Published in final edited form as:

J Neuroimaging. 2016 ; 26(1): 28–36. doi:10.1111/jon.12313.

A New MRI Masking Technique Based on Multi-Atlas Brain Segmentation in Controls and Schizophrenia: A Rapid and Viable Alternative to Manual Masking

Elisabetta C. del Re*, Yi Gao*, Ryan Eckbo, Tracey L. Petryshen, Gabriëlla A.M. Blokland, Larry J. Seidman, Jun Konishi, Jill M. Goldstein, Robert W. McCarley, Martha E. Shenton, and Sylvain Bouix

VA Boston Healthcare System, Brockton, MA (ECdR, JK, RWM, MES); Department of Psychiatry, Harvard Medical School, Boston, MA (ECdR, YG, RE, TLP, GAMB, LJS, JK, JMG, RWM, MES, SB); and Radiology, Harvard Medical School, Boston, MA (MES)

Abstract

Brain masking of MRI images separates brain from surrounding tissue and its accuracy is important for further imaging analyses. We implemented a new brain masking technique based on multi-atlas brain segmentation (MABS) and compared MABS to masks generated using FreeSurfer (FS; version 5.3), Brain Extraction Tool (BET), and Brainwash, using manually defined masks (MM) as the gold standard. We further determined the effect of different masking techniques on cortical and subcortical volumes generated by FreeSurfer.

METHODS—Images were acquired on a 3-Tesla MR Echosped system General Electric scanner on five control and five schizophrenia subjects matched on age, sex, and IQ. Automated masks were generated from MABS, FS, BET, and Brainwash, and compared to MM using these metrics: a) volume difference from MM; b) Dice coefficients; and c) intraclass correlation coefficients.

RESULTS—Mean volume difference between MM and MABS masks was significantly less than the difference between MM and FS or BET masks. Dice coefficient between MM and MABS was significantly higher than Dice coefficients between MM and FS, BET, or Brainwash. For subcortical and left cortical regions, MABS volumes were closer to MM volumes than were BET or FS volumes. For right cortical regions, MABS volumes were closer to MM volumes than were BET volumes.

CONCLUSIONS—Brain masks generated using FreeSurfer, BET, and Brainwash are rapidly obtained, but are less accurate than manually defined masks. Masks generated using MABS, in contrast, resemble more closely the gold standard of manual masking, thereby offering a rapid and viable alternative.

Keywords

Schizophrenia; MABS; Dice coefficient; FreeSurfer; BET; Brainwash

Correspondence: Address correspondence to Sylvain Bouix, Ph.D. Brigham and Women's Hospital, Harvard Medical School, Boston, MA. sylvain@bwh.harvard.edu.

*These authors contributed equally to the work.

Introduction

Brain masking of MRI images separates brain from surrounding tissue and its accuracy is critical for the further analyses of imaging data. Manual masking (MM) is generally considered the gold standard, but it is extremely time-consuming, taking up as much as several hours per case. As increasingly larger data sets are used in research studies, the need to reduce brain masking time has become ever more important. Several automated masking techniques are available to alleviate this problem. Some of these techniques are based on the detection of boundaries of the target image data without the use of external anatomical priors,¹⁻⁵ and among these, the Brain Extraction Tool (BET) masking technique⁶ is a widely popular methodology based on the surface reconstruction of each individual target image. The FreeSurfer technique⁷ is also commonly used. In contrast to the BET technique, it relies on the use of a single atlas that is adapted to all target images. More recently, with the advent of more powerful computing techniques, methodologies based on the adoption of multiple atlases have become available to address this issue.⁸⁻¹¹ In large part, the atlas selection dictates the outcome of the brain masking method in three ways. First, intrinsic errors of the atlases may be transferred to the target image. Second, atlases might be anatomically unsuitable for certain target images especially in clinical settings, where neuroanatomy may be affected by pathological processes. Lastly, individual differences between subjects might be averaged out.

Here, we present a recently implemented brain masking technique, which is based on the idea of multi-atlas brain segmentation (MABS), in which an initial number of MM are used to generate masks adapted to each new target image. This technique circumvents the basic issue of atlas selection by assigning weights to the atlases according to their resemblance to the target image. This latter feature of the MABS technique allows flexibility. Given a set of heterogeneous atlases, the atlases in the set most closely resembling the target will be given more weight than atlases with less resemblance to the target. In addition, together with the inclusion of both pathological and control images in the training set, the MABS technique reduces the risk of averaging out individual differences, important in cross-sectional or clinical studies.

In order to determine the validity of the MABS methodology, masks generated using this technique were compared to masks generated via other methodologies. MM was used as the gold standard. Statistically, Dice coefficients and intra-class correlation coefficients (ICCs) were employed to determine overlap and absolute agreement between MM masks and FreeSurfer (FS; version 5.3) masks,^{12,13} masks generated using the BET,⁶ masks generated through Brainwash (<http://www.nitrc.org/projects/art/>), and MABS masks. To determine whether brain pathology might affect outcome of different masking techniques, age-, gender-, and intelligence quotient (IQ)-matched control and schizophrenia subjects were included in the study design. Further, in order to determine the effect of different masking techniques on FreeSurfer 5.3 output, the volumes of 10 cortical and 4 subcortical regions, in addition to total gray matter volume, were extracted using each masking type and compared.

