frontiers in Neuroscience

TECHNOLOGY REPORT
published: 14 July 2015
doi: 10.3389/fnins.2015.00242

Preliminary analysis using multi-atlas labeling algorithms for tracing longitudinal change

Regina E. Y. Kim [1]*, Spencer Lourens [2], Jeffrey D. Long [1, 2], Jane S. Paulsen [1, 3, 4] and Hans J. Johnson [1, 5, 6]

[1] Department of Psychiatry, University of Iowa, Iowa City, IA, USA, [2] Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA, USA, [3] Department of Neurology, Carver College of Medicine, University of Iowa, Iowa City, IA, USA, [4] Neuroscience, Carver College of Medicine, University of Iowa, Iowa City, IA, USA, [5] Department of Electrical Engineering, University of Iowa, Iowa City, IA, USA, [6] Biomedical Engineering, University of Iowa, Iowa City, IA, USA

OPEN ACCESS

Edited by:
Kevin J. Black,
Washington University School of Medicine, USA

Reviewed by:
Jessica A. Church-Lang,
The University of Texas at Austin, USA
Mert Sabuncu,
Harvard Medical School, USA

***Correspondence:**
Regina E. Y. Kim,
Department of Psychiatry, University of Iowa, 200 Hawkins Dr., Iowa city, IA 52242, USA
eunyoung-kim@uiowa.edu

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 24 April 2015
Accepted: 24 June 2015
Published: 14 July 2015

Citation:
Kim REY, Lourens S, Long JD, Paulsen JS and Johnson HJ (2015) Preliminary analysis using multi-atlas labeling algorithms for tracing longitudinal change.
Front. Neurosci. 9:242.
doi: 10.3389/fnins.2015.00242

Multicenter longitudinal neuroimaging has great potential to provide efficient and consistent biomarkers for research of neurodegenerative diseases and aging. In rare disease studies it is of primary importance to have a reliable tool that performs consistently for data from many different collection sites to increase study power. A multi-atlas labeling algorithm is a powerful brain image segmentation approach that is becoming increasingly popular in image processing. The present study examined the performance of multi-atlas labeling tools for subcortical identification using two types of *in-vivo* image database: Traveling Human Phantom (THP) and PREDICT-HD. We compared the accuracy (Dice Similarity Coefficient; DSC and intraclass correlation; ICC), multicenter reliability (Coefficient of Variance; CV), and longitudinal reliability (volume trajectory smoothness and Akaike Information Criterion; AIC) of three automated segmentation approaches: two multi-atlas labeling tools, MABMIS and MALF, and a machine-learning-based tool, BRAINSCut. In general, MALF showed the best performance (higher DSC, ICC, lower CV, AIC, and smoother trajectory) with a couple of exceptions. First, the results of accumben, where BRAINSCut showed higher reliability, were still premature to discuss their reliability levels since their validity is still in doubt (DSC < 0.7, ICC < 0.7). For caudate, BRAINSCut presented slightly better accuracy while MALF showed significantly smoother longitudinal trajectory. We discuss advantages and limitations of these performance variations and conclude that improved segmentation quality can be achieved using multi-atlas labeling methods. While multi-atlas labeling methods are likely to help improve overall segmentation quality, caution has to be taken when one chooses an approach, as our results suggest that segmentation outcome can vary depending on research interest.

Keywords: brain MRI, longitudinal data analysis, multicenter study, machine learning, multi-atlas label fusion, validation

Introduction

Brain MRI analysis from longitudinal multicenter studies has become increasingly important in clinical studies of normal aging as well as in neurodegenerative disorders, such as Huntington (HD), Alzheimer's, and Parkinson's disease. Precise assessment of longitudinal changes of brain structures may provide a non-invasive means to monitor treatment effects of clinical intervention. During the last decade, many large-scale multicenter longitudinal studies have collected series of imaging data (Jack et al., 2008; Paulsen et al., 2008; Tabrizi et al., 2014) to understand how the human brain changes in the course of aging and disease progression. These studies of structural brain changes have provided a key insight into healthy development (Sullivan et al., 2011; Treit et al., 2013; Herting et al., 2014), normal aging (Tang et al., 2001; Resnick et al., 2003; Scahill et al., 2003; Mungas et al., 2005; Risacher et al., 2010), and disease progression (Ahdidan et al., 2011; Tabrizi et al., 2012; Takahashi et al., 2012; Weiner et al., 2012).

We have utilized two independent data sets to compare performance of automated segmentation methods in human MRI. The PREDICT-HD database provides multi-center longitudinal data collected for pre-symptomatic gene-positive HD (Pre-HD) individuals over a 10-year period (Paulsen et al., 2006). Traveling Human Phantom (THP) data was collected for the multicenter reliability study and includes repeated multi-modal MRIs (T1-weighted and T2-weighted) from same five healthy subjects at eight different sites that had either a Siemens 3T TIM Trio scanner or a Philips 3T Achieva scanner (Magnotta et al., 2012). The THP data provides valuable reproducibility insights since the same five individuals traveled to the eight different sites in a month, where no brain change was expected. In this study, PREDICT-HD data was used to compare segmentation accuracy and longitudinal reliability; and THP was used to investigate intra-subject multicenter reliability. We further limited our attention to subcortical structures, which are the main regions of interest in HD. The subcortical structures of interest include the accumben nucleus, caudate nucleus, putamen, globus pallidus, thalamus, and hippocampus. Volume changes in the basal ganglia have been repeatedly reported in several studies in Pre-HD subjects (Paulsen et al., 2014a,b) and HD patients (Rosas et al., 2011; Tabrizi et al., 2013) and are primarily considered part of the hypothesized main brain target region in HD pathology (Phillips et al., 2014).

The interpretation of volumetric change is greatly affected by the quality of the segmentation approach under consideration. It is easily agreeable that large-scale longitudinal data are potentially more powerful but also more prone to methodological bias. Therefore, volume measures for longitudinally collected MRIs are required to accurately capture how individual differences are related to brain structural changes over time. While the large-scale multicenter longitudinal design is increasingly popular, one main factor that limits the sensitivity of multicenter longitudinal studies is data variation. Variation in MRI-driven volumetric measures may result from biological differences between subjects as well as from image characteristic differences, such as intensity profiles, which depend on scanner types, acquisition protocols, field strength, and subject placement in the scanner. The challenge remains in providing segmentation techniques that work in all cases, regardless of type of scanner, progression of disease, or MRI protocol to identify biomarkers which model disease progression and to predict clinical outcomes.

A desire to attain precise and sensitive measurements of MRI-driven volume changes has spurred the creation of several MRI segmentation methods (Balafar et al., 2010). This continual development effort achieved reasonable cross sectional brain anatomy segmentation by adopting a single atlas-based labeling method (Cabezas et al., 2011), which identifies regions of interests by propagating atlas information to a target subject using an image registration technique. This single atlas-based approach requires that the brain morphology presented in the image be very similar between the target subject and the atlas MRI. The single atlas-driven approach becomes vulnerable when inter-image (subject) differences are too large to be captured by any given registration method. The issue becomes increasingly problematic for large-scale data or analysis, where data variation is inherently large while the method solely relies on one atlas.

