Bootstrapping Multi-atlas Hippocampal Segmentation with MAGeT-Brain

Pipitone J., Winterburn J., Lerch J., Pruessner J., Lepage M.,
Voineskos A., Chakravarty M.M., and
the Alzheimer's Disease Neuroimaging Initiative

April 9, 2013

Abstract

Neuroimaging research often relies on automated anatomical segmentations of MR images of the brain. Current multi-atlas based approaches provide accurate segmentations of brain images by propagating manually derived segmentations of specific neuroanatomical structures to unlabelled data. These approaches often rely on a large number of such manually segmented atlases that take significant time and expertise to produce. We present an algorithm for the automatic segmentation of the hippocampus that minimizes the number of atlases needed while still achieving similar accuracy to multi-atlas approaches.

finish

1 Introduction

The hippocampus is of particular interest to many researchers because it is implicated in forms of brain dysfunction such as Alzheimer's disease(Sabuncu et al., 2011) and schizophrenia(Narr et al., 2004; Karnik-Henry et al., 2012), and has functional significance in cognitive processes such as learning and memory(den Heijer et al., 2012; Scoville and Milner, 2000). For many research questions involving magnetic resonance imaging (MRI) data accurate identification of the hippocampus and its subregions is a necessary first step to better understand the individual neuroanatomy of subjects.

Currently, the gold standard for neuroanatomical segmentation is manual delineation by an expert human rater. This is problematic for hippocampal segmentation for several reasons. First, manual segmentation takes a significant investment of time and expertise (Hammers et al., 2003) which may not be readily available to researchers or clinicians. Second, the amount of data produced in neuroimaging experiments increasingly exceeds the capacity for identification of specific neuroanatomical structures by an expert manual rater. Third, the true definition of hippocampal anatomy in MR images is disputed (Geuze et al., 2004), as evidenced by efforts to create an unified segmentation protocol (Jack et al., 2011).

Compounding each of these problems is the significant neuroanatomical variability in the hippocampus throughout the course of aging, maturation, and neuropsychiatric disorders (**?**Sabuncu et al., 2011; **?**; Gogtay et al., 2006; Narr et al., 2002). The result is that existing hippocampal atlases available to a researcher may not accurately represent neuroanatomy of a specific population under study. Additionally, in the course of a research or clinical study, it may be necessary to make adjustments to hippocampal definition as a means of hypothesis testing. For example, Poppenk (Poppenk and Moscovitch, 2011) found that subdividing the hippocampus into anterior and posterior regions resulted in a predictive relationship between volume difference of those regions and recollection memory performance. Making such modifications to a set of MRI data segmentations requires additional manual effort.

Automated segmentation techniques do not require human intervention but do require *a priori* anatomical information to guide segmentations. In this paper we focus on methods that use manually segmented MRI atlases as anatomical priors, as these methods achieve some of the best automated hippocampal segmentation accuracies to-date. This technique was first developed using a single atlas prior (known as single-atlas, or model-based, segmentation)(**??**). Volumetric image registration is used to estimate a fit between the

neuroanatomy of an atlas and target images. Labelling of the target image is achieved by applying the resulting transformation to the atlas labels to bring them into the target image space (*label propagation*). This method is limited in accuracy by the introduction of estimation errors in registration and partial volume effects in label resampling, and errors introduced when the anatomy of the atlas is unrepresentative of the target anatomy.

Multi-atlas segmentation techniques address these limitations by combining segmentation information from a series of expertly segmented atlases (Heckemann et al., 2006, 2011; Collins and Pruessner, 2010; **?**; Aljabar et al., 2009; Leung et al., 2010; Wolz et al., 2010). Each atlas image is registered to a target image, and label propagation is performed to produce several labellings of the target image (one from each atlas). A *label fusion* technique, such as voxel-wise voting, is used merge these labels into a definitive segmentation for the target. In addition, *atlas selection* techniques are often used to exclude atlases from label fusion that are dissimilar to a target image in order to reduce error from unrepresentative anatomy (Aljabar et al., 2009). Cross-correlation or normalised mutual information of image intensities are common measures of image similarity used in atlas selection.

Multi-atlas methods have been very successfully applied to hippocampal segmentation. Collins et al. found near-manual segmentation performance using an atlas library of 80 T1-weighted atlas images from the ICBM152 dataset, the ANIMAL nonlinear registration algorithm, normalised mutual information as an atlas selection similarity metric, and majority vote for label fusion (Collins and Pruessner, 2010). The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a commonly used benchmarking dataset of MR images of controls and patients with MCI or Alzheimer's (see Methods for more information on the ADNI dataset). Leung et al. tuned parameters for registration and label fusion to the segmentation of ADNI1 1 year dataset of images with an atlas library of 55 images (Leung et al., 2010). The MAPER whole brain segmentation algorithm (Heckemann et al., 2006, 2011), using 30 atlases, on all ADNI1 baseline images (Heckemann et al., 2011). Lotjonen et al. use images from the ADNI1 baseline dataset and 30 atlases with a proprietary non-linear registration method based on intensity differences, and post-processing step using a graph cuts algorithm to optimise fused segmentations against a spatial intensity prior (**?**).

No story here

The LEAP algorithm is an elegant modification to the basic multi-atlas strategy (Wolz et al., 2010). The atlas library is grown, beginning with a set of manually labelled atlases, and successively incorporats unlabelled target images after being labelling using multi-atlas techniques. The sequence in which target images are labelled is chosen so that the similarity between the atlas images and the target images is minimised at each step, effectively allowing for deformations between very dissimilar images to be broken up into sequences of smaller deformations. With an atlas library of 30 MR images, LEAP was used to segment the ADNI1 baseline dataset, achieving a mean Dice score of 0.85 with manual segmentations.

While not purely nulti-atlas techniques, there are several important algorithms for hippocampal segmentation that inform our approach. The popular FreeSurfer application's whole brain segmentation algoritm uses a probabilistic atlas of anatomical and tissue classes along with spatial constraints for class labels encoded using a Markov random field model (Fischl et al., 2002). When segmenting hippocampal subfields, FreeSurfer employs a Bayesian inference algorithm using a probabilistic atlas of anatomical classes as a prior, and a likelihood model of how those classes translate into MR image intensities, both trained on manual segmentations of high resolution MR images (**?**). Yushkevitch et al. describe a semi-automated method for hippocampal subfield segmentation of focal T2 images(**?**). The unlabelled MR image must be manually partitioned into 'head', 'body' and 'tail', and then multi-atlas methods are used to segment the image. Finally, an AdaBoost-based bias correction classifier is trained on texture, spatial location, and intensities of manual segmentations and is applied to fix mislabelled voxels.

Aside from the algorithmic choices used in multi-atlas segmentation, it is natural to ask about how the features of the atlases themselves impact the resulting segmentations. As noted, by choosing atlases ranked most similar to a target image by voxel intensity profile, segmentation accuracy is improved, suggesting that neuroanatomical similarity plays a strong role (Aljabar et al., 2009). Carmichael et al. explored this directly and found that when using only one atlas the important factors leading to improved accuracy are that the atlas have neuroanatomical features that match the target, and that the atlas segmentation use the same protocol as the gold-standard (Carmichael et al., 2005). Nestor et al. found that hippocampal segmentation protocols that include more dorsal white-matter and posterior anatomy tended to produce higher overlap and better accuracy at distinguishing disease classes in the ADNI1 1 year dataset (Nestor et al., 2012). These results suggest both atlas library neuroanatomy and delineation protocol play a significant role in the

resulting segmentation.

Considered along with our earlier discussion on the difficulty of producing manual segmentations of MR images and the need for adaptable segmentation definitions in order to conduct research, this presents a real problem of labour and expertise when using existing multi-atlas segmentation methods which rely on relatively large atlas libraries (typically between 30 and 80 atlases). Indeed, it may be especially prohibitive to use these methods in situations where producing a single atlas is challenging (e.g. histology-based atlases, or atlases from very high resolution images). In this paper we address the problem of producing accurate segmentations using small numbers of manually segmented atlases.

Our algorith, called MAGeT-Brain (*M*ultiple *A*utomatically *G*enerated *T*emplates), is an extension to the basic multi-atlas-based segmentation schema(**?**). Principally, we explore the possibility of using a small atlas library to bootstrap a much larger *template library* composed of images taken from the target population. The template library is then used to segment the targets in a similar fashion to basic multi-atlas segmentation: by label propagation and label fusion. The intuition driving this approach is that by generating a template library we leverage the unique neuroanatomy of target population on hand to initialize the segmentation process and improve accuracy over direct propagation from the atlas library to unlabelled targets while also using fewer manually segmented atlases.

The insight of generating a template library is not new. Heckemann et al. compared "indirect" segmentation – taking a single atlas and propagating the labels to intermediate targets before fusing them in a target image space – to multi-atlas segmentation and found that the indirect approach performed worse (Heckemann et al., 2006). In this paper we continue the same line of investigation but explore the performance when using multiple atlases as well as the effect of different registration and fusion methods.

The LEAP algorithm (Wolz et al., 2010), described above, is another example of indirect segmentation previously explored. LEAP proceeds by iteratively segmenting unlabelled images most similar to the atlas library images and then incorporating the labelled images into the atlas library for future iterations. The novelty explored in our current work is to demonstrate the viability of achieving comparable segmentation accuracy using the basic multi-atlas schema and using significantly fewer manually created atlases.

In previous work (**?**), we applied MAGeT-Brain to segmentation of the human striatum, globus pallidus, and thalamus using a single histologically-derived atlas. The contribution of the present work is to extend our approach to the human hippocampus and perform a series of experiments to rigorous validate the method. First, we conduct an extensive cross-validation of MAGeT-Brain and basic multi-atlas segmentation on a subset of the ADNI1 dataset to assess the accuracy of MAGeT-Brain under various parameter settings (number of atlases and templates, registration and fusion methods). With the best performing parameter configuration discovered above, we estimate MAGeT-Brain intra-rater reliability by segmenting separately acquired T1 images of the atlas subjects. For this experiment, we use the Winterburn atlases: digital hippocampal subfield segmentations of five in-vivo high-resolution (300u isotropic) T1-weighted MR scans(**?**). To validate MAGeT-Brain in a real world situation, we segment the entire ADNI1 Complete 1Yr dataset and compare our segmentations to established automated and manual segmentations. Additionally, to ensure MAGeT-Brain accuracy across disease categories, we also compare MAGeT-Brain segmentations to manual segmentations of 139 first episode schizophrenic patients.