Materials and Methods

Subjects

All together 19 subjects participated in the study. MRI images from 10 subjects, five first episode schizophrenics (FESZs; two females) and five healthy controls (HCs; two females), were masked using each of five methods. The other nine subjects, including five FESZs and four HCs, were apportioned to the training set. All subjects were recruited as part of the Boston Center for Intervention Development and Applied Research (www.bostoncidar.org).^{14,15} HCs were recruited from the general community via Internet advertisements. FESZs were recruited from outpatient clinics affiliated with Harvard Medical School, or through referrals from clinicians. The study was approved by the local IRB committees at Harvard Medical School, Beth Israel Deaconess Medical Center, Massachusetts General Hospital, Brigham and Women's Hospital, and the Veteran Affairs Boston Healthcare System (Brockton campus). All study participants gave written informed consent and received payment for participation. All demographic and clinical data are summarized in Table 1.

MRI Image Acquisition

MR images were acquired on a 3-Tesla whole-body MRI EchoSpeed system General Electric scanner (GE Medical Systems, Milwaukee, USA) at Brigham and Women's Hospital. Each MRI scan was visually inspected for movement artifacts. An eight-channel coil was used to perform parallel imaging using array spatial sensitivity encoding techniques (ASSETs GE) with a SENSE acceleration factor of 2. The structural MRI acquisition protocol included the following pulse sequence and parameters: contiguous inversion-prepared spoiled gradient-recalled acquisition (T1W) and IP-SPGR (T2W) (fastSPGR), TR = 7.4 ms, TE = 3 ms, TI = 600, 10° flip angle, 25.6 cm² field of view, and matrix = 256 × 256. The voxel dimensions were 1 × 1 × 1 mm. Images were realigned to the anterior commissure–posterior commissure (AC-PC) line and to the sagittal sulcus to correct for head tilt.

MRI Image Processing

Manual Masking—Masks were first generated from T2-weighted images from 10 subjects using the BET software,⁶ part of the FSL software library v.5.0.^{16–18} The resulting masks were then carefully manually edited using 3-dimensional slicer, version 4.0.¹⁹ Holes in the mask were removed. In addition, extensions of the boundary beyond the edge of the brain gray matter were corrected. The brainstem was included only after the emergence of the cerebellum as seen in axial view. Cerebrospinal fluid inside the brain and in the subarachnoid space was included in the final masks. After an average processing time of ~8 hours, the final MM masks were used as the gold standard against which all masks generated using other methods were compared.

Multi-Atlas Brain Segmentation—A set of nine anatomical T2W images comprising a training set (different from the 10 subjects tested), each with an associated expert manual tracing of the brain mask, were nonlinearly registered to the target anatomical image to be segmented using the symmetric diffeomorphic image registration of the Advanced Normalization Tools (ANTs) software suite.²⁰ All targets used in the present series of

validations were T2W images and only one type of scan was used in the training set. The resulting transformations were used to transform the corresponding brain masks to the target image. The transformed masks were then combined using a weighted average and thresholded at 50% to create an estimate of the brain mask of the target image (Fig 1). The weights given to each atlas during averaging were determined by the quality of the registration of each atlas image to the target. Specifically, we computed the mutual information^{41,42} (MI) between each atlas image and the target image and weighted the contribution of each atlas mask to the target mask according to w_i , a measure of image similarity. The weight w_i of atlas A_i was calculated as:

$$W_i = \frac{1}{C} e^{-\alpha \frac{(d_i - d_{\min})}{(d_{\max} - d_{\min})}},$$

where $C = \sum_1^n W_i$, where $A_1, \dots, A_n$ are the atlases in target space, d_i is the MI between atlas A_i and the target T, and $d_{\min}$ and $d_{\max}$ are the minimum and maximum MI over all atlases. The parameter α is chosen by the user and influences how quickly the weights decay when the MI is not optimal. Figure 2 graphically depicts how quickly atlases weights decay in accordance with atlases MI values (Fig 2). After testing a range of values for α , we set $\alpha = 1.6$ in our experiments. Figure 3 illustrates Dice coefficient values between MABS obtained at different alpha values and manual masks (Fig 3). A repeated measures ANOVA analysis, where group was the between-subjects factor and the seven Dice coefficients calculated according to the seven α values were the within-subjects factor, demonstrated no significant difference for group [$F(1,8) = 1.23$; $p = .3$]. Instead, a main effect of Dice coefficients was found [$F(1,8) = 26.5$; $p = .001$]. Dedicated post hoc analyses demonstrated that a significantly higher Dice coefficient was obtained between MABS constructed with an $\alpha = 1.6$ and manual masks, than between MABS generated with each of the other alpha values and manual masks, including the $\alpha = 0$ value. As shown in the figure, Dice coefficients significantly decrease from α value of 1.6 up to α value of 40. The lowest Dice coefficient was that obtained between MABS generated with $\alpha = 0$ and manual masks.