To overcome the above-mentioned issues of a single atlas-based approach, a multi-atlas labeling approach has been proposed (Rohlfing and Maurer, 2004). The multi-atlas labeling method employs several different atlases to cover a variety of MR data characteristics. Utilization of multiple atlases in a segmentation method takes account for image profile differences and large intersubject anatomical variation that naturally occurs in the human brain (Cabezas et al., 2011). There is also increasing evidence that multi-atlas labeling improves segmentation accuracy in several studies (Wang et al., 2012; Chakravarty et al., 2013; Sjoberg and Ahnesjo, 2013), and consequently, this approach is rapidly gaining popularity (Sabuncu et al., 2010; Zhang et al., 2011a; Jimenez del Toro and Muller, 2014).

The present study was designed to systematically contrast three methods for subcortical segmentation on identical data sets. Three publicly available open-source segmentation tools are utilized in this comparative study: BRAINSCut (Kim et al., 2014), ANTs MALF (Wang et al., 2012), and 3DSlicer's MABMIS (Jia et al., 2012) (Also see **Table 1**). We are specifically interested in how tools perform on *in-vivo* human MRI studies with respect to at least three aspects: segmentation accuracy, multi-center reliability, and longitudinal reliability. To capitalize on our knowledge and experience in automated segmentation tools, we have evaluated our in-house tool, BRAINSCut, in addition to two distinct techniques that are based on the multi-atlas labeling approach: MABMIS and MALF.

BRAINSCut is an open-source machine-learning-based segmentation software targeted for processing of multicenter large-scale MRI. The core of the segmentation algorithm implements a machine-learning technique called random-forest to delineate target structures. BRAINSCut excels in processing large-scale multicenter data reliably and efficiently, and has been used extensively by the PREDICT-HD (Paulsen et al., 2013) and TRACK-ON (Tabrizi et al., 2009) research teams. The latest version of BRAINSCut was evaluated using both PREDICT-HD

TABLE 1 | A brief summary of three automated segmentation tools investigated in this study: MALF (Wang and Yushkevich, 2013), MABMIS (Jia et al., 2012), and BRAINS Cut (Kim et al., 2014).

Tool	General approach	Remark
MALF	Multi-atlas labeling based	Joint fusion algorithm with Advanced SyN-based registration (Avants et al., 2008)
MABMIS	Multi-atlas labeling based	Expedite multiple registrations using a tree-based group-wise registration method and naïve label voting
BRAINS Cut	Machine-learning based	Machine-learning, specifically a Random-forest-based method, which outperformed other classification methods in multicenter large-scale MRI processing in terms of segmentation accuracy and generalizability of large scale data

and TRACK-ON data to assess its accuracy and multi-center reliability (Kim et al., 2014).

Multi-atlas based multi-image segmentation (MABMIS) (Jia et al., 2012) proposes an efficient way to expedite multiple registrations between target and atlases. MABMIS aims to address the bottleneck of multi-atlas labeling methods: computationally expensive registrations from multiple atlases to a target image. MABMIS reduces registration time by constructing a hierarchical registration tree between the atlas and target images.

Finally, multi-atlas based label fusion (MALF) provides a great implementation of the multi-atlas labeling approach in conjunction with the advanced normalization tools (ANTs) development framework. The MALF algorithm advances segmentation accuracy via weighted voting, assuming conditional independence between atlases. The approach utilizes ANTs symmetric image normalization (SyN)-based registration (Avants et al., 2008), endowing it with great potential to be a powerful tool in the field. The parameter profiles of MALF are well explained in Wang and Yushkevich (2013) and its performance is formally reported in Yushkevich et al. (2012).

This paper aims to provide validation for several aspects regarding assessment of automated segmentation performance, potentially leading to more powerful tool development in the future. Although we believe that key indicators of the quality of automated segmentation outcomes are their accuracy and reliability, only a few studies address both accuracy and reliability, and their assessment is often limited to short-term period data (Babalola et al., 2009; Wonderlick et al., 2009). Segmentation accuracy warrants the validity of the identified structures to be used for brain research as their definition corresponds to the research intent. On the other hand, reliability means the extent of measurement stability, e.g., across sites (multicenter reliability) or across time (longitudinal reliability), so that outcomes can be used to detect differences between groups or over time. Validity requires that the measurement is reliable, but the measurement can be reliable without being valid (Kimberlin and Winterstein, 2008). Therefore, we sought to investigate both aspects of segmentation quality, accuracy and reliability, in order to compare the three different approaches.

Finally, a sample size analysis was carried out to establish the minimum sample size necessary for detecting changes at the caudate nucleus and putamen level presented in PREDICT-HD MRI data with 80% power. These regions were used because they are established as the most prominent candidates for measuring longitudinal change in HD (Paulsen et al., 2014a). This sample size estimation provides crucial information to give one an idea of what to expect with the current tools available in the field as well as to guide future direction of tool development and study design.

Thus, the goal of this technical report is to provide insight into performance of different brain MRI segmentation approaches, including two emerging multi-atlas labeling techniques, focusing on *in-vivo* longitudinal multicenter MRI data. With the growing demand for a reliable segmentation technique and with attention to multi-atlas labeling methods spreading, this technical report investigated how multi-atlas labeling works on a multicenter longitudinal MRI data set. By utilizing the multicenter longitudinal data, we present quantitative and qualitative assessments of how individual trajectories of subcortical volume relate to the choice of methodology. Although this study utilized the PREDICT-HD data set, the outcomes of this study are generalizable to other study domains involving longitudinal and/or multicenter MRI studies. We hope that the validation and results in this paper will draw attention to the behavior of techniques as a useful reference to future neuroimaging studies where appropriate.

Materials and Methods

The data set used in this study is described, followed by a description of the MR image pre-processing that we applied to all our experiments. Finally, we calculate evaluation criteria used to contrast performance of MALF, MABMIS, and BRAINSCut.

Data Description

Three subsets were investigated to assess different aspects of segmentation quality resulting from MALF, MABMIS, and BRAINSCut methods: two from PREDICT-HD (Paulsen et al., 2008), and one from THP data (Magnotta et al., 2012). PREDICT-HD collected T1-weighted (and T2-weighted) MRI data at 24 sites. The sites involved in PREDICT-HD had Siemens, GE, or Phillips scanners, and some sites upgraded their scanners from 1.5 to 3.0 T during the study period. For each of the data used in this study, a summary of data is given in **Table 2** and the detailed image acquisition protocol is described elsewhere (Paulsen et al., 2008; Magnotta et al., 2012). A subset of 35 scans with manual traces (PHD35), THP multicenter data (THP), and 13 subjects of a longitudinal (L-PHD13) data set were used to assess segmentation quality in terms of accuracy, multicenter reliability, and longitudinal consistency, respectively (see **Table 2**).

