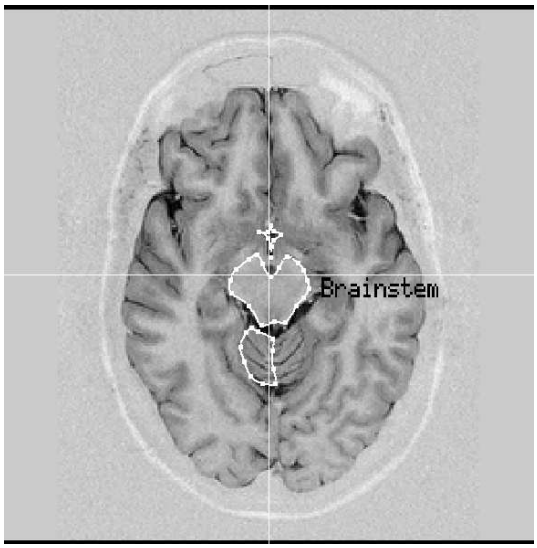
Slice 9



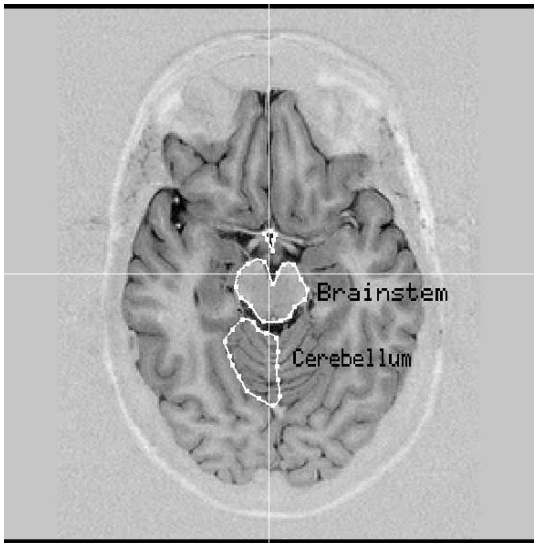
Slice 10



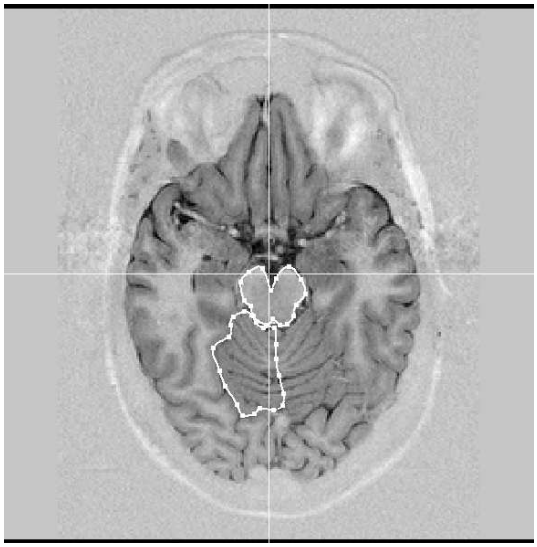
Slice 11



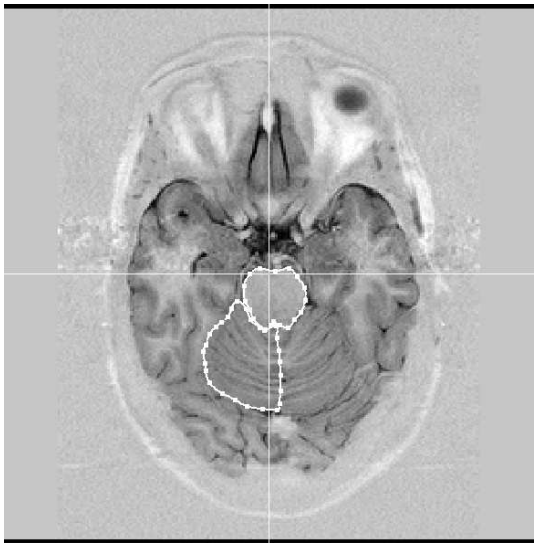
Slice 12