The exponential weighing procedure included in MABS creation is particularly useful in the presence of atlases that poorly match the target.

Standard Brain Segmentation Tools—FreeSurfer v5.3³⁹ (FS), the BET2,⁶ and Brainwash (<http://www.nitrc.org/projects/art/>) were also used to generate automated masks to test MABS against the current state of the art. All these programs were used with their default parameters without any type of editing. The Bet2 option was utilized with the $-m$ input to generate binary masks. A comparison of masking outcomes is presented in Figure 4.

FreeSurfer Image Analysis—FreeSurfer v5.3 was utilized to obtain region of interest (ROI) volumes from images generated using MM masks, BET masks, FreeSurfer masks, and MABS masks. Brainwash masks were not included in these latter experiments because they were deemed too inaccurate for further FreeSurfer segmentation comparisons based on visual inspection (eg, masks sometimes missed whole regions) and the mean overlap with MM (Dice coefficient; Fig 5). Total gray matter volume, right and left hippocampus, and

right and left amygdala were extracted by FreeSurfer. In addition, the following cortical ROIs (right and left) were extracted by FreeSurfer using the Desikan–Killiany atlas:⁴⁰ caudal and rostral anterior cingulate, caudal and rostral middle frontal gyrus, superior temporal gyrus, Heschl's gyrus, entorhinal cortex, latero-orbital cortex, medial orbital cortex, and insula.

Metrics Used to Compare Accuracy of Masks—Four metrics were used to compare the accuracy of each of the automated masking tools (MABS, FreeSurfer, BET, and Brainwash) against the gold standard MM: a) mask volume difference from MM; b) number of masks in each set of 10 masks created with one of the four masking methods, where volume differed more than 3% from the MM; c) Dice coefficient (an overlap measure) applied to MM; and d) ICCs calculated between the volumes of each cortical and subcortical region obtained from FreeSurfer analysis using each mask type and from FreeSurfer analysis using MM masks. The Dice coefficient d was calculated as $[d = 2|X \cap Y| / (|X| + |Y|)]$, where X and Y are two voxel sets, one from a mask generated by one of the automated techniques and the other from the MM mask.

Statistical Analyses

Subject sociodemographic and clinical data were compared using analysis of variance (ANOVA). ANOVA was also used for analyzing volume differences and Dice coefficient differences among masks with Group (ie, HC and FESZ) as a between-subjects factor and Mask-type as a within-subjects factor. Means were compared using protected paired T-tests. Absolute volume differences of cortical and subcortical regions obtained using different masking techniques were compared using MANOVA with Group as a between-subjects factor, and Mask type and cortical or subcortical ROIs as within-subjects factors. Fisher's exact test was used for categorical data. p -Values reported for post hoc analyses were Bonferroni corrected.

Results

Subjects

No significant differences were found between HC and FESZ groups for the following variables: age ($F[1,8] = .03$, $p = .87$), premorbid IQ ($F[1,8] = .016$, $p = .9$), current IQ ($F[1,7] = 2.67$, $p = .15$), parental socioeconomic status ($F[1,8] = 1.4$, $p = .273$), or proportion of males and females ($\chi^2 [1] = .4$; $p = 1.0$). These data are summarized in Table 1.

Masks Volume Differences

In order to determine whether the volume of masks generated by different methodologies differed significantly from the volume of manually traced masks, an ANOVA was performed with Group (HC and FESZ) as a between-subjects factor and Volume Difference of each mask type from MM mask as a within-subjects factor. ANOVA showed no significant main effect of Group [$F(1,8) = .09$; $p = .772$] but there was a significant main effect of Mask type [$F(2,16) = 34.5$; $p < .001$]. Post hoc analysis showed that the mean volume difference between MABS and MM masks ($19,389.1 \text{ mm}^3$) was significantly smaller than that between

FS masks and MM masks (142,558.3 mm³) and between BET masks and MM masks (99,005.2 mm³) ($p < .01$ for both comparisons).

Using a cutoff of 3% volume difference from MM, it was also determined that 10/10 FS masks, 9/10 BET masks, and none of MABS masks differed by >3% from MM volumes (Fisher test; $p < .001$ for FS vs. MABS and for BET vs. MABS). A visual comparison of different masks is depicted in Figure 4.

Dice Coefficients

Dice coefficients were calculated for FS, BET, Brainwash, and MABS masks, each relative to the MM masks. Using ANOVA, no significant main effect of Group was found [$F(1,8) = 3.8$; $p = .09$], but there was a significant main effect for Mask type [$F(3,24) = 8.9$; $p < .001$]. Post hoc analysis showed that the MABS-MM Dice coefficient ($d = .997 \pm .002$) was significantly higher than the FS-MM coefficient ($d = .93 \pm .02$; $p = .001$), the BET-MM coefficient ($d = .97 \pm .006$; $p = .04$), and the Brainwash-MM coefficient ($d = .85 \pm .13$; $p = .04$). Dice coefficients are depicted in Figure 5.