TABLE 2 | Image profile for three subsets used in this study is presented.

	n	Notes	Scanner type each scan	TE (ms, T1 w)	TE (ms, T2 w)	Field strength
PHD35	35	1 scan/subject	GE (2) Siemens (28) Phillips (5)	2.804, 2.82 1.93∼3.09 3.5	88.919, 79.97 430,433 182.566∼185.971	3.0 T
THP	5	Repeated scans at 8 sites per subject	Siemens (5) Phillips (3)	(Magnotta et al., 2012)		3.0 T
L-PHD13	13	8∼10 longitudinal scans per subject	GE (49) Siemens (175) Phillips (5)	3, 5 2.87∼4.75 2.925∼3	28∼98 430∼4800 NA	1.5 T/3.0 T

The detailed image acquisition protocol is described elsewhere (PHD: Paulsen et al., 2008; THP: Magnotta et al., 2012).

PHD35

The 35-scan set of PREDICT-HD was selected to assess segmentation accuracy against manual traces. Thirty-five scans were selected by varying acquisition site as well as tissue ratio, which is generally thought to correlate with brain atrophy. Their MRI scans were manually delineated for all 12 structures of interest: the accumben nucleus, caudate nucleus, globus palladium, putamen, thalamus, and hippocampus in the left and right hemispheres.

THP

THP data from a multicenter reliability study (Magnotta et al., 2012) was incorporated into this study to compute measurement variation from MRI-driven volumetric measurements with multicenter data collection from the same subjects. The THP data provides a series of repeated scans of fives subjects at eight different sites over a short time period where biological changes would be negligible.

L-PHD13

Longitudinal reliability within subjects utilizes 13 PREDICT-HD subjects that repeatedly collected MR data for more than three time points. The subjects were also selected to include various disease burden statuses [CAG-Age Project or CAP score (Zhang et al., 2011b)], which are generally known to have different brain atrophy levels.

Image Processing

MR images were pre-processed using tools from BRAINSTools suite. Preprocessing of MR images consists of AC-PC spatial alignment (Lu, 2010; Ghayoor et al., 2013), co-registration between T1-weighted (T1-w) and T2-weighted (T2-w) images, and multimodal bias-field correction (Kim and Johnson, 2013).

The segmentation was performed on the bias-field corrected T1-w and T2-w images. The automatic segmentation tools used in this study are all publicly available and their characteristics are summarized in **Table 1**.

Evaluation

To evaluate the performance of the subcortical segmentation results, we analyzed their accuracy, multicenter reliability, and longitudinal reliability using three sets of *in-vivo* MRI data.

Segmentation accuracy in this study is a measure of how similar automated segmentation is compared to manual segmentation (the *de facto* gold standard). Using a 10-fold cross-validation approach, Dice Similarity Coefficient (DSC), and intraclass correlation (ICC) were computed by contrasting automated segmentation against manual traces. For 10-fold cross-validation, 35 subjects are roughly subdivided into 10 subsets (three or four subjects per set) and cross-validation is conducted to estimate accurate segmentation performance (more details available in the Supplemental Materials). DSC is a measure of how much two segmentations overlap in volume, and a higher DSC indicates a better correspondence in volume between two raters. ICC measures a correlation between two independent approaches on a series of data, and the approaches are generally accepted as equivalent if the ICC is higher than 0.75 (Shrout and Fleiss, 1979). A higher DSC and ICC together indicate better segmentation accuracy when compared to the gold standard.

To assess multicenter reliability, the coefficient of variation (CV) was calculated as: CV% = (SD volume/Mean volume) * 100. The CV was calculated from these THP scans. Note that the CV does not measure the correctness of segmentation, only the variability of segmentation algorithm across sites within subjects. CV compares the variability of a measurement to it's mean, which gives a much better idea of the signal than assessing each alone (large variability along with a large mean is not as worrisome as large variability with a small mean).

We attempted to quantify the longitudinal reliability using Akaike Information Criterion (AIC) from the restricted maximum likelihood (REML) approach assuming linear changes in subcortical volumes, if they exist, in the course of disease progression. As mentioned in DeShon et al. (1998), longitudinal data present many challenges for analyses and, in particular, the estimation of reliability. Thus, we also reviewed a visualization of the trajectory of each volume to ensure that the AIC, as a longitudinal reliability estimate, does not bias the interpretation of our results in any undesirable direction; the AIC can only be considered as a measure of the longitudinal reliability of the tool when the approach presents valid segmentation (higher DSC and ICC) and smooth trajectory on the plot.

Results

Segmentation Accuracy

The three methods explored differed for segmentation accuracy as measured by DSC (**Figure 1**) and ICC (**Figure 2**) against manual segmentation. For all subcortical structures, BRAINSCut and MALF presented higher DSC and ICC than MABMIS. Furthermore, MALF had generally higher DSC and ICC than BRAINSCut.

Reliability Across Centers

The multicenter reliability investigation is summarized in **Figure 3**. Segmentations from BRANSCut and MALF had lower CV values than those obtained from MABMIS in all subcortical structures (**Figure 3**). This clearly indicated higher reliability for BRAINSCut and MALF. BRAINSCut presented lower CV values for hippocampus segmentation, while for all other regions, MALF presented lower CV values than BRAINSCut. An examination of CV revealed a significant improvement of multicenter reliability when using MALF rather than BRAINSCut or MABMIS.

Longitudinal Reliability within Subjects across Disease Burden

Longitudinal segmentation reliability is contrasted and summarized in **Figure 4** and **Table 3**. MALF showed a significantly smoother trajectory (**Figure 4**) and the smallest AIC (**Table 3**) (The smaller AIC, the better fit of the model) compared to BRAINSCut and MABMIS in Pre-HD subjects. Except for the accumben nucleus and hippocampus, MALF presented very stable trajectories, as shown in **Figure 4** and a Supplemental Figure.

For the caudate and putamen, which are of foremost importance in HD study (Paulsen et al., 2014a), **a sample size analysis** was conducted in order to determine the minimum sample size necessary for detecting (with 80% power) slopes equal in magnitude to those observed in the pilot data, L-PHd13, at the 0.05 significance level. The analysis was conducted only for MALF and BRAINSCut, because estimated annual