2 Methods

2.1 MAGeT-Brain Algorithm

In this paper we will be making a distinction between an atlas and a template – typically these terms are used roughly interchangeably. The term *atlas* is taken to refer to two image volumes: an intensity image (*atlas image*) and a corresponding manual segmentation image (*atlas labels*). *Template* refers more generically to any image and corresponding labelling, manual or computed, when it is *used as* as a model in the segmentation of another image. The terms *atlas library* and *template library* mean a set of such images. Additionally, we will use the terms *target* to refer to an intensity image for which we would like an segmentation.

The simplest form of multi-atlas segmentation combines labellings derived from several atlases by way of label fusion (Heckemann et al., 2006, 2011). We will refer to this as *basic multi-atlas segmentation*. The schema as for this method is as follows:

1. An atlas library and set of target images are given as input. The atlas library is used as a template library in the following steps;

2. Each atlas intensity image is nonlinearly registered to each target intensity image;

3. Label images from each atlas are propagated via the resulting transformations to the target image space; and

4. the resulting labels are fused to produce a single, definitive segmentation.

The particular registration and voting method used are left unspecified.

MAGeT-Brain is best understood as an extension of the basic multi-atlas segmentation schema. Instead of using the atlas library to directly label the target images, a subset of the input images are selected as template images and then labelled. The choice of targets used in the template library can be made to reflect the neuroanatomy or demographics of the target set as a whole (for instance, by sampling equally from cases and controls). Once the template library images have been chosen, a truncated version of basic multi-atlas segmentation is used to label the template library images without performing label fusion. Instead, each template image receives multiple labellings: one from each atlas image. A second round of basic multi-atlas segmentation uses the template library to segment the entire set of target images (including those images used in the template library). Label fusion in this final step fuses all labels from all templates. To summarize, figure 1 describes the MAGeT-Brain algorithm in pseudocode.

Source code for MAGeT-Brain can be found at http://github.com/pipitone/MAGeTbrain.

Algorithm 1 Pseudocode for the MAGeT-Brain algorithm

```
function BasicMultiAtlasSegmentation(Templates, Subjects)
    for all target do
        for all template do
            propagate all labels for template to target space
            store target labels
        end for
        fuse target labels
    end for
end function

function MAGeTBrain(Subjects, Atlases, n)
    for i = 1 → n do
        choose a target to be used as a template
        propagate labels from each atlas to template space
        store the template with all of its labels
    end for
    MultiAtlas(Templates, Subjects)
end function
```

2.2 Subjects

Our experiments use three distinct subject datasets.

2.2.1 ADNI1 1.5T Complete 1Yr Dataset

Clinical, demographic and pre-processed T1-weighted MRI were downloaded by the authors from the ADNI1 database (adni.loni.ucla.edu) between March 2012 and August 2012. The image dataset download was the "ADNI1:Complete 1Yr 1.5T" standardized dataset available from ADNI [1] (Wyman et al., 2012). This image collection contains uniformly preprocessed images which have been designated to be the "best" after quality control. All images were acquired using 1.5T scanners (General Electric Healthcare, Philips Medical Systems or Siemens Medical Solutions) at multiple sites using the protocol described in (**?**). Representative 1.5T imaging parameters were TR = 2400ms, TI = 1000ms, TE = 3.5ms, flip angle = 8○, field of view = 240 x 240mm, a 192 x 192 x 166 matrix (x, y, and z directions) yielding a voxel resolution of 1.25 x 1.25 x 1.2 mm^3. Clinical and demographic data are shown in table **??**.

[1] http://adni.loni.ucla.edu/methods/mri-analysis/adni-standardized-data/

Table 1: Schizophrenia First Episode Patient Demographics

	N	FEP
		$N = 81$
Age	80	21 23 26
Gender : M	81	63% (51)
Handedness : ambi	81	6% (5)
left		5% (4)
right		89% (72)
Education	81	11 13 15
SES : lower	81	31% (25)
middle		54% (44)
upper		15% (12)
FSIQ	79	88 102 109

$_a$ b $_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables.
N is the number of non–missing values.
Numbers after percents are frequencies.

For a subset of ADNI1 images, labels of the left and right hippocampi are available (herein refered to as SNT labels). Semi-automated hippocampal volumetry was carried out using a commercially available high dimensional brain mapping tool (Medtronic Surgical Navigation Technologies, Louisville, CO), that has previously been validated and compared to manual tracing of the hippocampus (Hsu et al., 2002). Measurement of hippocampal volume is achieved first by placing manually 22 control points as local landmarks for the hippocampus on the individual brain MRI data: one landmark at the hippocampal head, one at the tail, and four per image (i.e., at the superior, inferior, medial and lateral boundaries) on five equally spaced images perpendicular to the long axis of the hippocampus. Second, fluid image transformation is used to match the individual brains to a template brain (Christensen et al., 1997). The pixels corresponding to the hippocampus are then labeled and counted to obtain volumes. This method of hippocampal voluming has a documented reliability of an intraclass coefficient better than .94 (Hsu et al., 2002).

```
Error:  invalid subscript type 'list'
Error:  object 'tab' not found
```

2.2.2 Schizophrenia First Episode Patients

All patients were recruited and treated through the Prevention and Early Intervention Program for Psychoses (PEPP-Montreal), a specialized early intervention service at the Douglas Mental Health University Institute in Montreal, Canada. People aged 1530 years from the local catchment area suffering from either affective or non-affective psychosis who had not taken antipsychotic medication for more than one month with an IQ above 70 were consecutively admitted as either in- or out-patients. Of those treated at PEPP, only patients aged 18 to 30 years with no previous history of neurological disease or head trauma causing loss of consciousness were eligible for the neuroimaging study; only those suffering from schizophrenia spectrum disorders were considered for this analysis. For complete program details see Malla et al. (**?**).

Scanning was carried out at the Montreal Neurological Institute on a 1.5-T Siemens whole body MRI system. Structural T1 volumes were acquired for each participant using a three-dimensional (3D) gradient echo pulse sequence with sagittal volume excitation (repetition time=22ms, echo time=9.2ms, flip angle=30∘, 180 1mm contiguous sagittal slices). The rectangular field-of-view for the images was 256mm (SI)204mm (AP). Subject demographics are shown in table 1.

The hippocampus were traced following a validated protocol developed by Dr Jens Pruessner (Pruessner et al., 2000). A recent update to this protocol by Dr J Pruessner in 2006 allows to accurately and consistently subdivide the hippocampus into three different subregions: head, body, and tail.

Table 2: ANIMAL registration parameters

Parameters	Stage 1	Stage 2	Stage 3
Model Blur (FWHM)	8	8	4
Input Blur (FWHM)	8	8	4
Iterations	30	30	10
Step	8x8x8	4x4x4	2x2x2
Sub-Lattice	6	6	6
Lattice Diameter	24x24x24	12x12x12	6x6x6

2.2.3 Winterburn Atlases

The Winterburn atlases (**?**) are digital hippocampal segmentations of of five in-vivo 300u isotropic T1-weighted MR images. The segmentations include subfield segmentations for the cornus ammonis (CA) 1, CA4, dentate gyrus, subiculum, and CA 2 and 3 combined. Subjects in the Winterburn atlases range in age from 29-57 years (mean age of 37), and include two males and three females.

In addition to the high-resolution scans distributed as part of the Winterburn atlases, we also obtained additional T1 BRAVO scans of four of the five subjects.

demographics

explain scan in more detai

2.3 Registration Methods

Before registration, all images underwent preprocessing with the N3 algorithm (Sled et al., 1998) to minimize intensity nonuniformity. In our experiments we use one of two non-linear image registration methods.

2.3.1 Automatic Normalization and Image Matching and Anatomical Labeling (ANIMAL)

The ANIMAL algorithm carries out image registration is in two phases. In the first, a 12-parameter linear transformation (3 translations, rotations, scales, shears) is estimated between images using an algorithm that maximizes the correlation between blurred MR intensities and gradient magnitude over the whole brain (Collins et al.). In the second phase, nonlinear registration is completed using the ANIMAL algorithm (Collins et al., 1995): an iterative procedure that estimates a 3D deformation field between two MR images. At first, large deformations are estimated using blurred version of the input data. These larger deformations are then input to subsequent steps where the fit is refined by estimating smaller deformations on data blurred with a Gaussian kernel with a smaller FWHM. The final transformation is a set of local translations defined on a bed of equally spaced nodes that were estimated through the optimization of the correlation coefficient. For the purposes of this work we used the regularization parameters optimized in Robbins et al. (Robbins et al., 2004), displayed in table 2.3.1.

2.3.2 Automatic Normalization Tools (ANTS)

ANTS is a diffeomorphic registration algorithm which provides great flexibility over the choice of transformation model, objective function, and the consistency of the final transformation. The transformation is estimated in a hierarchical fashion where the MRI data is subsampled, allowing large deformations to be estimated and successively refined at later hierarchical stages (where the data is subsampled to a finer grid). The deformation field and the objective function are regularized with a Gaussian kernel at each level of the hierarchy. The ANTS algorithm is freely available http://www.picsl.upenn.edu/ANTS/. We used an implementation of the ANTS algorithm compatible with the MINC data format, mincANTS https://github.com/vfonov/mincANTS.

We used the following command line when running ANTS :

```
mincANTS 3 -m PR[target_file.mnc,source_file.mnc,1,4]
 --number-of-affine-iterations 10000x10000x10000x10000x10000
 --affine-gradient-descent-option 0.5x0.95x1.e-4x1.e-4
 --use-Histogram-Matching --MI-option 32x16000
 -r Gauss[3,0] -t SyN[0.5] -i 100x100x100x20
```

```
-o transformation.xfm
```

These settings were adapted from the "reasonable starting point" given in the ANTS manual [2].

2.4 Label Fusion

Label fusion is a term given to the process of combining the information from several candidate labellings for an intensity image into a single labelling. In this paper we explore three fusion methods.

2.4.1 Voxel-wise Majority Vote

Labels are propagated from all template library images to a target. Each output voxel is given the most frequent label at that voxel location amongst all candidate labellings. Ties are broken arbitrarily.