FreeSurfer Volume Differences Based on Different Masking Methods

In order to determine whether the masking technique affects FreeSurfer output, FreeSurfer v5.3 reconstruction was run separately for each mask type (except BrainWash, for reasons outlined above). As noted previously, 10 cortical, 4 subcortical, and total gray matter volumes were extracted. Absolute volume differences were calculated by subtracting the MABS, FS, and BET volumes for each cortical and subcortical region from MM volumes. Percent volume differences were also calculated to determine their magnitude relative to the total size of the structure; however, the results were consistent, and therefore, only the analyses using absolute volume differences are detailed below. All results are summarized in Table 2.

MANOVA of absolute volume differences of left hemisphere cortical ROI did not detect a main effect of Group [$F(1,8) = .012$; $p = .92$], while a main effect of Mask type was significant [$F(2,16) = 26.7$; $p = .002$]. Using post hoc analysis, the mean volume difference between MM and MABS was significantly lower than the difference between MM and FS ($p = .001$) and between MM and BET ($p = .042$). A significant interaction between Mask type difference and ROI was also present [$F(18,144) = 16.9$; $p < .001$] and was followed up by post hoc analyses. BET volume differences were greater than MABS differences in six cortical regions, and FS volume differences were greater than MABS differences in two cortical regions. As in the left hemisphere, no main effect of Group was found for right hemisphere for the cortical ROIs [$F(1,8) = .068$; $p = .8$], while there was a main effect of Mask [$F(2,16) = 20.8$; $p < .001$]. Post hoc comparison showed that the mean volume difference between MM and MABS in the right hemisphere did not differ significantly from the mean volume difference between MM and FS ($p = .25$). Nonetheless, both MM-MABS mean volume differences and MM-FS mean volume differences were significantly lower than the mean volume difference between BET and MM masks ($p = .004$ and $p = .006$, respectively).

For subcortical regions, no significant main effect of Group was present [$F(1,8) = .52$; $p = .49$]; however, a significant main effect of Mask type was found [$F(2,16) = 11.1$; $p < .001$]. Post hoc analysis showed that the mean volume difference between MM and MABS was less than the mean volume difference calculated between FS and MM ($p = .031$). In addition, the mean volume difference between MM and MABS was also less than the mean volume difference between MM and BET ($p = .005$). Among the findings, notice that total gray matter volume difference between MABS and MM was significantly less than total gray matter volume difference calculated between MM and BET ($p = .005$) and between FS and MM ($p = .03$).

All data and p values are summarized in Table 2.

Intraclass Correlation Coefficients

Two-way mixed ICCs with measures of absolute agreement were calculated between the volume of each cortical and subcortical region obtained from FreeSurfer analysis using different mask types and from FreeSurfer analysis using MM masks. As summarized in Table 3, the ICC calculated for MM and MABS across cortical and subcortical regions and total gray volume were consistently higher than the ICC calculated for MM and either FS or BET.

Discussion

Automated masking methods have recently become popular and are often preferred to MM, which requires substantial time and man power, and is especially taxing when large datasets are processed. Nonetheless, MM remains the gold standard for quality over automated masking techniques. Several automated masking techniques are freely available and widely used. These include the FreeSurfer masking technique,⁷ the BET masking tool,⁶ the Brainwash masking tool (<http://www.nitrc.org/projects/art/>), the Exbrain method,²¹ the SFU method,²² the multi-atlas skull-stripping technique, MASS,²³ the BREED masking method,²⁴ and the MBRASE method,²⁵ among others.^{21,26–29}

We have presented research data demonstrating that the MABS masking technique, recently developed by our group yields superior results compared to other readily available automated masking techniques. We compared MABS to FreeSurfer v5.3, BET, and BrainWash masking methods, as representative examples of a single atlas masking technique, a brain surface reconstruction technique, and a multi-atlas technique, respectively. We adopted manual masking as the gold standard, bypassing the issues of comparing different masking techniques without the use of an unbiased standard.³⁰ Four different metrics, including Dice overlap of automated masks with manual masks, ICCs between automated masks and manual masks, and volume differences between automated masks and manual masks, all indicate that the MABS approach better resembles manual masks compared to other available automated masking techniques. The use of a mixed population sample, composed of healthy controls and age-, IQ-, gender-matched schizophrenia patients, demonstrate that MABS are suited to healthy and clinical populations. Functionally, we show that FreeSurfer v5.3 reconstruction performed separately for each mask type is

affected by the initial masking choice and that MABS faithfully recreates FreeSurfer volume data obtained by using manual masks.

MABS masking is implemented by: 1) creating a training set of 8 to 20 manual masks each associated with the corresponding brain; 2) registration of each of the training set of masks to a new brain; 3) weighted averaging the registered training set of masks; and 4) thresholding the average to produce a binary mask (Fig 1). The automated MABS technique thereby substantially reduces the effort required to create accurate brain masks and, as we have demonstrated in this study, it better resembles that of MM compared to other automated masking techniques (eg, FS, BET, and BrainWash; see Figs 4 and 5).