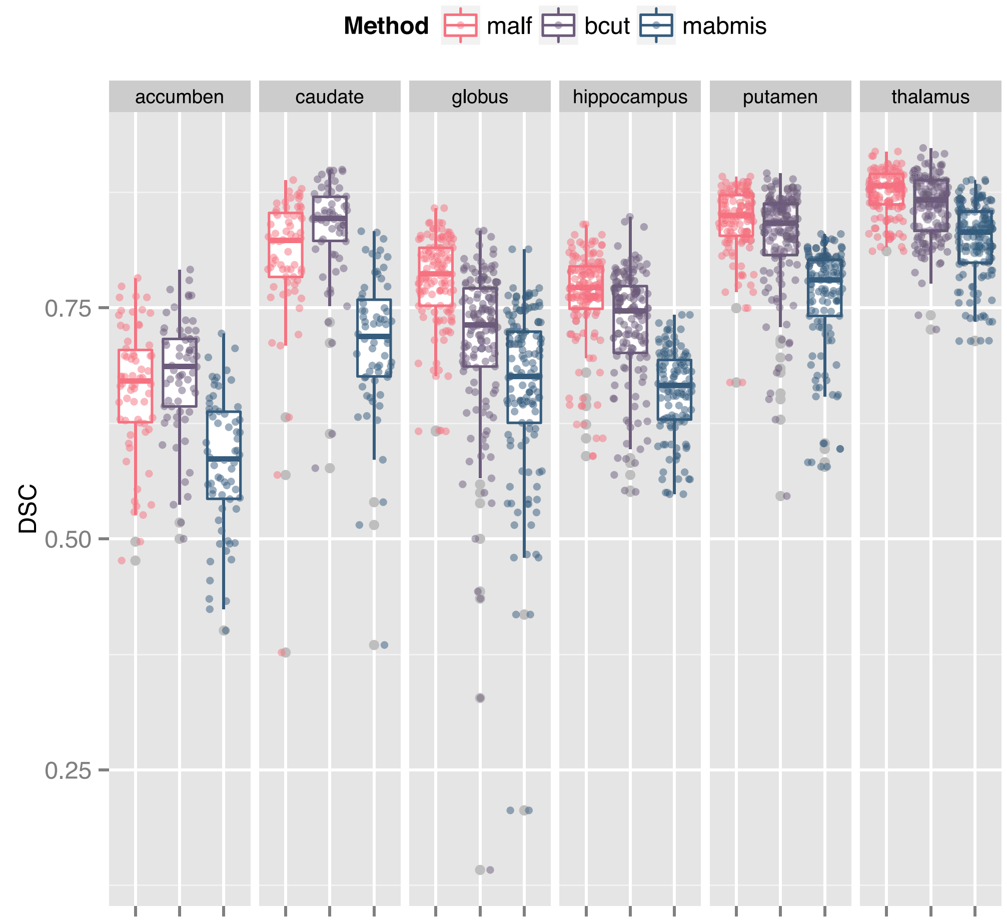

FIGURE 1 | Comparison of the Dice Similarity Coefficient (DSC) of three methods to the manual traces. A higher DSC indicates better accuracy. For all six subcortical stuctures, BRAINSCut, and MALF presented higher DSC values than MABMIS. Other than accumbens and caudates, MALF had higher DSC values than BRAINSCut.

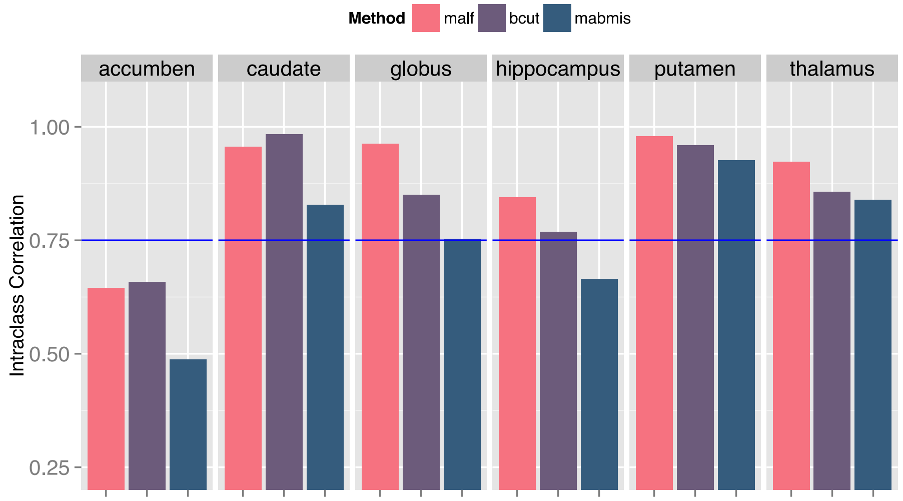

FIGURE 2 | Comparison of the intraclass correlation (ICC) of three methods to the manual traces. A higher ICC means better correspondence with the manaul traces and, therefore, improved accuracy. ICC = 0.75 (blue line) was a suggested bound by Shrout et al. (Shrout and Fleiss, 1979) for two independent measruements to be equivalent.

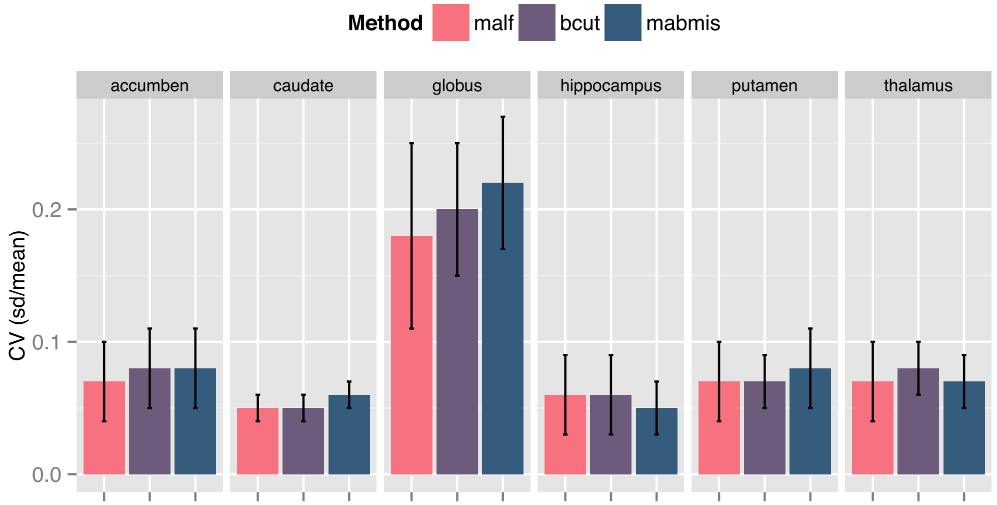

FIGURE 3 | A comparison of the Coefficient of Variation (CV) for the segmentation approaches from traveling human phantom (THP) data. MALF presented the lowest CV, the better reliability across centers, in general. BRAINSCut (bcut) outperformed only for hippocampus in terms of multicenter reliability measured in CV.

change from MABMIS was positive in the pilot data. This highly suspicious result implied that volumes of putamen and caudate increased over time, which contradicted previous findings and underlying scientific consensus regarding HD. In addition, the low segmentation accuracy (Result Segmentation Accuracy) and multicenter reliability (Result Reliability across Centers) of MABMIS indicate that its segmentation outcomes are less valid for further analysis, when compared to MALF and BRAINSCut.