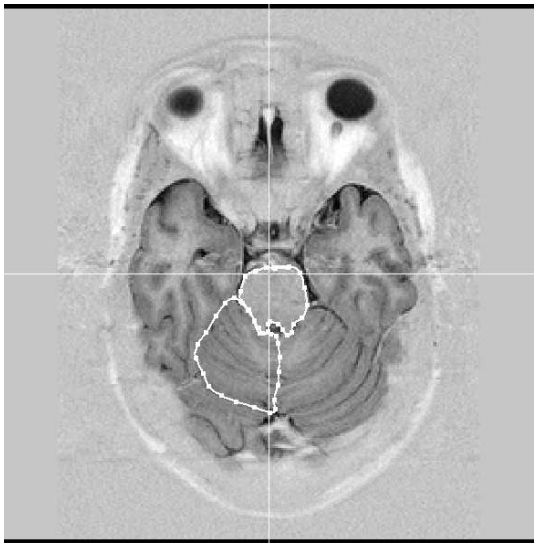
Slice 13



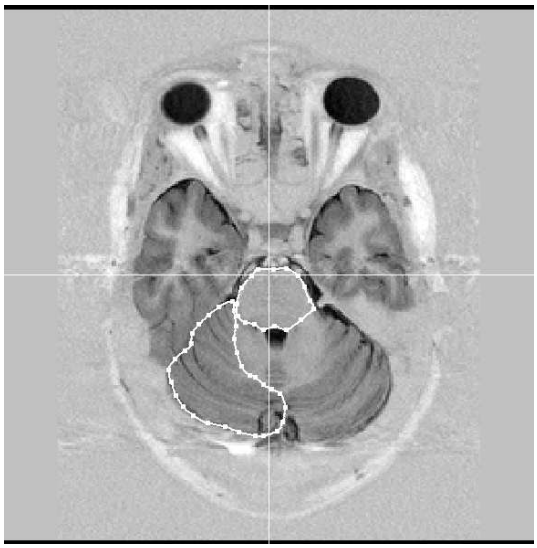
Slice 14



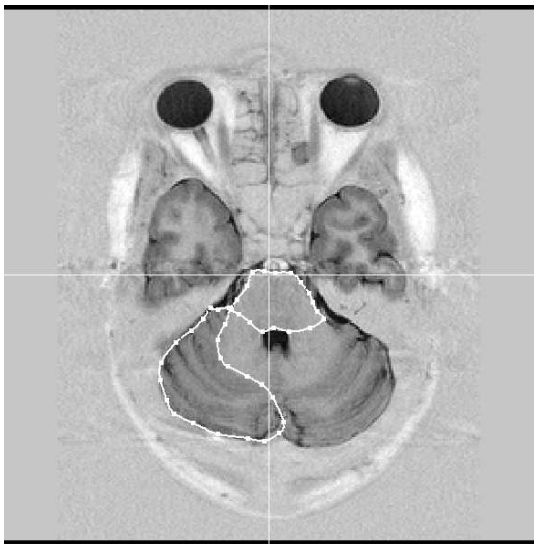
Slice 15



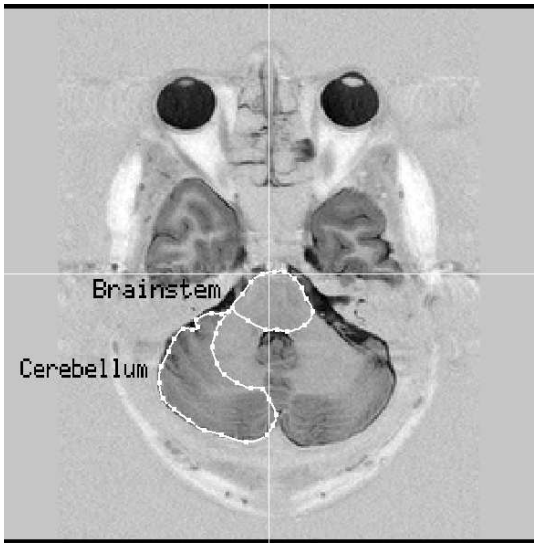
Slice 16



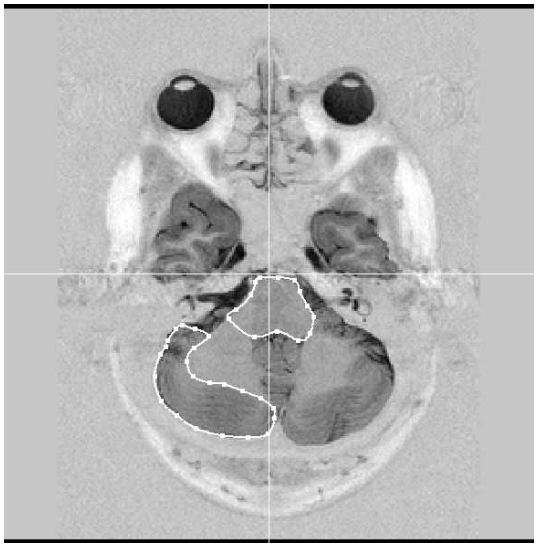