Atlas-based segmentation is not a new concept. Such techniques, in fact, were originally developed on the basis of one single atlas.^{31–34} These techniques relied on the similarity and nonlinear registration of the target image and the atlas. However, because of the inherent diversity of individual anatomical specimens, the results were not always optimal. A more recent, further development of atlas-based techniques, is represented by multi-atlas segmentation.^{5,35–38} This methodology takes advantage of multiple atlases that are nonlinearly registered to the target image. Use of multiple atlases allows for more of the individual variability of the individual targets to be predicted than is possible with the use of one single atlas.⁵ All in all, two steps seem to be critical to a good outcome of the multi-atlas methodology and these are atlas selection and registration to the target. These two steps should successfully maintain individual images' differences, while achieving brain masking quality comparable to that achieved with manual masking. The method should also take into account the risk implicit in multi-atlas brain masking methodology of averaging out individual differences in brain morphology. This latter risk is an especially detrimental drawback, when the aim is to detect small pathological abnormalities or differences in cross-sectional studies that utilize controls and clinical populations to detect control case differences. Among multi-atlas techniques, these aspects have been recognized and approached with different tools.⁸ For example, the ROBEX approach combines the multi-atlas approach with an algorithm that uses information from the atlases to grow a shape of the target brain; but also,²¹ while the MASS approach²³ has made use of a registration algorithm that adaptively aligns different regions of two images based on their similarity and on the reliability of the matching between images. In contrast, the MABS methodology handles the issue of atlas selection and potential inaccuracies in nonlinear registration by taking a two-pronged approach: 1) the initial training set includes both control and pathological specimens, and 2) atlases are exponentially weighted according to their resemblance to the target image, thus reducing the impact of atlases that poorly match the target. One further advantage of the MABS methodology is the ability to add new subjects to the study after the initial training set has been constructed, and again allowing individual differences to be preserved and not averaged out.

Inclusion of pathological scans in the initial training set in conjunction with exponential weighting allows for use of MABS on a wider range of brain images than previous multi-atlas masking methodologies. Although the danger in including pathological images lays in having scans that are inappropriate as atlas choices, the weighting would correct for the matching to the target while allowing “the training set” to recognize a wider range of brain

abnormalities. The present manuscript includes data from first episode schizophrenia and HC subjects, nonetheless we have used MABS masking technique to mask brains in chronic schizophrenia populations and in subjects at genetic high risk of developing schizophrenia with optimal results. The latter outcomes indicate that MABS technique is applicable to a range of brain pathologies.

Our study is limited by the fact that while all of the masking methods were compared to MM, only MABS was trained using masks that follow the same anatomical conventions as the gold standard manual masks. This is not the case for FreeSurfer and BET masks, which were produced using conventions defined by their originating laboratories. However, we feel confident in our data because, as is evident in Figure 3, not only was there a large difference in the mean Dice coefficients between manual masks and either FS, BET, or Brainwash, but the standard deviations for the latter three methods were also large, indicating a lack of robustness of these masking techniques. That is notwithstanding initial input, MABS remain very reproducible, something that instead is missing in the other automated techniques shown. Indeed, we believe that irrespective of the conventions from which those masks were produced, the results would have remained highly variable and less accurate than the MABS technique.

A further drawback of the MABS technique might be the need of AC-PC realignment of the images. Nonetheless, we consider this a minor weakness considering the high-quality results obtained by using MABS methodology. The MABS technique represents a compromise that combines both manual masks and automation for the production of high-quality masks, where the quality of a small number of manual masks to be included in the training set determines the quality of the resulting final MABS.

In summary, although many automated masking techniques somewhat compromise quality for speed, we demonstrated that the MABS methodology achieves high-quality results, comparable to those obtained with manual masks, while greatly reducing processing time. More specifically, our results using Dice coefficients demonstrate that MABS better corresponds to manual masks than do FS, BET, or Brainwash masks, and further suggests that these latter techniques produce results with higher variability.

Importantly, volumetric data extracted by FreeSurfer analysis of brains masked with MABS, FS, or BET also demonstrate that MABS masking produces results that are closer to those obtained after manual masks than the other two automated techniques. All in all, our data demonstrate that the MABS technique is advantageous over other automated masking techniques both in terms of time and in terms of quality/accuracy. Scripts and procedures are available from Sylvain Bouix, Ph.D., and the Psychiatry Neuroimaging Laboratory, on request.

References

1. Zeng X, Staib LH, Schultz RT, et al. Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation. *IEEE Trans Med Imaging*. 1999; 18:927–37. [PubMed: 10628952]