For each region/method of interest, the simulation proceeded as follows: intercepts and slopes (change over time) were estimated for each region/method using linear mixed models (LMMs). LMMs were used because subjects had multiple measurements, leading to correlated data within subjects. The estimated intercepts and slopes were then used to generate a data set with a fixed sample size while assuming the data follow a linear mixed model (Psioda, 2012). This means that we used the pilot study estimates as population parameters, and is analogous to sample size analyses utilized in simpler settings, i.e., assessing differences in means at one observation time, where pilot data are used to infer a value for the variance (σ^2). Next, an approximate Z-test was employed in order to test the null hypothesis $\mathbf{H_O}$:

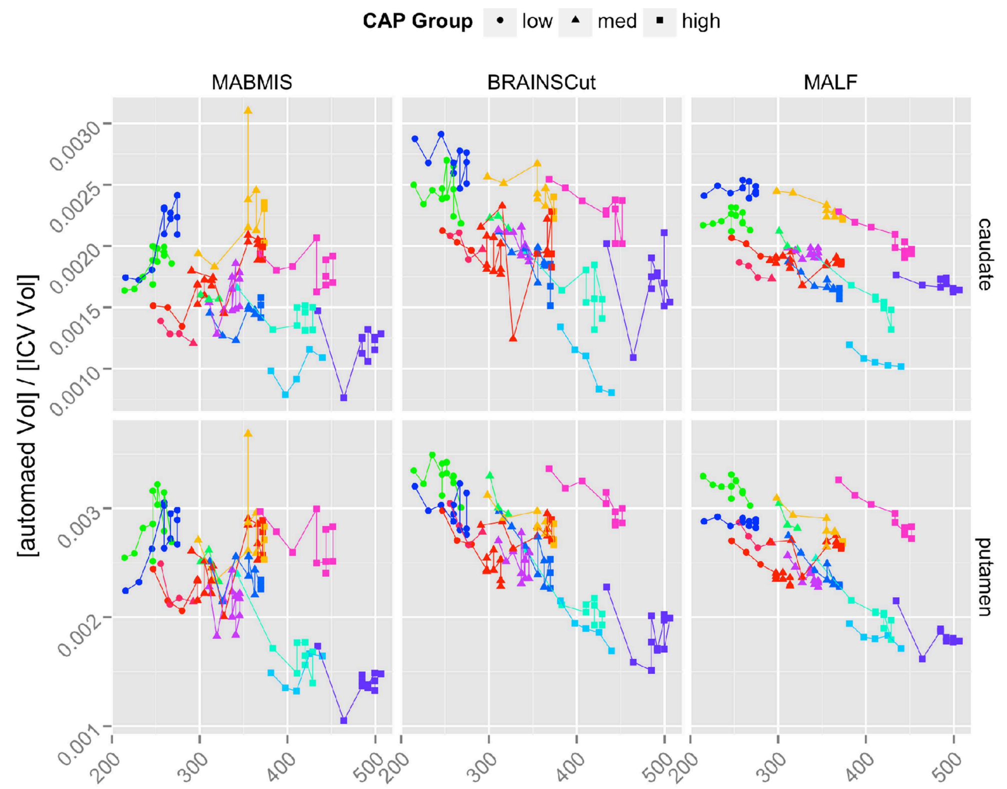

FIGURE 4 | The longitudinal trajectories of caudate and putamen volumes in left hemisphere according to Low (Circle), Medium (Triangle), and High (Square) CAP groups. Each line segment represents one of 13 longitudinal subjects. The results of all six subcortical structures are also provided in the Supplemental Figure.

TABLE 3 | The table presents comparisons of the Akaike Information Criterion (AIC) and log-likelihoods for the longitudinal model using three different methods.

ROI (hemisphere)		AIC			Log Likelihood		
		BRAINSCut	MALF	MABMIS	BRAINSCut	MALF	MABMIS
Accumben	(L)	2320.20	*2204.82	2542.02	−1151.10	−1093.41	−1262.01
	(R)	*2124.47	2180.40	2535.45	−1053.24	−1081.20	−1258.73
Caudate	(L)	3297.80	*2863.82	3336.75	−1639.90	−1422.91	−1659.38
	(R)	3252.29	*2867.40	3251.63	−1617.15	−1424.70	−1616.81
Globus	(L)	2958.17	*2634.54	3134.22	−1470.08	−1308.27	−1558.11
	(R)	2806.80	*2614.14	2992.29	−1394.40	−1298.07	−1487.14
Hippocampus	(L)	2837.26	2798.11	*2756.88	−1409.63	−1390.05	−1369.44
	(R)	2769.51	*2646.68	2786.49	−1375.75	−1314.34	−1384.25
Putamen	(L)	3195.97	*2982.74	3453.81	−1588.99	−1482.37	−1717.90
	(R)	3123.95	*2945.32	3451.46	−1552.97	−1463.66	−1716.73
Thalamus	(L)	3152.77	*3100.40	3383.60	−1567.38	−1541.20	−1682.80
	(R)	3122.50	*3055.30	3481.18	−1552.25	−1518.65	−1731.59

The minimum AICattained by the preferred model, was achieved for the structure by the approach marked with.*

slope equals zero. This process was repeated M times, and the percentage of cases for which the null hypothesis was rejected was calculated, i.e., R/M, where R was the number of simulated samples for which the null was rejected. The number M used for all analyses was 1000, and this process was repeated for a sequence of sample sizes that increased in magnitude, i.e., 10, 20, 30, etc. In order to adjust for missing data, the minimum sample size with adequate power was inflated, assuming 10% missing data as shown in **Tables 4**, **5**. In specific, the final sample size necessary was $N_1 = \frac{1}{0.9}N_0$, where N_0 was the minimum

TABLE 4 | Estimated intercept and slope for the power analysis and sample size estimation for caudate and putamen in left and right hemisphere.

Method	Side	Structure	Intercept *[V_P=V_{ROI}/ICV (%)]*	Slope *[V_P/Year]*
MALF	Left	Putamen	0.2748	−0.0026
		Caudate	0.1989	−0.0014
	Right	Putamen	0.2672	−0.0023
		Caudate	0.1992	−0.0014
BRAINSCut	Left	Putamen	0.2894	−0.0038
		Caudate	0.2123	−0.0029
	Right	Putamen	0.2767	−0.0031
		Caudate	0.2204	−0.0041

The power and sample size analysis performed on data expressed as a percent of total intercranial volume along years offollow-up. Note that slope magnitude from BRAINSCut is larger than that estimated by MALF. These slopes and intercepts were utilized as population parameters in the sample size estimation with 80% power. (V_P, Percent volume of total ICV; V_{ROI}, Volume of region of interest; ICV, intercranial volume).

TABLE 5 | Sample size estimation to attain 80% power for caudate and putamen is reported.

Required sample size *N*	Putamen		Caudate	
	Left	Right	Left	Right
MALF	45	56	62	84
BRAINSCut	45	62	123	62

*The slope and intercept from **Table 4** were used. Numbers are inflated from the minimum required sample size found in thepower analysis that is reported in the supplement.*

uninflated sample size with adequate power, and N_1 was the sample size adjusted for missing data.

The left and right hemispheres were analyzed separately, under both the MALF and BRAINSCut segmentation methods. The following table summarizes the necessary sample size for detecting differences observed in the pilot data with at least 80% power.

Discussion

We sought to characterize the performance of three different automated segmentation tools using *in-vivo* MRI data focusing on the potential of multi-atlas labeling approaches for use in large-scale multicenter longitudinal studies. Segmentation accuracy, multicenter reliability, and longitudinal reliability were investigated. Several major findings emerged: (1) multi-atlas labeling methods can improve the segmentation outcome in general when considering segmentation accuracy, multicenter reliability, and longitudinal reliability; (2) among the three methods, MALF performed the best for most structures, followed by BRAINSCut and then MABMIS; and (3) the sample size analysis for detection of a decrease in volume of caudate and putamen serves as a useful guide for future researchers who want to assure adequate power for detecting structural volume changes. It is worth noting that MALF did not outperform the other methods for all the structures, while our results show widespread improvement of segmentation quality using MALF. We suggest that the multi-atlas labeling approach can be one of the main focuses of future studies.