2.4.2 Cross-correlation Weighted Majority Vote

An optimal combination of targets from the template library has previously been shown to improve segmentation accuracy (Aljabar et al., 2009; Collins and Pruessner, 2010). In this method, each template library image is ranked in similarity to each unlabelled image by the normalized cross-correlation (CC) of image intensities after linear registration, over a region of interest (ROI) generously encompassing the hippocampus. Only the top ranked template library image labels are used in a voxel-wise majority vote. The ROI is heuristically defined as the extent of all atlas labels after linear registration to the template, dilated by three voxels (**?**). The number of top ranked template library image labels is a configurable parameter.

The `xcorr_vol` utility from the ANIMAL toolkit is used to calculate the cross-correlation similarity measure.

2.4.3 Normalised Mutual Information Weighted Majority Vote

This method is similar to cross-correlation weighted voting except that image similarity is calculated by the normalised mutual information score over the region of interest(Studholme et al., 2001). The `itk_similarity` utility from the EZMinc toolkit[3] is used to calculate the normalised mutual information measure between to images.

2.5 Assessing Segmentation Agreement

The Dice similarity coefficient (DSC) assesses the agreement between two segmentations. It is one of the most widely used measures of segmentation performance, and we use it as the basis of comparison throughout this paper. Additionally, we report the Jaccard index, another commonly used similarity measure:

$$\text{Dice's coefficient (DSC)} = \frac{2|A \cap B|}{|A| + |B|}$$

$$\text{Jaccard (J)} = \frac{|A \cap B|}{|A \cup B|} = \frac{DSC}{(2 - DSC)}$$

where A and B are the regions being compared, and the cardinality is the volume measured in voxels.

2.6 Experiments

The following experiments were performed to assess the performance of MAGeT-Brain with various parameter settings as well as on diverse datasets. In each experiment we contrast the performance of MAGeT-Brain with that standard single- and multi-atlas segmentations derived from the same atlas library.

[2] https://sourceforge.net/projects/advants/files/Documentation/
[3] https://github.com/vfonov/EZminc

Table 3: ADNI-1 cross-validation subset demographics

	CN $N = 23$	LMCI $N = 23$	AD $N = 23$	Combined $N = 69$
Age at baseline Years	72.2 75.5 78.5	71.0 77.1 81.4	71.7 77.8 81.8	71.5 76.6 81.3
Sex : Female	43% (10)	43% (10)	43% (10)	43% (30)
Education	16.0 16.0 18.0	15.0 16.0 18.0	12.0 16.0 16.5	14.0 16.0 18.0
Ethnicity : Unknown	0% (0)	0% (0)	0% (0)	0% (0)
Not Hisp/Latino	100% (23)	100% (23)	100% (23)	100% (69)
Hisp/Latino	0% (0)	0% (0)	0% (0)	0% (0)
CDR-SB	0.00 0.00 0.00	0.75 1.50 1.50	4.00 4.50 5.00	0.00 1.50 4.00
ADAS 13	4.67 5.67 12.34	14.34 16.00 20.50	23.83 29.00 31.66	10.00 16.00 25.33
MMSE	28.5 29.0 30.0	25.0 27.0 28.0	21.0 23.0 24.0	24.0 27.0 29.0

a b c represent the lower quartile a, the median b, and the upper quartile c for continuous variables.
Numbers after percents are frequencies.

2.6.1 ADNI1 cross-validation

To test the effect of parameter settings on the MAGeT-Brain algorithm, Monte Carlo Cross-Validation (MCCV) (**?**) was performed using a subset of the ADNI1 dataset for validation. This form of cross-validation (described in detail below) allows us to validate MAGeT-Brain with various atlas and template library sizes, registration and label fusion methods.

Dataset evaluated. 69 1.5T images were arbitrarily selected from the *ADNI1:Complete 1Yr 1.5T* standardized dataset. 23 subjects were chosen from each disease category (cognitively normal (CN), MCI and AD). Demographics for this subset are shown in Table 3.

Atlas and template libary. The atlas and template libraries are selected from the dataset evaluated. Atlases consist of images taken from the dataset, with corresponding manual labels provided by SNT. Template library images were selected at random from remaining images not used as atlases. The atlas library size ranged from 1 to 9 images. Template library size ranged from 1 to 20 images.

Registration method. Both the ANTS and ANIMAL registration methods were used.

Label fusion. Majority vote, cross-correlation weighted majority vote, and Normalized Mutual Information weighted majority vote were used. With the weighted majority vote fusion methods, the number of top labels used in the fusion was varied from 1 to 20 images.

Evaluation. Monte Carlo Cross-Validation (MCCV), also known as repeated random sub-sampling cross-validation, consists of repeated rounds of validation in which items from the dataset are randomly sampled (without replacement) and assigned to a training set or validation set **?**.

In this experiment, each validation round consists of the following steps. First, an atlas library is selected (the training set) and, from the remaining images, a template library and subject to be segmented is chosen (the validation set). Second, MAGeT-Brain and basic multi-atlas segmentation are performed on the subject using the selected atlas and template libraries. Third, the accuracy of the resulting segmentations are measured against the SNT labels.

A total of ten validation rounds are perfomed on each subject in the dataset, for each combination of parameter settings. Reported subject segmentation accuracy is the averaged over the ten validation rounds. The parameter settings we explore are: atlas library size (1-9), template library size (1-20), registration method (ANTS vs. ANIMAL), and label fusion method (MV, XC-WV, NMI-WV). A total of $10 * 69 * 9 * 20 * 2 * 3 = 7452000$ validation rounds were conducted, resulting in a total of 1'490'400 segmentations analysed.

2.6.2 ADNI-1 Complete 1Yr Validation

To test the accuracy of MAGeT-Brain on a real-world task we segment the entire ADNI-1 dataset using an atlas set that is not representative of the target set.

Dataset evaluated. All images from the *ADNI1:Complete 1Yr 1.5T* standardized dataset.

Atlas and template libary. The Winterburn atlases were used as the atlas library. The template library consisted of 21 randomly selected images from the ADNI1 data dataset (7 healthy, MCI and AD subjects). Clinical and demographic data for the template library subjects are shown in Table 4.

Table 4: ADNI1 Complete 1Yr Template Library demographics

	CN $N = 7$	LMCI $N = 7$	AD $N = 7$	Combined $N = 21$
Age at baseline Years	81.1 84.6 86.7	67.2 71.1 77.9	74.2 79.5 81.2	71.1 79.5 84.6
Sex : Female	71% (5)	43% (3)	86% (6)	67% (14)
Education : 10	0% (0)	0% (0)	14% (1)	5% (1)
12	0% (0)	0% (0)	29% (2)	10% (2)
13	0% (0)	0% (0)	29% (2)	10% (2)
14	0% (0)	14% (1)	0% (0)	5% (1)
16	57% (4)	29% (2)	29% (2)	38% (8)
17	0% (0)	14% (1)	0% (0)	5% (1)
18	29% (2)	29% (2)	0% (0)	19% (4)
20	14% (1)	14% (1)	0% (0)	10% (2)
Ethnicity : Unknown	0% (0)	0% (0)	0% (0)	0% (0)
Not Hisp/Latino	100% (7)	100% (7)	100% (7)	100% (21)
Hisp/Latino	0% (0)	0% (0)	0% (0)	0% (0)
CDR-SB	0.00 0.00 0.00	0.75 1.50 1.75	4.50 5.00 7.25	0.00 1.50 4.00
ADAS 13	4.67 5.33 7.00	16.34 18.00 22.50	26.33 30.33 33.00	7.33 17.00 28.33
MMSE : 21	0% (0)	0% (0)	29% (2)	10% (2)
22	0% (0)	0% (0)	14% (1)	5% (1)
23	0% (0)	0% (0)	14% (1)	5% (1)
25	0% (0)	14% (1)	29% (2)	14% (3)
26	0% (0)	14% (1)	14% (1)	10% (2)
27	0% (0)	14% (1)	0% (0)	5% (1)
28	14% (1)	14% (1)	0% (0)	10% (2)
29	29% (2)	43% (3)	0% (0)	24% (5)
30	57% (4)	0% (0)	0% (0)	19% (4)

a b c represent the lower quartile a, the median b, and the upper quartile c for continuous variables.
Numbers after percents are frequencies.

Registration method. The ANTS non-linear registration algorithm was used, as it performed best in the ADNI1 cross-validation experiment (see Results).

Label fusion. Majority vote was used performed equally well in the ADNI1 cross-validation experiment and is the easiest label fusion method to operate.

Evaluation. Since the hippocampal segmentation protocols differ between the ADNI labels and Winterburn atlases, this poses a problem for direct evaluation between labels produced by MAGeT-Brain and the ADNI labels in terms of overlap; we would not expect different segmentation protocols to have a high degree of overlap. Instead, to evaluate the performance of MAGeT-Brain we compare the correlation of MAGeT-Brain segmentation volumes with manual segmentation (SNT) volumes. Additionally, we correlate the hippocampal volumes of established automated segmentation methods to MAGeT-Brain segmentations.

2.6.3 SZ First Episode Patient Validation

The previous experiments explore MAGeT-Brain on an Alzheimer's disease dataset (ADNI). To validate that MAGeT-Brain algorithm works effectively with other diseased brain images, the performance of MAGeT-Brain is measured on a Schizophrenia dataset.

Dataset evaluated. The entire Schizophrenia First Episode Patient (SZFEP) dataset (see section 2.2.2).

Atlas and template libary. The atlas library is composed of the Winterburn T1 atlas images, with whole hippocampus segmentations (no subregions). The template library is composed of 21 images selected at random from the SZFEP dataset.

Registration method. The ANTS registration algorithm is used for image registration.

Label fusion. The majority-vote method is used for label fusion.

Evaluation. The manual segmentation protocol used to segment the Winterburn atlases is similar to, but different from, the protocol used to segment the SZFEP dataset. Therefore, rather than use an overlap metric, MAGeT-Brain hippocampal volumes are compared to the corresponding manual segmentation volumes.

2.6.4 Winterburn Atlases Cross-Validation

In this experiment, the accuracy of the MAGeT-Brain algorithm on hippocampal subregion segmentation is tested using a leave-one-out cross-validation (LOOCV) design.