2. Shattuck DW, Sandor-Leahy SR, Schaper KA, et al. Magnetic resonance image tissue classification using a partial volume model. *Neuroimage*. 2001; 13:856–76. [PubMed: 11304082]
3. Huh S, Ketter TA, Sohn KH, et al. Automated cerebrum segmentation from three-dimensional sagittal brain MR images. *Comput Biol Med*. 2002; 32:311–28. [PubMed: 12102751]
4. Huang A, Abugharbieh R, Tam R, et al. MRI brain extraction with combined expectation maximization and geodesic active contours. *IEEE International Symposium on Signal Processing and Information Technology*. 2006:107–11.
5. Wang Y, Nie J, Yap P-T, et al. Robust deformable-surface-based skull-stripping for large-scale studies. *Med Image Comput Comput Assist Interv*. 2011; 14:635–42. [PubMed: 22003753]
6. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapping*. 2002; 17:143–55.
7. Ségonne F, Dale AM, Busa E, et al. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*. 2004; 22:1060–75. [PubMed: 15219578]
8. Carass A, Cuzzocreo J, Wheeler MB, et al. Simple paradigm for extra-cerebral tissue removal: algorithm and analysis. *Neuroimage*. 2011; 56:1982–92. [PubMed: 21458576]
9. Eskildsen SF, Coupé P, Fonov V, et al. BEaST: brain extraction based on nonlocal segmentation technique. *Neuroimage*. 2012; 59:2362–73. [PubMed: 21945694]
10. Iglesias JE, Liu CY, Thompson PM, et al. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Trans Med Imaging*. 2011; 30:1617–34. [PubMed: 21880566]
11. Leung KK, Barnes J, Modat M, et al. Brain MAPS: an automated, accurate and robust brain extraction technique using a template library. *Neuroimage*. 2011; 55:1091–108. [PubMed: 21195780]
12. Fischl B, van der Kouwe A, Destrieux C, et al. Automatically parcellating the human cerebral cortex. *Cereb Cortex*. 2004; 14:11–22. [PubMed: 14654453]
13. Fischl B. FreeSurfer. *Neuroimage*. 2012; 62:774–81. [PubMed: 22248573]
14. Del Re EC, Bergen SE, Meshulam-Gately RI, et al. Analysis of schizophrenia-related genes and electrophysiological measures reveals ZNF804A association with amplitude of P300b elicited by novel sounds. *Transl Psychiatry*. 2014; 4:e346. [PubMed: 24424392]
15. del Re EC, Spencer KM, Oribe N, et al. Clinical high risk and first episode schizophrenia: auditory event-related potentials. *Psychiatry Res*. 2015; 231:126–33. [PubMed: 25557063]
16. Jenkinson M, Beckmann CF, Behrens TE, et al. FSL. *Neuroimage*. 2012; 62:782–90. [PubMed: 21979382]
17. Woolrich MW, Jbabdi S, Patenaude B, et al. Bayesian analysis of neuroimaging data in FSL. *Neuroimage*. 2009; 45(suppl 1):S173–86. [PubMed: 19059349]
18. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 2004; 23(suppl 1):S208–19. [PubMed: 15501092]
19. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*. 2012; 30:1323–41. [PubMed: 22770690]
20. Avants BB, Epstein CL, Grossman M, et al. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008; 12:26–41. [PubMed: 17659998]
21. Wang Y, Nie J, Yap PT, et al. Robust deformable-surface-based skull-stripping for large-scale studies. *Med Image Comput Comput Assist Interv*. 2011; 14:635–42. [PubMed: 22003753]
22. Atkins MS, Mackiewicz BT. Fully automatic segmentation of the brain in MRI. *IEEE Trans Med Imaging*. 1998; 17:98–107. [PubMed: 9617911]
23. Doshi J, Erus G, Ou Y, et al. Multi-atlas skull-stripping. *Acad Radiol*. 2013; 20:1566–76. [PubMed: 24200484]
24. Miura N, Taneda A, Shida K, et al. Automatic brain tissue extraction method using erosion-dilation treatment (BREED) from three-dimensional magnetic resonance imaging T1-weighted data. *J Comput Assist Tomogr*. 2002; 26:927–32. [PubMed: 12488737]
25. Stokking R, Vincken KL, Viergever MA. Automatic morphology-based brain segmentation (MBRASE) from MRI-T1 data. *Neuroimage*. 2000; 12:726–38. [PubMed: 11112404]