We found an increase in the DSC and ICC and a decrease in the CV for most structures when using MALF, indicating that better segmentation quality can be obtained using MALF as compared to BRAINSCut or MABMIS. A higher DSC indicates greater segmentation similarity between the automated and manual methods (the gold standard) that is usually interpreted as providing better accuracy. Furthermore, ICC, which is a measure of how two independent measures resemble each other, was increased for most subcortical structures of interest by using MALF. This increase in ICC, which has been known to be sensitive to intra-method variances as well as inter-method correlation, also reflected improved measurement accuracy. CV in this study is a measure of how reliable the tools are across centers. Reduction in CV indicates better inter-center reliability, i.e., less variation between multiple measurements on the same subject acquired at five different sites. In summary, the best performance in terms of higher DSC and ICC and lower CV was achieved using MALF; BRAINSCut appeared to be better in a few cases: higher DSC and ICC for the caudate nucleus and lower CV for hippocampus. Thus, our data (higher DSC, ICC, and lower CV with MALF) suggests possible segmentation superiority of the multi-atlas labeling approach in segmenting subcortical structures from human brain MRIs.

Our result suggests that MALF benefits the segmentation outcomes the most, but the choice of methods should depend on the researchers' aims since there exists a performance variation. For example, if one considers caudate for a region of interests, the choice of methods could depend on the study design. If the study design does not involve longitudinal data collection, BRAINSCut would be a better choice because of higher segmentation accuracy, which will give more sensitive outcomes. If the study design, however, expects longitudinal data acquisition, it might be wise to use MALF as it has smoother volumetric trajectory across years (**Figure 4**) with accurate segmentation results based on DSC and ICC (>0.75). For accumben nucleus, however, it is yet premature to use the data as a volumetric measurements since the accuracy of segmentation is very low, regardless the choice of the method (DSC < 0.75 and ICC < 0.7). Please note accumben is a notoriously small structure to be identified from our 1.5 or 3 Tesla MRI, ∼480 mm^3, about a size of a bean, connecting caudate and putamen. A somewhat similar argument is valid for hippocampus. Although hippocampus showed better multicenter reliability with MABMIS (lower CV), hippocampus from MABMIS may not be valid enough to be used as measurements (low DSC and low ICC). In summary, we recommend prioritizing performance criteria to choose a proper method that best suites ones study design. There is no one method that is always best. That is, depending on the purpose of the study, researchers may choose the method gives better accuracy, multicenter or longitudinal reliability for a given indication of valid segmentation.

The advantages of multicenter collaboration in observational studies include increased generalizability of results, a larger sample size, and improved efficiency, as discussed in Sprague et al. (2009) and widely practiced for rare or hard-to-recruit cases, such as HD (Paulsen et al., 2008; Tabrizi et al., 2009), Parkinson's Disease (Spencer et al., 2005), and Alzheimer's Disease (Jack et al., 2008). In such multicenter studies, across-scanner variations might interfere with the detection of disease-specific structural abnormalities (Bendfeldt et al., 2012), thereby potentially limiting the use of group analysis collected at several centers. Our data shows an increase in multicenter reliability for most subcortical structures when using MALF.

Multicenter reliability is also related to the generalizability of the tool, i.e., how much data the tool can handle, since utilization of multiple centers almost always decreases the amount of homogeneity in the data, sometimes considerably. The *"boosting"* theory in machine learning can be used to explain our results regarding superior multicenter reliability when using MALF. According to this theory, a collection of weak learning algorithms, which independently perform only slightly better than random guesses, can be converted into a highly accurate and generalizable algorithm [a better bounded generalization error (Mannor and Meir, 2001)]. Thus, our data on multicenter reliability seems to be in line with the formation of strong learners based on several weak learners; that is, the multi-atlas labeling method, where an atlas can be analogous to a weak learner, can outperform other methods. This finding is also in agreement with our previous success in subcortical segmentation using the Random-forest method, which is also a boosting method (Kim et al., 2014).

Longitudinal reliability of multi-atlas labeling tools has only been investigated in a few imaging studies (Bernal-Rusiel et al., 2012; Reuter et al., 2012) and is also limited to short-term (<2–3 years) data. The superiority of multi-atlas labeling tools for cross-sectional brain morphology investigation has been confirmed in a few studies (Wang et al., 2012; Chakravarty et al., 2013; Wu et al., 2014). The previous studies in the literature on MRI segmentation quality using the multi-atlas labeling method generally describe accuracy improvement in terms of similarity to the manually traced gold standard. We found that a multi-atlas labeling approach can also improve the longitudinal reliability of subcortical segmentation.

In the present study, visual inspection showed that MALF presented a more stable trajectory of subcortical volumes than other methods in the data collected over 10 years (**Figure 4**). We also investigated longitudinal reliability assuming linear changes of subcortical volume in the course of disease progression in Pre-HD subjects. MALF also presented the minimum AIC, which is considered the best model fit, for all subcortical structures except for the accumben in the right hemisphere and the hippocampus in the left hemisphere (**Table 3**). Although our analysis of longitudinal performance on subcortical structures showed the excellence of MALF when assuming a linear trajectory, according to the longitudinal modeling suggestions in the literature (DeShon et al., 1998), careful attention should be paid as longitudinal data present many challenges for analysis. However, our trajectory plot in **Figure 4** (and Supplemental Figure) demonstrates the superior stability of MALF in comparison to BRAINCut and MABMIS.

It is interesting to note that estimation accuracy and sample size may seem counter-intuitive at first: MALF requires larger sample to obtain 80% power at 0.05 significant level for caudate in right hemisphere (**Table 5**) even though MALF outperformed BRAINSCut in with respect to longitudinal reliability. This is because the power analysis and sample size estimation is based on segmentation results from pilot (LPH-13) data set. For both caudate and putamen in each hemisphere, the estimated slopes from BRAINSCut were larger than those estimated by MALF (**Table 4**). This means that the MALF sample size analysis was powered for detecting a much finer change than that for the BRAINSCut and thus MALF requires slightly larger sample size to achieve 80% power.

Before drawing conclusions, some limitations of the present study must be acknowledged. First, more implementations of the multi-atlas labeling method have to be incorporated to investigate which is the best method for subcortical segmentation. Second, the registration technique used in each segmentation tool can be investigated for better performance. Third, studies using other psychiatric conditions are required to generalize our findings beyond HD. This will allow for a better understanding of multi-atlas labeling approach behavior for the automated processing of human MRIs.

In conclusion, we have presented evidence that multi-atlas labeling methods, which fall under an emerging segmentation approach in the field, can improve segmentation quality in terms of accuracy and reliability. Other methods can also be useful, depending on the regions of interest, study design, and implementation. However, it is likely multi-atlas labeling helps to improve the overall quality of segmentation outcomes.