Dataset evaluated. MAGeT-Brain segmentations are evaluated on two datasets. The WA-BRAVO dataset consists of separately acquired 3T T1 BRAVO images (0.9mm-isotropic voxel) of four of the five Winterburn atlas subjects. The WA-Resampled dataset consists of the Winterburn atlas images and segmentations downsampled to 0.9mm-isotropic voxel resolution.

Atlas and template libary. The atlas library is composed of the Winterburn T1 atlas images and hippocampal subregion segmentations. The template library consists of 15 3T T1 images (0.9mm-isotropic voxels) of healthy subjects as well as all of the images from the dataset being evaluated.

.

reference, der
ics for health
images

Registration method. The ANTS image registration method is used.

Label fusion. The majority vote label fusion method is used.

Evaluation. Leave-one-out-cross-validation (LOOCV) is performed separately for each dataset evaluated (WA-BRAVO, and WA-Resampled) as follows. Each subject in the dataset is segmented by MAGeT-Brain using an atlas library in which that subject's image is excluded. The segmentation volumes of each hippocampal subregion are compared to the expert manual segmentations of the unmodified Winterburn atlases. As a point of comparison, we also calculate the subregion volumes from the Winterburn atlas segmentations after downsampling to 0.9mm-isotropic voxels.

3 Results

3.1 ADNI1 Cross-Validation

In this experiment we conducted 10 rounds of MAGeT-Brain and multi-atlas segmentation of each of 69 subjects at a range of atlas and template library sizes, registration algorithm (ANTS or ANIMAL), and three label fusion techniques. Hippocampal MAGeT-Brain -based segmentations using both ANIMAL and ANTS registration algorithm demonstrate good overlap with manually derived gold-standards (Figure **??**). Qualitatively, both ANIMAL and ANTS -based segmentations demonstrate trend overlap accuracy that increases with the size of atlas library and template library. Improvement in accuracy diminishes noticeably with template libraries larger than ten images.

No marked difference in segmentation accuracy is seen when either ANIMAL or ANTS registration is used with different label fusion techniques, at any atlas or template library sizes. In every parameter configuration, the use of MAGeT-Brain with ANTS registration shows a pronounced increase in segmentation accuracy over MAGeT-Brain with ANIMAL registration. In the remainder of this section, only results using the ANTS registration algorithm will be shown.

It is interesting to note that with an even number of templates, MAGeT-Brain shows a small decrease in performance relative to when one fewer template image is used. See section for a discussion of this behaviour. In the remainder of this section, only results from odd-sized template libraries will be shown.

ref

With an increasing number of templates, MAGeT-Brain -based segmentating using ANTS registration and majority vote label fusion shows improvement in overlap accuracy over multi-atlas-based segmentation, using the same number of atlases and voting method (Figure **??**). The magnitude of improvement over multi-atlas-based segmentation decreases with an increasing number of atlases, with accuracy converging with 7 atlases. Peak improvement in MAGeT-Brain accuracy (0.02 DSC) is found when one atlas is used with a template library of 20 images.

mention failu

In addition to an improvement in accuracy over multi-atlas-based segmentation, MAGeT-Brain also shows a decrease in the variability of segmentation accuracy (Figure). The size of template library necessary to reach a significant (p¡0.5) decrease in variance and standard deviation grows with the size of atlas library used. A template library of 19 images is sufficient to show significant decrease in variance and standard deviation for 3-7 atlases.

ADNI1-xval-

Finally, MAGeT-Brain segmentation volumes show differential agreement with manual (SNT) volumes throughout the range of hippocampal volumes (Figure 4). MAGeT-Brain shows a negligble fixed bias towards producing smaller segmentations, and shows a slight proportional bias towards smaller segmentations with larger hippocampi.

Note limits o
ment?

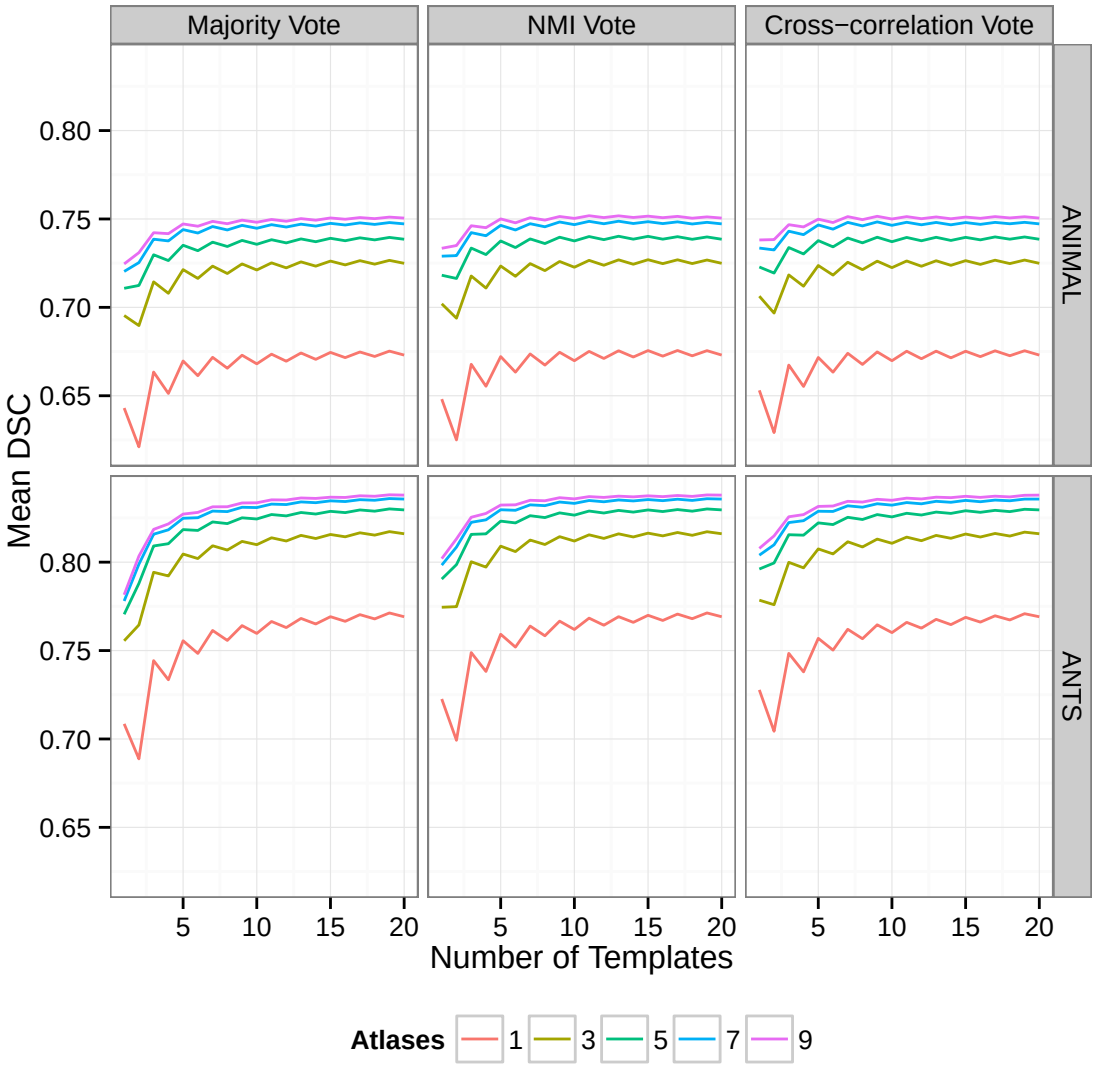

Figure 1: **Mean DSC by atlas and template library size, registration algorithm, and label fusion method.** The mean DSC score is calculated over all subjects from the ADNI1 Cross-Validation dataset for each parameter setting.

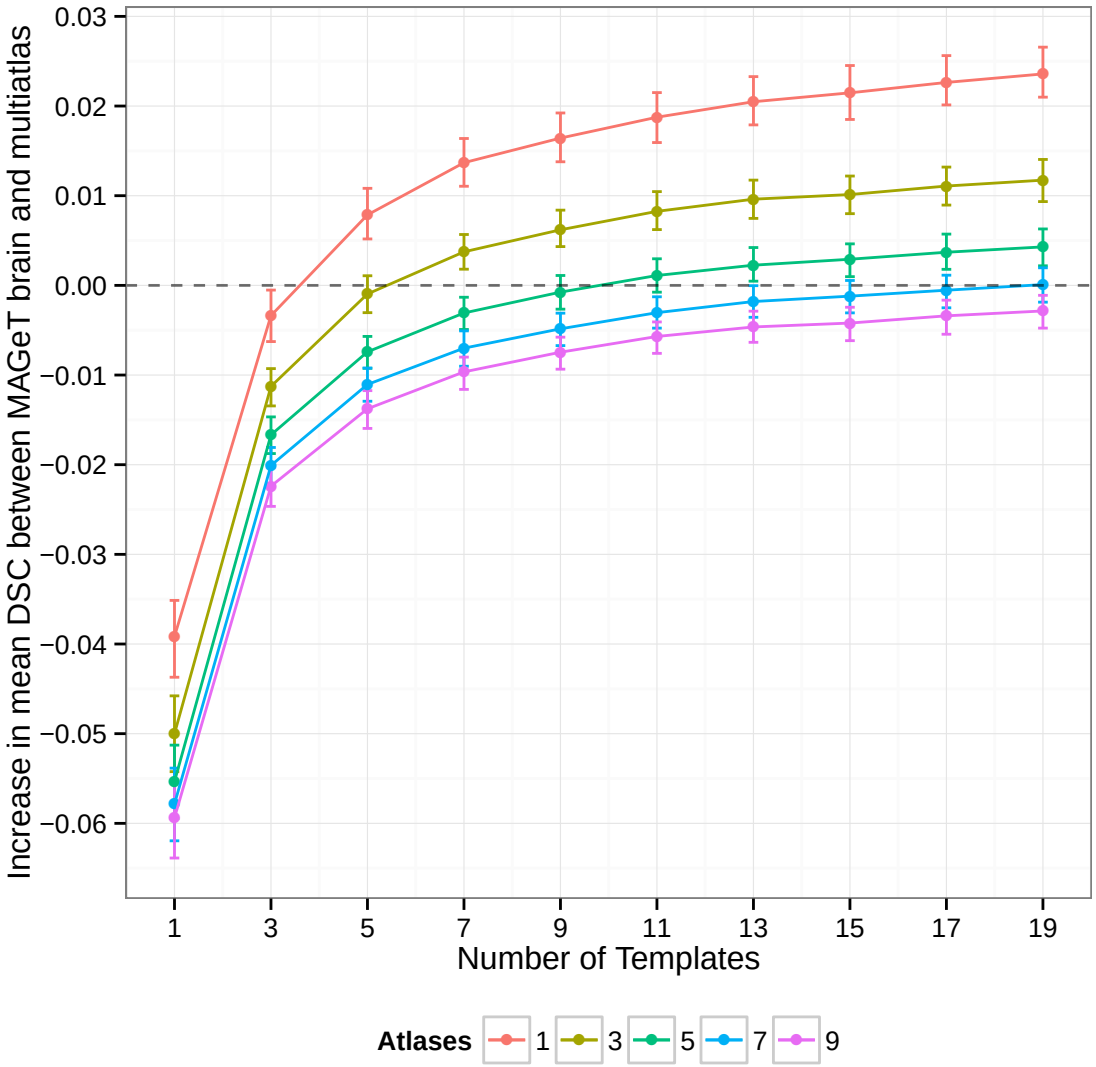

Figure 2: The difference in mean DSC between MAGeT-Brain and multi-atlas segmentations for a range of parameter settings.