26. Bomans M, Hohne KH, Tiede U, et al. 3-D segmentation of MR images of the head for 3-D display. *IEEE Trans Med Imaging*. 1990; 9:177–83. [PubMed: 18222762]
27. Höhne KH, Hanson WA. Interactive 3D segmentation of MRI and CT volumes using morphological operations. *J Comput Assist Tomogr*. 1992; 16:285–94. [PubMed: 1545027]
28. Shan ZY, Yue GH, Liu JZ. Automated histogram-based brain segmentation in T1-weighted three-dimensional magnetic resonance head images. *Neuroimage*. 2002; 17:1587–98. [PubMed: 12414297]
29. Zhuang AH, Valentino DJ, Toga AW. Skull-stripping magnetic resonance brain images using a model-based level set. *Neuroimage*. 2006; 32:79–92. [PubMed: 16697666]
30. Bouix S, Martin-Fernandez M, Ungar L, et al. On evaluating brain tissue classifiers without a ground truth. *Neuroimage*. 2007; 36:1207–24. [PubMed: 17532646]
31. Bajcsy R, Lieberman R, Reivich M. A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. *J Comput Assist Tomogr*. 1983; 7:618–25. [PubMed: 6602820]
32. Gee JC, Reivich M, Bajcsy R. Elastically deforming 3D atlas to match anatomical brain images. *J Comput Assist Tomogr*. 1993; 17:225–36. [PubMed: 8454749]
33. Collins DL, Neelin P, Peters TM, et al. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr*. 1994; 18:192–205. [PubMed: 8126267]
34. Van Leemput K, Maes F, Vandermeulen D, et al. Automated model-based tissue classification of MR images of the brain. *IEEE Trans Med Imaging*. 1999; 18:897–908. [PubMed: 10628949]
35. Rohlfing T, Russakoff DB, Maurer CR Jr. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Trans Med Imaging*. 2004; 23:983–94. [PubMed: 15338732]
36. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging*. 2004; 23:903–21. [PubMed: 15250643]
37. Heckemann RA, Hajnal JV, Aljabar P, et al. Multiclassifier fusion in human brain MR segmentation: modelling convergence. *Med Image Comput Comput Assist Interv*. 2006; 9:815–22. [PubMed: 17354848]
38. Asman AJ, Landman BA. Non-local statistical label fusion for multi-atlas segmentation. *Med Image Anal*. 2013; 17:194–208. [PubMed: 23265798]
39. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33:341–55. [PubMed: 11832223]
40. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006; 31:968–80. [PubMed: 16530430]
41. Mattes D, Haynor DR, Vesselle H, et al. Nonrigid multimodality image registration. *Proceedings of SPIE 4322, Medical Imaging*. 2001:1609–20.
42. Mattes D, Haynor DR, Vesselle H, et al. PET-CT image registration in the chest using free-form deformations. *IEEE Trans Med Imaging*. 2003; 22:120–8. [PubMed: 12703765]
43. Stoll, AL. *The Psychopharmacology Reference Card*, 1989–2009. McLean Hospital; Belmont, MA: 2009.
44. Woods SW. Chlorpromazine equivalent doses for the newer atypical antipsychotics. *Journal of Clinical Psychiatry*. 2003; 64:663–7. [PubMed: 12823080]

Fig 1.
MABS implementation steps. Thickness of arrows indicates weighing of atlases before averaging step.

Fig 2.

The weights of each atlas decay according to the atlases MI values and speed of weights decay when the MI is not optimal, is modulated by the parameter alpha.

Fig 3.

Dice coefficients values obtained between MABS and manual masks where MABS were created by employing seven different alpha values, with increasing values from 0 up to 40.

Fig 4.

Left panel. Representative T2-weighted image scan of a control subject (A). The same scan was masked with manual mask (B); MABS (C); BET (D); FreeSurfer v5.3 (E); and Brainwash (F). Red circles indicate obvious mask defects. Right panel. Masks are shown without the skull and without nonbrain matter. Arrows indicate where masks are erroneous in defining the brain. The darker margin in (D) indicates that the BET masks include nonbrain tissue, while both the FS masks (E) and Brainwash masks (F) are missing entire parts of the brain.

Fig 5.

Dice coefficients for each masking technique relative to MM. The horizontal line in the boxes indicates the median value, while the cross indicates the mean value. Whiskers indicate the 5%–95% confidence interval. MABS = Multi-atlas brain segmentation; MM = manual masking; FS = FreeSurfer v5.3; BET = Brain Extraction Tool. * $p < .05$; ** $p < .01$.

Table 1

Sociodemographic and Clinical Information

	HC (N = 5)	FESZ (N = 5)	<i>p</i>
Mean age (SD)	22.5 (5.3)	21.9 (5.9)	.55
Gender (male/female)	3/2	3/2	1.0
Premorbid IQ (Oral reading)	108.0 (20.6)	106.8 (5.2)	.9
Current IQ	120.8 (11.4)	105.0 (17.6)	.15
PSES	1.4 (0.55)	2.0 (1.0)	.27
CPZ	–	201.7 (134.9)	N.A.

Values are mean (SD); HC = Healthy controls; FESZ = First episode schizophrenia patients; CPZ = Chlorpromazine; PSES = Parental socioeconomic status; N.A. = not applicable. CPZ equivalents were calculated according to Stoll⁴³ and Woods.⁴⁴

Table 2

Absolute and Percent Volume Differences (Mean [SD]) for ROI Obtained from FreeSurfer Analysis Using Masks Generated by MABS, BET, and FreeSurfer Compared with Manual Masking