Acknowledgments

This paper is funded by multiple grants: Huntington's Disease Society of America (Human Biology Project Fellowship), BRAINS (R01 NS050568), Validation of Structural/Functional MRI Localization (R01 EB000975), 3D Shape Analysis for Computational Anatomy (R01 EB008171), Neurobiological Predictors of HD (R01 NS040068), Cognitive and Functional Brain Changes in Preclinical HD (R01 NS054893), Algorithms for Functional and Anatomical Brain Analysis (P41 RR015241), Enterprise Storage in a Collaborative Neuroimaging Environment (S10 RR023392), Core 2b HD (U54 EB005149), and NIPYPE (R03 EB008673).

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnins.2015.00242

References

Ahdidan, J., Hviid, L. B., Chakravarty, M. M., Ravnkilde, B., Rosenberg, R., Rodell, A., et al. (2011). Longitudinal MR study of brain structure and hippocampus volume in major depressive disorder. *Acta Psychiatr. Scand.* 123, 211–219. doi: 10.1111/j.1600-0447.2010.01644.x

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi: 10.1016/j.media.2007.06.004

Babalola, K. O., Patenaude, B., Aljabar, P., Schnabel, J., Kennedy, D., Crum, W., et al. (2009). An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *Neuroimage* 47, 1435–1447. doi: 10.1016/j.neuroimage.2009.05.029

Balafar, M. A., Ramli, A. R., Saripan, M. I., and Mashohor, S. (2010). Review of brain MRI image segmentation methods. *Artif. Intell. Rev.* 33, 261–274. doi: 10.1007/s10462-010-9155-0

Bendfeldt, K., Hofstetter, L., Kuster, P., Traud, S., Mueller-Lenke, N., Naegelin, Y., et al. (2012). Longitudinal gray matter changes in multiple sclerosis–differential scanner and overall disease-related effects. *Hum. Brain Mapp.* 33, 1225–1245. doi: 10.1002/hbm.21279

Bernal-Rusiel, J. L., Greve, D. N., Reuter, M., Fischl, B., and Sabuncu, M. R. (2012). Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *Neuroimage* 66C, 249–260. doi: 10.1016/j.neuroimage.2012.10.065

Cabezas, M., Oliver, A., Llado, X., Freixenet, J., and Cuadra, M. B. (2011). A review of atlas-based segmentation for magnetic resonance brain images. *Comput. Methods Programs Biomed.* 104, e158–e177. doi: 10.1016/j.cmpb.2011.07.015

Chakravarty, M. M., Steadman, P., van Eede, M. C., Calcott, R. D., Gu, V., Shaw, P., et al. (2013). Performing label-fusion-based segmentation using multiple automatically generated templates. *Hum. Brain Mapp.* 34, 2635–2654. doi: 10.1002/hbm.22092

DeShon, R. P., Ployhart, R. E., and Sacco, J. M. (1998). The estimation of reliability in longitudinal models. *Int. J. Behav. Dev.* 22, 493–515. doi: 10.1080/016502598384243

Ghayoor, A., Vaidya, J. G., and Johnson, H. J. (2013). Development of a novel constellation based landmark detection algorithm. *SPIE Med. Imaging* 8669, 86693F-6. doi: 10.1117/12.2006471

Herting, M. M., Gautam, P., Spielberg, J. M., Kan, E., Dahl, R. E., and Sowell, E. R. (2014). The role of testosterone and estradiol in brain volume changes across adolescence: a longitudinal structural MRI study. *Hum. Brain Mapp.* 35, 5633–5645. doi: 10.1002/hbm.22575

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., et al. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691. doi: 10.1002/jmri.21049

Jia, H., Yap, P.-T., and Shen, D. (2012). Iterative multi-atlas-based multi-image segmentation with tree-based registration. *Neuroimage* 59, 422–430. doi: 10.1016/j.neuroimage.2011.07.036

Jimenez del Toro, O. A., and Muller, H. (2014). "Multi atlas-based segmentation with data driven refinement," in *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on* (Valencia), 605–608. doi: 10.1109/BHI.2014.6864437

Kim, E. Y., and Johnson, H. J. (2013). Robust multi-site MR data processing: iterative optimization of bias correction, tissue classification, and registration. *Front. Neuroinform.* 7:29. doi: 10.3389/fninf.2013.00029

Kim, E. Y., Magnotta, V. A., Liu, D., and Johnson, H. J. (2014). Stable Atlas-based Mapped Prior (STAMP) machine-learning segmentation for multicenter large-scale MRI data. *Magn. Reson. Imaging* 32, 832–844. doi: 10.1016/j.mri.2014.04.016

Kimberlin, C. L., and Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *Am. J. Health. Syst. Pharm.* 65, 2276–2284. doi: 10.2146/ajhp070364

Lu, W. (2010). A *Method* for *Automated Landmark Constellation Detection Using Evolutionary Principal Components and Statistical Shape Models*. Master of Science thesis, University of Iowa.

Magnotta, V. A., Matsui, J. T., Liu, D., Johnson, H. J., Long, J. D., Bolster, B. D. Jr., et al. (2012). Multicenter reliability of diffusion tensor imaging. *Brain Connect.* 2, 345–355. doi: 10.1089/brain.2012.0112

Mannor, S., and Meir, R. (2001). "Weak learners and improved rates of convergence in boosting," in *Advances in Neural Information Processing Systems 13: Proc. NIPS'2000* (Denver, CO), 280–286.

Mungas, D., Harvey, D., Reed, B. R., Jagust, W. J., DeCarli, C., Beckett, L., et al. (2005). Longitudinal volumetric MRI change and rate of cognitive decline. *Neurology* 65, 565–571. doi: 10.1212/01.wnl.0000172913.88973.0d

Paulsen, J. S., Langbehn, D. R., Stout, J. C., Aylward, E., Ross, C. A., Nance, M., et al. (2008). Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *J. Neurol. Neurosurg. Psychiatry* 79, 874–880. doi: 10.1136/jnnp.2007.128728

Paulsen, J. S., Long, J. D., Johnson, H. J., Aylward, E. H., Ross, C. A., Williams, J. K., et al. (2014a). Clinical and biomarker changes in premanifest Huntington disease show trial feasibility: a decade of the PREDICT-HD study. *Front. Aging Neurosci.* 6:78. doi: 10.3389/fnagi.2014.00078

Paulsen, J. S., Long, J. D., Ross, C. A., Harrington, D. L., Erwin, C. J., Williams, J. K., et al. (2014b). Prediction of manifest Huntington's disease with clinical and imaging measures: a prospective observational study. *Lancet Neurol.* 13, 1193–1201. doi: 10.1016/S1474-4422(14)70238-8

Paulsen, J. S., Magnotta, V. A., Mikos, A. E., Paulson, H. L., Penziner, E., Andreasen, N. C., et al. (2006). Brain structure in preclinical Huntington's disease. *Biol. Psychiatry* 59, 57–63. doi: 10.1016/j.biopsych.2005.06.003

Paulsen, J. S., Smith, M. M., and Long, J. D. (2013). Cognitive decline in prodromal Huntington Disease: implications for clinical trials. *J. Neurol. Neurosurg. Psychiatry* 84, 1233–1239. doi: 10.1136/jnnp-2013-305114

Phillips, O., Squitieri, F., Sanchez-Castaneda, C., Elifani, F., Griguoli, A., Maglione, V., et al. (2014). The corticospinal tract in huntington's disease. *Cereb. Cortex.* doi: 10.1093/cercor/bhu065. [Epub ahead of print].