Slice 17



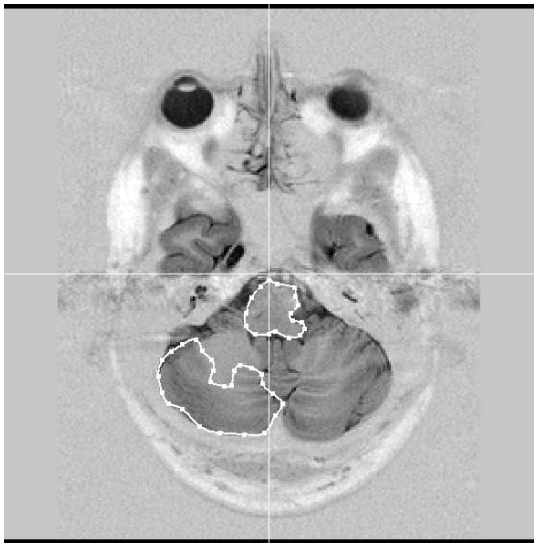
Slice 18



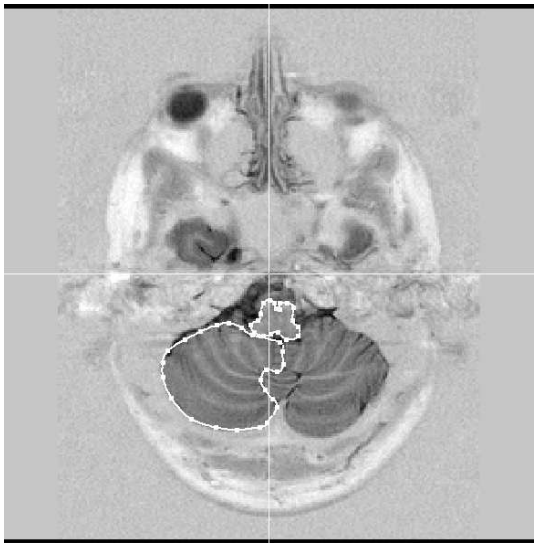
Slice 19



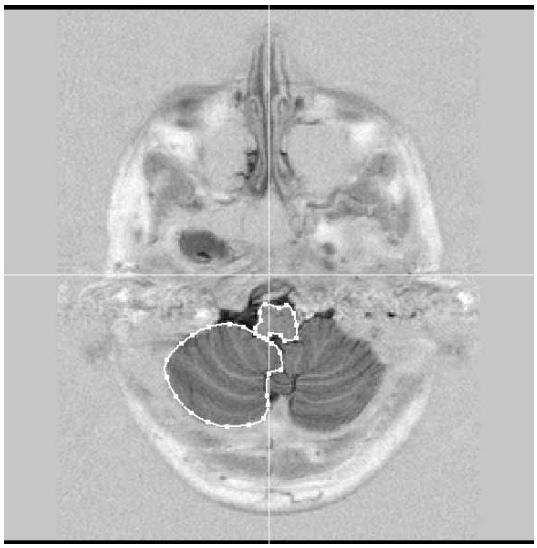
Slice 20



Slice 21



Slice 22



Slice 23