	MABS		BET		FS	
		%		%		%
Global volume						
Total gray matter volume	5,133.2 (5.386)	.6 (.9)	21,893 (8,562)**	3.0 (1.2)	16,390.8 (9,037)*	2.4 (1.2)
Subcortical regions						
L Hippocampus	83.6 (53.9)	2 (1.3)	95.4 (60.34)	2.3 (1.4)	164.1 (64.5)*	3.9 (1.6)
L Amygdala	100.9 (105.2)	6.6 (6.9)	119.2 (91.4)	8.2 (6.4)	119.2 (91.4)	8.4 (6.7)
R Hippocampus	90.6 (59)	2.1 (1.4)	85.8 (46.3)	2.0 (1.0)	133.4 (107)	3.1 (2.4)
R Amygdala	83.4 (55.1)	5.4 (3.8)	103.1 (73)	6.7 (4.7)	144.2 (73.1)	9.2 (5.0)
Cortical regions						
L Caudal anterior cingulate	76.2 (66.8)	4.5 (5.3)	79.5 (61.2)*	4.5 (3.7)	74.9 (34.1)	3.7 (1.9)
R Caudal anterior cingulate	100 (78.5)	4.5 (3.4)	140 (140.8)	6.3 (5.1)	86 (65.5)	3.7 (2.5)
L Caudal middle frontal	146(188)	1.8 (2.2)	181(176)	2.3 (2.1)	373(263)*	5.1 (3.3)
R Caudal middle frontal	143(76.8)	1.9 (1.7)	205(122)	3.0 (2.1)	210(126)	3.1 (1.7)
L Rostral anterior cingulate	170.4(189)	6.3 (7.3)	211 (164)*	7.9 (6.2)	156.6(167.1)	5.7 (6.6)
R Rostral anterior cingulate	120(84)	5.8 (3.2)	211 (164)	7.9 (6.9)	171(145)	6.2 (3.5)
L Rostral middle frontal	326.6(218)	1.7 (1.3)	2,377(1,031)**	14.0 (6.3)	441.1(171.4)	2.4 (1.1)
R Rostral middle frontal	441(320)	2.3 (1.6)	3,136(1,556)*	19.0 (10)	471(238)	2.6 (1.9)
L Superior temporal gyrus	342(299)	3.0 (3.)	596(509)	5.0 (4.1)	430(284)	3.4 (2.0)
R Superior temporal gyrus	327(210)	2.7 (2.0)	507(470)	4.5 (4.3)	325(212)	2.6 (1.9)
L Heschl's gyrus	40.6(47.4)	3.2 (3.1)	48(44.6)*	3.9 (3.2)	36.5(23)	3.0 (1.8)
R Heschl's gyrus	231(568)	7.2 (5.3)	231(588)	6.0 (4.2)	196(551)	4.8 (8.6)
L Insula	279.4(312)	4.1 (4.3)	156.4(98.6)	2.3 (1.6)	194.9(234.9)	2.7 (3.1)
R Insula	300(297)	4.0 (3.9)	285(187)	4.0 (2.5)	357(274)	4.9 (3.9)
L Entorhinal cortex	167(138)	9.3 (7.2)	197(174)	11.0 (11)	201(205)	11.0 (11)
R Entorhinal cortex	92.3(70.7)	5.8 (4.6)	172(197)	10.0 (11.7)	108(90.7)	7.0 (6.7)
L Lateral orbital cortex	128(96)	1.5 (1.5)	530(307)**	7.2 (4.8)	640(425)**	5.0 (4.0)
R Lateral orbital cortex	305(248)	3.6 (2.9)	632(532)*	8.5 (7.3)	514(521)	6.6 (7.6)

	MABS	%	BET	%	FS	%
L Medial orbital cortex	209(145)	3.6 (2.7)	756(521)**	14.2 (9.4)	276(237)	4.7 (4.0)
R Medial orbital L R Medial orbital cortex	234(137)	4.4 (2.7)	795(334)**	15.7 (72)	201(139)	3.6 (2.4)

Percent volume difference (%) of each ROI was calculated according to (2*(Method – MM)/(Method +MM)). MABS = Multi-atlas brain segmentation; BET = Brain Extraction Tool; FS = FreeSurfer v5.3; L = Left; R = Right.

* $p < .05$,
** $p < .01$.

Table 3

Intraclass Correlation Coefficients for ROI Volumes Extracted Using MABS, BET, and FreeSurfer Masking Methods Relative to Manual Masking

	MABS	BET	FS
Global volume			
Total gray matter volume	.99	.96	.94
Subcortical regions			
L Hippocampus	.97	.95	.89
L Amygdala	.7	.5	.6
R Hippocampus	.95	.9	.95
R Amygdala	.92	.8	.89
Mean ICC (S.D.)	.9066 (.12)	.822 (.19)*	.8534 (.144)*
Cortical regions			
L Caudal anterior cingulate	.99	.99	.99
L Caudal middle frontal	.99	.988	.99
L Rostral anterior cingulate	.86	.84	.84
L Rostral middle frontal	.99	.7	.7
L Superior temporal gyrus	.98	.9	.9
L Heschl's gyrus	.97	.97	.97
L Insula	.9	.9	.9
L Entorhinal cortex	.84	.8	.7
L Lateral orbital cortex	.98	.9	.9
L Medial orbital cortex	.91	.81	.24
L Mean ICC (S.D.)	.941 (.058)	.8798 (.093)*	.813 (.23)*
R Caudal anterior cingulate	.99	.99	.97
R Caudal middle frontal	.98	.96	.96
R Rostral anterior cingulate	.97	.95	.95
R Rostral middle frontal	.9	.63	.99
R Superior temporal gyrus	.97	.89	.7
R Heschl's gyrus	.92	.88	.78
R Insula	.86	.91	.84
R Entorhinal cortex	.98	.87	.97
R Lateral orbital cortex	.95	.83	.76
R Medial orbital cortex	.86	.37	.76
R Mean ICC (S.D.)	.94 (.05)	.83 (.2)*	.87 (.11)*

MABS = Multi-atlas brain segmentation; BET = Brain Extraction Tool; FS = FreeSurfer v5.3; L = Left; R = Right.