Psioda, M. (2012). *Random Effects Simulation for Sample Size Calculations Using SAS ® Matthew Psioda.* Department of Biostatistics, The University of North Carolina at Chapel Hill (Chapel Hill, NC).

Resnick, S. M., Pham, D. L., Kraut, M. A., Zonderman, A. B., and Davatzikos, C. (2003). Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *J. Neurosci.* 23, 3295–3301. Available online at: http://www.jneurosci.org/content/23/8/3295.full.pdf+html

Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 61, 1402–1418. doi: 10.1016/j.neuroimage.2012.02.084

Risacher, S. L., Shen, L., West, J. D., Kim, S., McDonald, B. C., Beckett, L., et al. (2010). Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. *Neurobiol. Aging* 31, 1401–1418. doi: 10.1016/j.neurobiolaging.2010.04.029

Rohlfing, T., and Maurer, C. R. (2004). Multi-classifier framework for atlas-based image segmentation. *Proc. 2004 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition.* 1, 255–260. doi: 10.1109/cvpr.2004.1315040

Rosas, H. D., Reuter, M., Doros, G., Lee, S. Y., Triggs, T., Malarick, K., et al. (2011). A tale of two factors: what determines the rate of progression in Huntington's disease? A longitudinal MRI study. *Mov. Disord.* 26, 1691–1697. doi: 10.1002/mds.23762

Sabuncu, M. R., Yeo, B. T. T., Van Leemput, K., Fischl, B., and Golland, P. (2010). A generative model for image segmentation based on label fusion. *IEEE Trans. Med. Imaging* 29, 1714–1729. doi: 10.1109/TMI.2010.2050897

Scahill, R. I., Frost, C., Jenkins, R., Whitwell, J. L., Rossor, M. N., and Fox, N. C. (2003). A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. *Arch. Neurol.* 60, 989–994. doi: 10.1001/archneur.60.7.989

Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420

Sjoberg, C., and Ahnesjo, A. (2013). Multi-atlas based segmentation using probabilistic label fusion with adaptive weighting of image similarity measures. *Comput. Methods Programs Biomed.* 110, 308–319. doi: 10.1016/j.cmpb.2012.12.006

Spencer, S. S., Berg, A. T., Vickrey, B. G., Sperling, M. R., Bazil, C. W., Shinnar, S., et al. (2005). Predicting long-term seizure outcome after resective epilepsy surgery: the multicenter study. *Neurology* 65, 912–918. doi: 10.1212/01.wnl.0000176055.45774.71

Sprague, S., Matta, J. M., Bhandari, M., Dodgin, D., Clark, C. R., Kregor, P., et al. (2009). Multicenter collaboration in observational research: improving generalizability and efficiency. *J. Bone Joint Surg. Am.* 91(Suppl. 3), 80–86. doi: 10.2106/JBJS.H.01623

Sullivan, E. V., Pfefferbaum, A., Rohlfing, T., Baker, F. C., Padilla, M. L., and Colrain, I. M. (2011). Developmental change in regional brain structure over 7 months in early adolescence: comparison of approaches for longitudinal atlas-based parcellation. *Neuroimage* 57, 214–224. doi: 10.1016/j.neuroimage.2011.04.003

Tabrizi, S., Craufurd, D., Dürr, A., Fox, N., Frost, C., Johnson, H. et al. (2014). TRACK-HD. Available online at: http://www.track-hd.net.

Tabrizi, S. J., Langbehn, D. R., Leavitt, B. R., Roos, R. A., Durr, A., Craufurd, D., et al. (2009). Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurol.* 8, 791–801. doi: 10.1016/S1474-4422(09)70170-X

Tabrizi, S. J., Reilmann, R., Roos, R. A. C., Durr, A., Leavitt, B., Owen, G., et al. (2012). Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. *Lancet Neurol.* 11, 42–53. doi: 10.1016/S1474-4422(11)70263-0

Tabrizi, S. J., Scahill, R. I., Owen, G., Durr, A., Leavitt, B. R., Roos, R. A., et al. (2013). Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *Lancet Neurol.* 4422, 1–13. doi: 10.1016/s1474-4422(13)70088-7

Takahashi, T., Kido, M., Nakamura, K., Furuichi, A., Zhou, S.-Y., Kawasaki, Y., et al. (2012). Longitudinal MRI study of the pituitary volume in chronic schizophrenia: a preliminary report. *Psychiatry Res.* 202, 84–87. doi: 10.1016/j.pscychresns.2011.11.008

Tang, Y., Whitman, G. T., Lopez, I., and Baloh, R. W. (2001). Brain volume changes on longitudinal magnetic resonance imaging in normal older people. *J. Neuroimaging* 11, 393–400. doi: 10.1111/j.1552-6569.2001.tb00068.x

Treit, S., Lebel, C., Baugh, L., Rasmussen, C., Andrew, G., and Beaulieu, C. (2013). Longitudinal MRI reveals altered trajectory of brain development during childhood and adolescence in fetal alcohol spectrum disorders. *J. Neurosci.* 33, 10098–10109. doi: 10.1523/JNEUROSCI.5004-12.2013

Wang, H., Suh, J. W., Das, S. R., Pluta, J., Craige, C., and Yushkevich, P. A. (2012). Multi-Atlas Segmentation with Joint Label Fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 611–623. doi: 10.1109/TPAMI.2012.143

Wang, H., and Yushkevich, P. (2013). Multi-atlas segmentation with joint label fusion and corrective learning—an open source implementation. *Front. Neuroinform.* 7:27. doi: 10.3389/fninf.2013.00027

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2012). The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimers Dement.* 8, S1–S68. doi: 10.1016/j.jalz.2011.09.172

Wonderlick, J. S., Ziegler, D. A., Hosseini-Varnamkhasti, P., Locascio, J. J., Bakkour, A., van der Kouwe, A., et al. (2009). Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. *Neuroimage* 44, 1324–1333. doi: 10.1016/j.neuroimage.2008.10.037

Wu, G., Wang, Q., Zhang, D., Nie, F., Huang, H., and Shen, D. (2014). A generative probability model of joint label fusion for multi-atlas based brain segmentation. *Med. Image Anal.* 18, 881–890. doi: 10.1016/j.media.2013.10.013

Yushkevich, P. A., Pluta, J., and Avants, B. B. (2012). "From label fusion to correspondence fusion: a new approach to unbiased groupwise registration," in *Proceedings / CVPR, Computer Society Conference on Computer Vision and Pattern Recognition* (Providence, RI), 956–963. doi: 10.1109/CVPR.2012.6247771

Zhang, D., Wu, G., Jia, H., and Shen, D. (2011a). Confidence-guided sequential label fusion for multi-atlas based segmentation. *Med. Image Comput. Comput. Assist. Interv.* 14, 643–650. doi: 10.1007/978-3-642-23626-6_79

Zhang, Y., Long, J. D., Mills, J. A., Warner, J. H., Lu, W., and Paulsen, J. S. (2011b). Indexing disease progression at study entry with individuals at-risk for Huntington disease. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 156B, 751–763. doi: 10.1002/ajmg.b.31232

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.