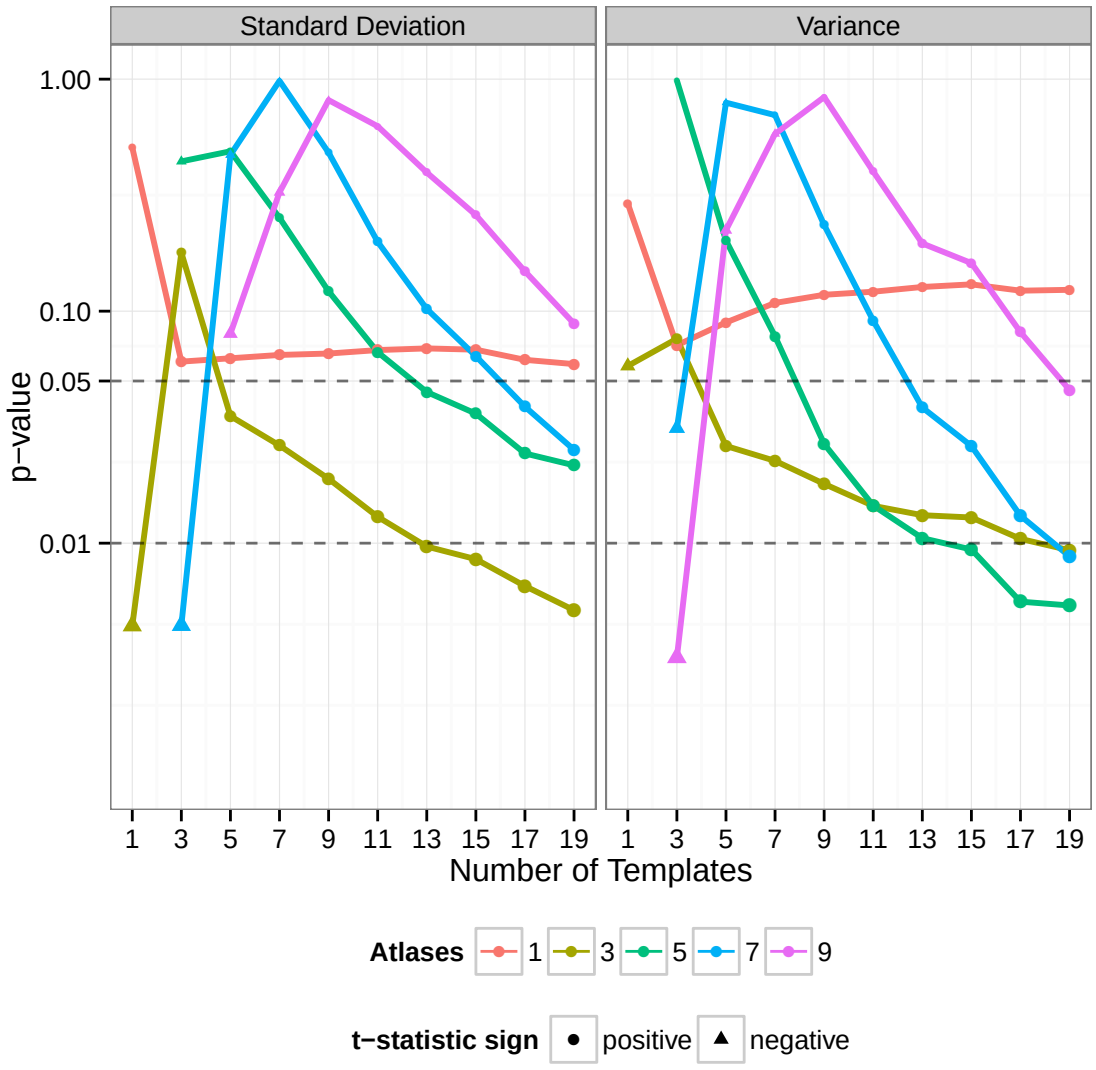

Figure 3: **Difference in variability of MAGeT-Brain vs. multi-atlas segmentation accuracy.** Variability of segmentation accuracy within 10 rounds of validation per subject. Standard deviation and variance is computed per subject for both MAGeT-Brain and multi-atlas, and the distribution of these test statistics is compared with a t-test. The p-value of this test, is shown on the y-axis (scaled logarithmically).

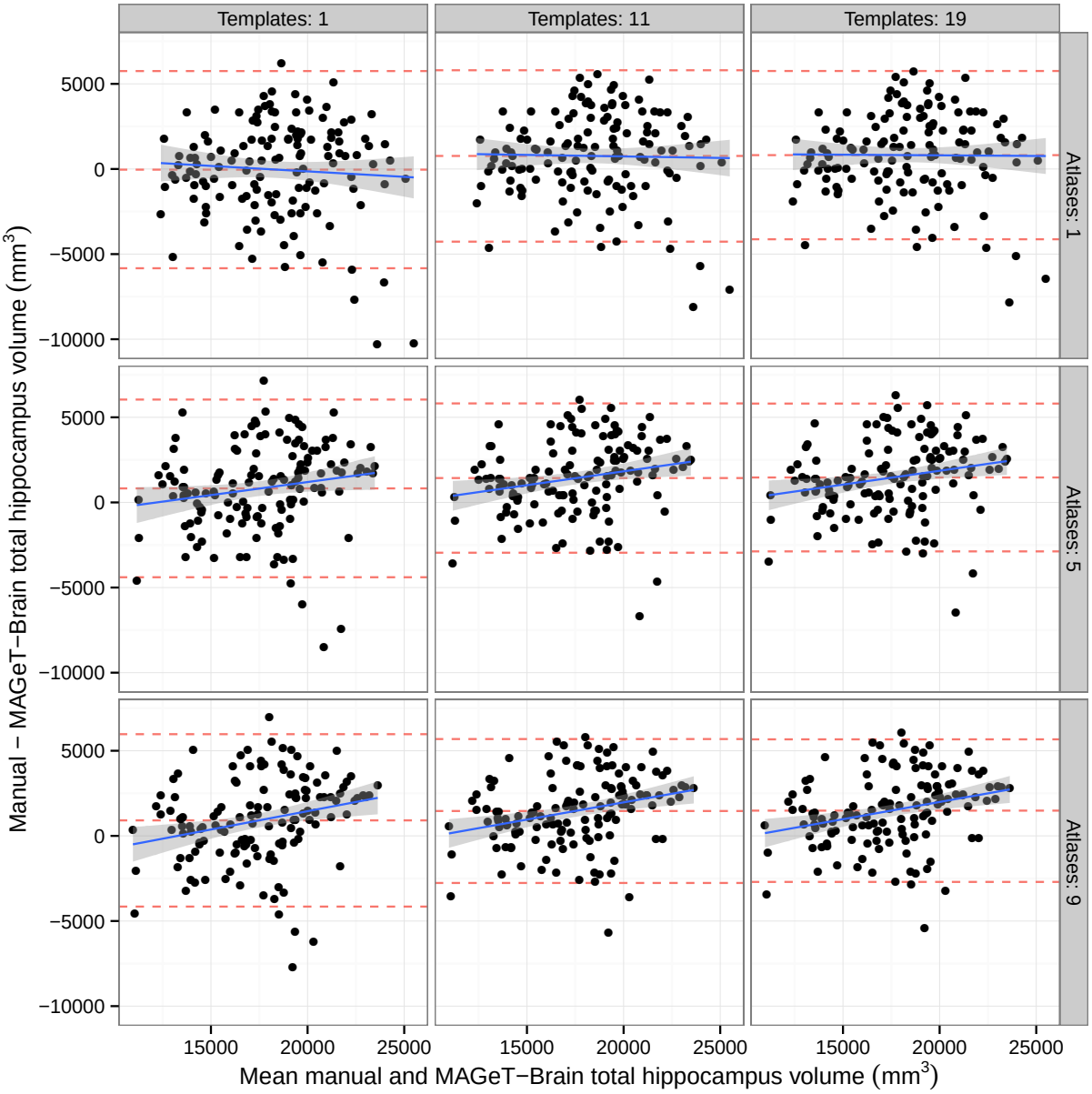

Figure 4: Bland-Altman plots comparing manual and MAGeT-Brain hippocampal volumes when using a varying number of atlases and templates.

Table 5: Pass/fail quality control indicators were supplied with the FreeSurfer volumes downloaded from the ADNI website (we used the temporal lobe quality control indicator, TEMPQC). One of the authors (MP) performed visual quality inspection for MAGeT and FSL segmentations.

	Total	SNT	MAGeT	MAPER	FSL	FS
Images	1909	1445	1909	636	1876	1530
Failures	N/A	–	34	–	27	304

3.2 ADNI1 Complete 1Yr Validation

Based on the results from the ADNI1 Cross-Validation experiment, in this experiment MAGeT-Brain was configured with a template library of 21 randomly chosen subject images (7 from each disease class) and used majority vote label fusion. The entire ADNI1 Complete 1Yr 1.5T dataset was segmented by MAGeT-Brain and we now compare the resulting volumes with those obtained by manual segmentation (SNT), and other automated segmentation techniques (MAPER, FreeSurfer, and FSL). Table **??** shows the total count of segmentations available, including a count of those which have failed a quality control inspection. Only those images which had segmentations from every method are included in the following analysis (a total of 345 images; Table 5).

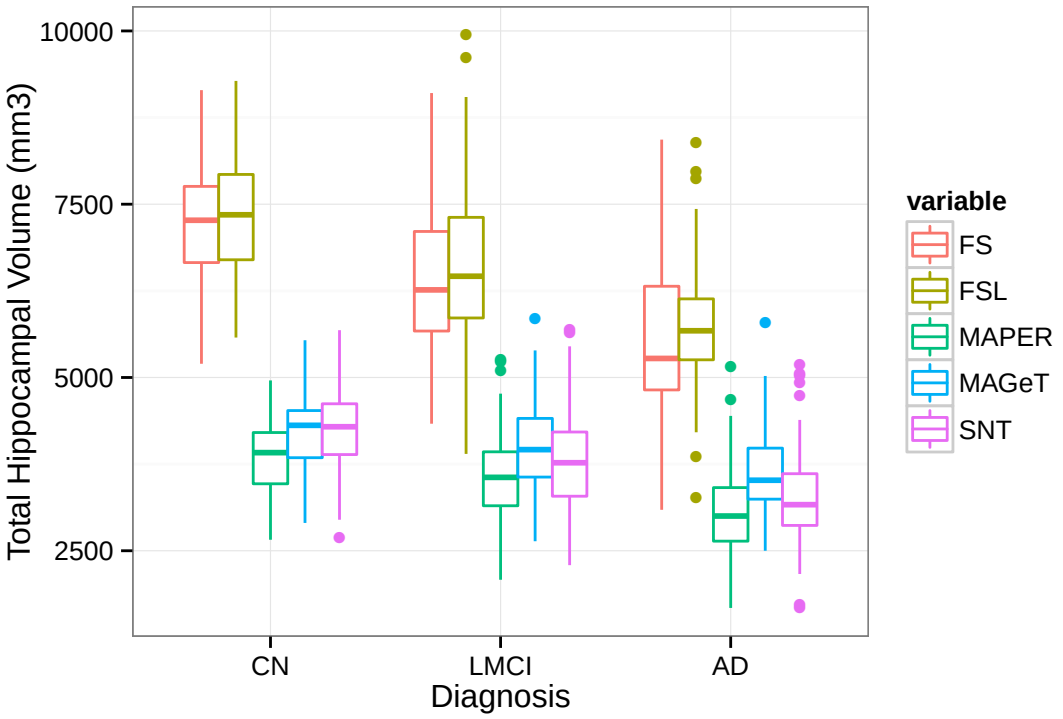

Figure 5: Comparison of hippocampus volumes obtained by FreeSurfer (FS), FSL, MAPER, MAGeT-Brain (MAGeT) and manual (SNT).

Comparing total hippocampal volume, we find close agreement between manual volumes, MAPER and MAGeT volumes across disease categories (Figure 5, Figure 6). FSL and FreeSurfer both produce volumes in close agreement with each other, but which are consistently larger than manually segmented volumes in each disease category.

3.3 First Episode Schizophrenic Patients

In this experiment we move from an Alzheimer's disease dataset to a dataset of first episode schizophrenia parients. This dataset is segmented using MAGeT-Brain with the Winterburn atlases, and a template library of 21 subject images selected at random. Expert manual whole hippocampal segmentations are used as gold standards.

MAGeT-Brain produces hippocampus segmentation volumes that are highly correlated with manual segmentation volumes (Figure 7). Additionally, as we saw in the previous experiments, MAGeT-Brain volumes show a fixed bias towards smaller volumes (although the bias is negligble), and a proportional bias towards producing smaller segmentations for larger hippocampi (Figure 8.

3.4 Winterburn Atlases Cross-Validation

The final experiment explores MAGeT-Brain segmentations of hippocampal subfields. To achieve this, a leave-one-out validation is conducted in which lower-resolution images ($0.9mm^3$ voxels) of each Winterburn atlas subject is segmented using the remaining Winterburn atlases. As a point of comparison, volumes of Winterburn atlases when downsampled to $0.9mm^3$ voxels are also computed.

In general, across hippocampal subregions the percent error in volume between MAGeT-Brain segmentations and the manual Winterburn atlas segmentations compares favourably to error when resampling the atlas segmentations (Figure 9). In particular, the CA1, CA4, and Dentate subregions all show near or smaller percent errors. The Sibiculum and CA2/CA3 subregions show distinctly larger than resampling error.

4 Discussion

5 Conclusion

References

P Aljabar, R a Heckemann, a Hammers, J V Hajnal, and D Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage*, 46(3):726–38, July 2009. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2009.02.018. URL http://www.ncbi.nlm.nih.gov/pubmed/19245840.

Owen T Carmichael, Howard A Aizenstein, Simon W Davis, James T Becker, Paul M Thompson, Carolyn Cidis Meltzer, and Yanxi Liu. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 27(4):979–90, October 2005. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2005.05.005. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2862692&tool=pmcentrez&rendertype=abstract.

G E Christensen, S C Joshi, and M I Miller. Volumetric transformation of brain anatomy. *IEEE transactions on medical imaging*, 16(6):864–77, December 1997. ISSN 0278-0062. doi: 10.1109/42.650882. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=650882.

D L Collins, P Neelin, T M Peters, and A C Evans. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of computer assisted tomography*, 18(2):192–205. ISSN 0363-8715. URL http://www.ncbi.nlm.nih.gov/pubmed/8126267.

D Louis Collins and Jens C Pruessner. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *NeuroImage*, 52 (4):1355–66, October 2010. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2010.04.193. URL http://www.ncbi.nlm.nih.gov/pubmed/20441794.

D. Louis Collins, C. J. Holmes, T. M. Peters, and A. C. Evans. Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping*, 3(3):190–208, October 1995. ISSN 10659471. doi: 10.1002/hbm.460030304. URL http://doi.wiley.com/10.1002/hbm.460030304.

T den Heijer, F Van der Lijn, M W Vernooij, M de Groot, P J Koudstaal, a Van der Lugt, G P Krestin, a Hofman, W J Niessen, and M M B Breteler. Structural and diffusion MRI measures of the hippocampus and memory performance. *NeuroImage*, 63(4):1782–9, December 2012. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2012.08.067. URL http://www.ncbi.nlm.nih.gov/pubmed/22960084.

Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre van der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, Albert Montillo, Nikos Makris, Bruce Rosen, and Anders M Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–55, January 2002. ISSN 0896-6273. URL http://www.ncbi.nlm.nih.gov/pubmed/11832223.

E. Geuze, E. Vermetten, and J D Bremner. MR-based in vivo hippocampal volumetrics: 2. Findings in neuropsychiatric disorders. *Molecular Psychiatry*, 10(2):160, September 2004. doi: 10.1038/sj.mp.4001579. URL http://www.nature.com.myaccess.library.utoronto.ca/mp/journal/v10/n2/full/4001579a.htmlhttp://www.nature.com.myaccess.library.utoronto.ca/mp/journal/v10/n2/pdf/4001579a.pdf.

Nitin Gogtay, Tom F Nugent, David H Herman, Anna Ordonez, Deanna Greenstein, Kiralee M Hayashi, Liv Clasen, Arthur W Toga, Jay N Giedd, Judith L Rapoport, and Paul M Thompson. Dynamic mapping of normal human hippocampal development. *Hippocampus*, 16(8):664–72, January 2006. ISSN 1050-9631. doi: 10.1002/hipo.20193. URL http://www.ncbi.nlm.nih.gov/pubmed/16826559.

Alexander Hammers, Richard Allom, Matthias J Koepp, Samantha L Free, Ralph Myers, Louis Lemieux, Tejal N Mitchell, David J Brooks, and John S Duncan. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human brain mapping*, 19(4):224–47, August 2003. ISSN 1065-9471. doi: 10.1002/hbm.10123. URL http://www.ncbi.nlm.nih.gov/pubmed/12874777.

Rolf A. Heckemann, Joseph V. Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 46 (3):726–38, July 2006. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2009.02.018. URL http://www.ncbi.nlm.nih.gov/pubmed/19245840.

Rolf A Heckemann, Shiva Keihaninejad, Paul Aljabar, Katherine R Gray, Casper Nielsen, Daniel Rueckert, Joseph V Hajnal, and Alexander Hammers. Automatic morphometry in Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 56(4):2024–37, July 2011. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2011.03.014. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3153069&tool=pmcentrez&rendertype=abstract.

Yuan-Yu Hsu, Norbert Schuff, An-Tao Du, Kevin Mark, Xiaoping Zhu, Dawn Hardin, and Michael W Weiner. Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia. *Journal of magnetic resonance imaging : JMRI*, 16(3):305–10, September 2002. ISSN 1053-1807. doi: 10.1002/jmri.10163. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1851676&tool=pmcentrez&rendertype=abstract.

Clifford R Jack, Frederik Barkhof, Matt A Bernstein, Marc Cantillon, Patricia E Cole, Charles Decarli, Bruno Dubois, Simon Duchesne, Nick C Fox, Giovanni B Frisoni, Harald Hampel, Derek L G Hill, Keith Johnson, Jean-François Mangin, Philip Scheltens, Adam J Schwarz, Reisa Sperling, Joyce Suhy, Paul M Thompson, Michael Weiner, and Norman L Foster. Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 7(4):474–485.e4, July 2011. ISSN 1552-5279. doi: 10.1016/j.jalz.2011.04.007. URL http://www.ncbi.nlm.nih.gov/pubmed/21784356.

Meghana S Karnik-Henry, Lei Wang, Deanna M Barch, Michael P Harms, Carolina Campanella, and John G Csernansky. Medial temporal lobe structure and cognition in individuals with schizophrenia and in their non-psychotic siblings. *Schizophrenia research*, 138(2-3):128–35, July 2012. ISSN 1573-2509. doi: 10.1016/j.schres.2012.03.015. URL http://www.ncbi.nlm.nih.gov/pubmed/22542243.

Kelvin K Leung, Josephine Barnes, Gerard R Ridgway, Jonathan W Bartlett, Matthew J Clarkson, Kate Macdonald, Norbert Schuff, Nick C Fox, and Sebastien Ourselin. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *NeuroImage*, 51(4):1345–59, July 2010. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2010.03.018. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2873209&tool=pmcentrez&rendertype=abstract.

Katherine L. Narr, Theo G.M. van Erp, Tyrone D. Cannon, Roger P. Woods, Paul M. Thompson, Seonah Jang, Rebecca Blanton, Veli-Pekka Poutanen, Matti Huttunen, Jouko Lönnqvist, Carl-Gustav Standerksjöld-Nordenstam, Jaakko Kaprio, John C. Mazziotta, and Arthur W. Toga. A Twin Study of Genetic Contributions to Hippocampal Morphology in Schizophrenia. *Neurobiology of Disease*, 11(1):83–95, October 2002. ISSN 09699961. doi: 10.1006/nbdi.2002.0548. URL http://dx.doi.org/10.1006/nbdi.2002.0548.

Katherine L Narr, Paul M Thompson, Philip Szeszko, Delbert Robinson, Seonah Jang, Roger P Woods, Sharon Kim, Kiralee M Hayashi, Dina Asunction, Arthur W Toga, and Robert M Bilder. Regional specificity of hippocampal volume reductions in first-episode schizophrenia. *NeuroImage*, 21(4):1563–75, April 2004. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2003.11.011. URL http://www.ncbi.nlm.nih.gov/pubmed/15050580.

Sean M Nestor, Erin Gibson, Fu-Qiang Gao, Alex Kiss, and Sandra E Black. A Direct Morphometric Comparison of Five Labeling Protocols for Multi-Atlas Driven Automatic Segmentation of the Hippocampus in Alzheimer's Disease. *NeuroImage*, November 2012. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2012.10.081. URL http://www.ncbi.nlm.nih.gov/pubmed/23142652.

Jordan Poppenk and Morris Moscovitch. A Hippocampal Marker of Recollection Memory Ability among Healthy Young Adults: Contributions of Posterior and Anterior Segments. *Neuron*, 72(6):931–937, December 2011. ISSN 0896-6273. doi: 10.1016/j.neuron.2011.10.014. URL http://www.sciencedirect.com/science/article/pii/S089662731100924Xhttp://pdn.sciencedirect.com.myaccess.library.utoronto.ca/science?_ob=MiamiImageURL&_cid=272195&_user=994540&_pii=S089662731100924X&_check=y&_origin=article&_zone=toolbar&_coverDate=22-Dec-2011&view=c&originContentFamily=serial&wchp=dGLbVlV-zSkzk&md5=e75d94a1de9d5c31e146f910b38468da/1-s2.0-S089662731100924X-main.pdfhttp://www.sciencedirect.com.myaccess.library.utoronto.ca/science/article/pii/S089662731100924X.

Steven Robbins, Alan C Evans, D Louis Collins, and Sue Whitesides. Tuning and comparing spatial normalization methods. *Medical image analysis*, 8(3):311–23, September 2004. ISSN 1361-8415. doi: 10.1016/j.media.2004.06.009. URL http://www.ncbi.nlm.nih.gov/pubmed/15450225.

Mert R Sabuncu, Rahul S Desikan, Jorge Sepulcre, Boon Thye T Yeo, Hesheng Liu, Nicholas J Schmansky, Martin Reuter, Michael W Weiner, Randy L Buckner, Reisa a Sperling, and Bruce Fischl. The dynamics of cortical and hippocampal atrophy in Alzheimer disease. *Archives of neurology*, 68(8):1040–8, August 2011. ISSN 1538-3687. doi: 10.1001/archneurol.2011.167. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3248949&tool=pmcentrez&rendertype=abstract.

W B Scoville and B Milner. Loss of recent memory after bilateral hippocampal lesions. 1957. *The Journal of neuropsychiatry and clinical neurosciences*, 12(1):103–113, 2000. URL http://www.ncbi.nlm.nih.gov/pubmed/10678523.

J G Sled, a P Zijdenbos, and a C Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, 17(1):87–97, February 1998. ISSN 0278-0062. doi: 10.1109/42.668698. URL http://www.ncbi.nlm.nih.gov/pubmed/9617910.

C Studholme, E Novotny, I G Zubal, and J S Duncan. Estimating tissue deformation between functional images induced by intracranial electrode implantation using anatomical MRI. *NeuroImage*, 13(4):561–76, April 2001. ISSN 1053-8119. doi: 10.1006/nimg.2000.0692. URL http://www.ncbi.nlm.nih.gov/pubmed/11305886.

Robin Wolz, Paul Aljabar, Joseph V Hajnal, Alexander Hammers, and Daniel Rueckert. LEAP: learning embeddings for atlas propagation. *NeuroImage*, 49(2):1316–25, January 2010. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2009.09.069. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3068618&tool=pmcentrez&rendertype=abstract.

Bradley T Wyman, Danielle J Harvey, Karen Crawford, Matt A Bernstein, Owen Carmichael, Patricia E Cole, Paul K Crane, Charles Decarli, Nick C Fox, Jeffrey L Gunter, Derek Hill, Ronald J Killiany, Chahin Pachai, Adam J Schwarz, Norbert Schuff, Matthew L Senjem, Joyce Suhy, Paul M Thompson, Michael Weiner, and Clifford R Jack. Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, October 2012. ISSN 1552-5279. doi: 10.1016/j.jalz.2012.06.004. URL http://www.ncbi.nlm.nih.gov/pubmed/23110865.

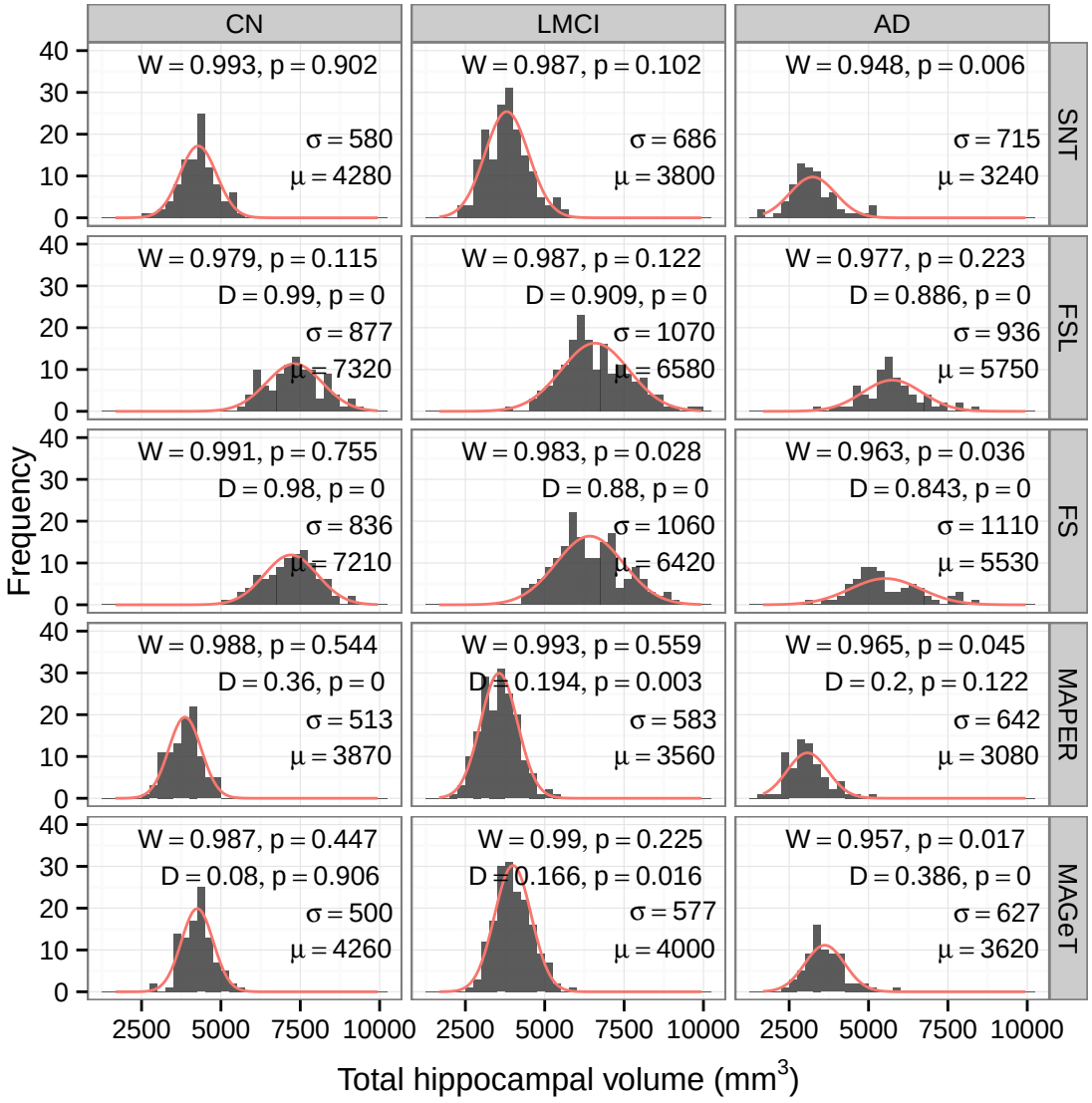

Figure 6: **ADNI Baseline cohort.** Comparison of total hippocampal volumes as measured by SNT, FSL, FreeSurfer (FS), MAPER, and MAGeT-Brain (MAGeT). A fitted normal curve is shown in red. *W* is the Shapiro-Wilkes test statistic measuring normality of the data (signficance indicates non-normality). *D* is the Kolmogorov-Smirnov test statistic measuring the goodness-of-fit between the distribution of measured volumes and SNT volumes (significance indicates a difference).

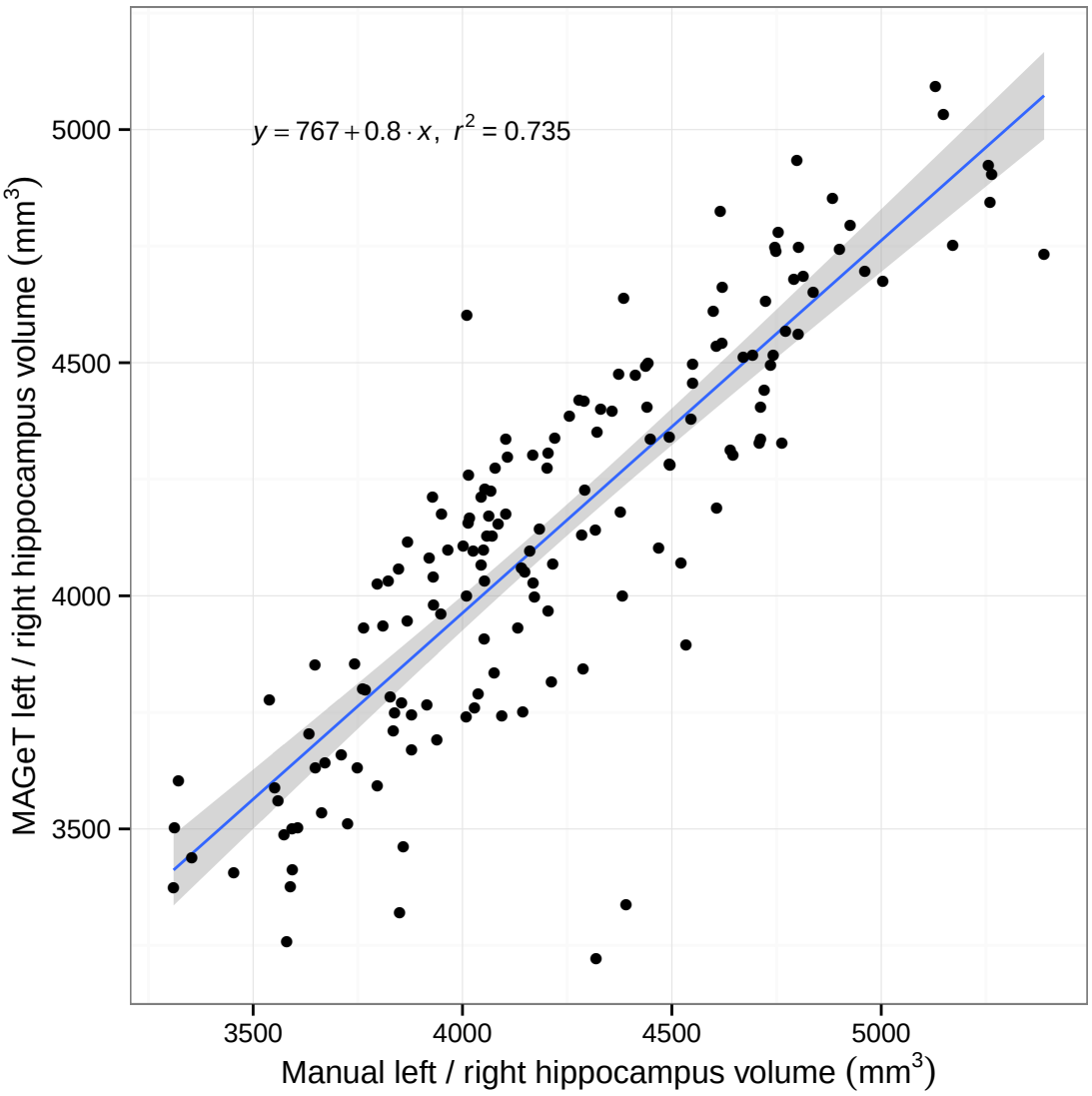

Figure 7: **First Episode Schizophrenic Patients.** Comparison of total HC volumes for MAGeT-Brain against manually rated Hippocampal volumes

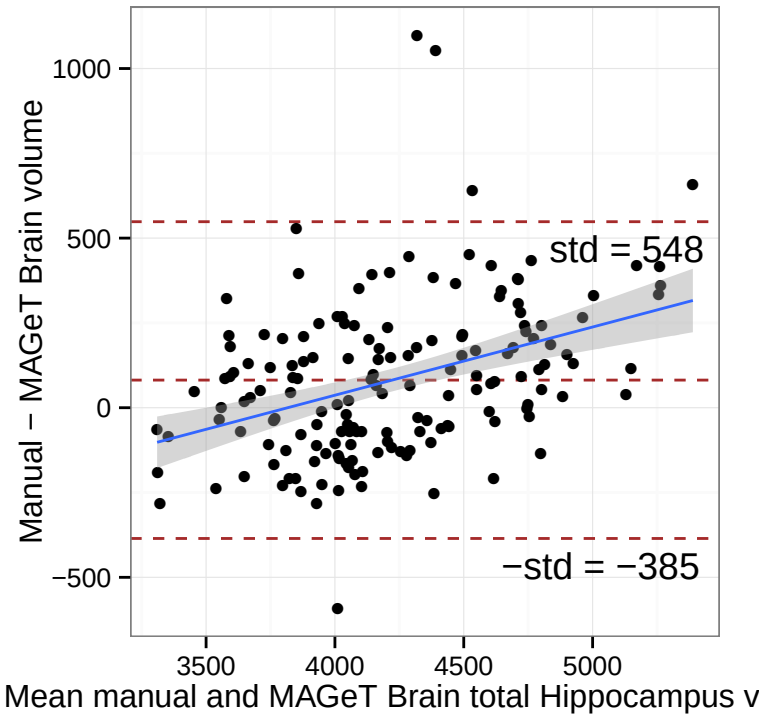

Figure 8: Bland-Altman plot comparing manual and MAGeT-Brain hippocampal volumes when using five Winterburn atlases, and a 21 image template library.

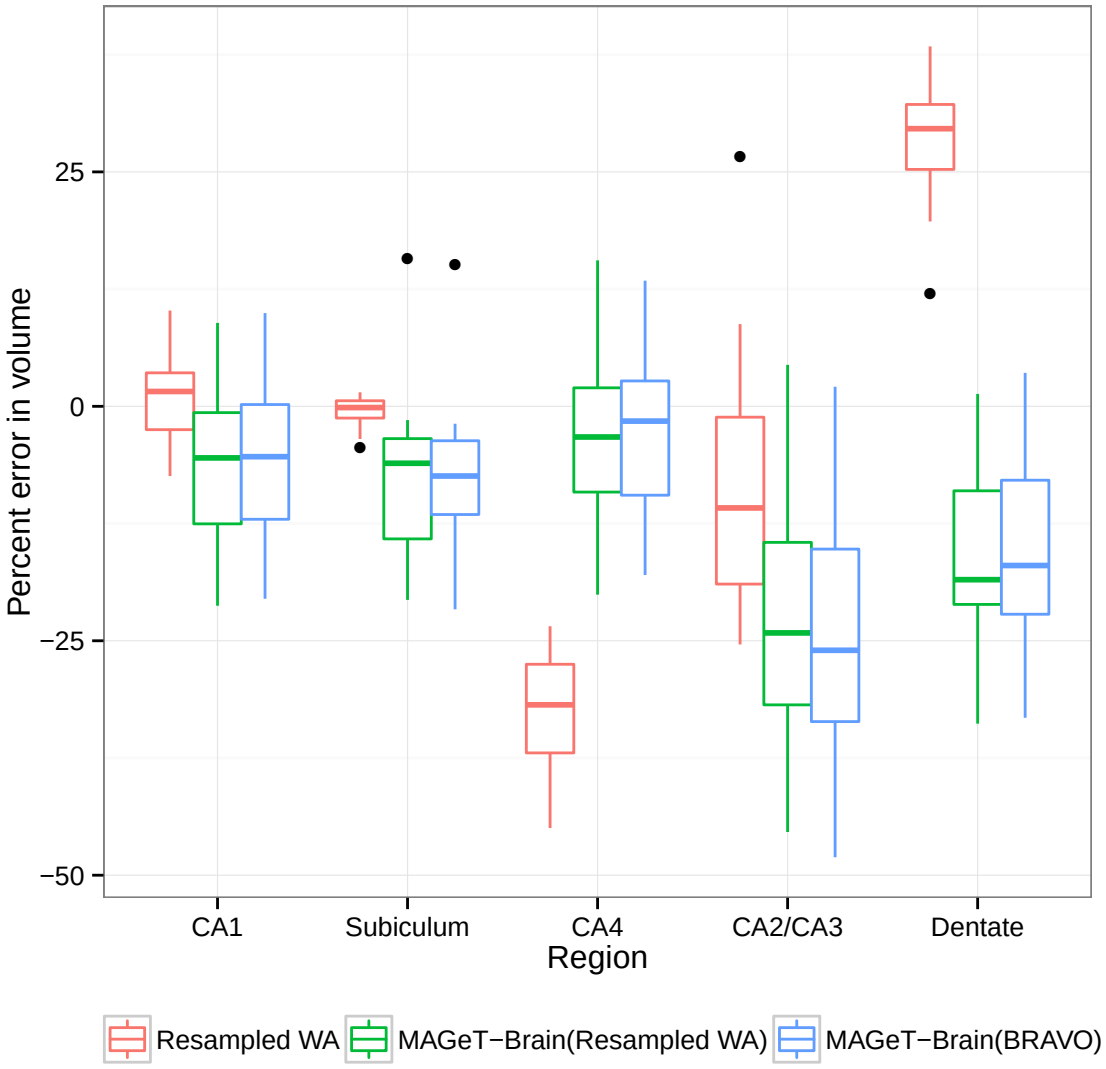

Figure 9: **Percent error in segmentation volume by hippocampus subregion.** *Resampled WA* are volumes of the manual segmentations of the Winterburn atlases after resampling to $0.9mm^3$. *MAGeT-Brain (Resampled WA)* volumes are MAGeT-Brain segmentations of the Winterburn atlas images after resampling to $0.9mm^3$ voxels. *MAGeT-Brain (BRAVO) MAGeT-Brain* volumes are MAGeT-Brain segmentations of T1 BRAVO images ($0.9mm^3$ voxels) acquired separately of four of the five Winterburn atlas subjects. Percent error is measured against the volumes of of the unmodified Winterburn atlas segmentations.