Neurobiology of Aging 84 (2019) 9–16

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

Normative brain volumetry derived from different reference populations: impact on single-subject diagnostic assessment in dementia

Elisabeth J. Vinke [a,b], Wyke Huizinga [c], Martin Bergtholdt [d], Hieab H. Adams [a,b], Rebecca M.E. Steketee [a,b], Janne M. Papma [e,b], Frank J. de Jong [e,b], Wiro J. Niessen [b,c,f], M. Arfan Ikram [a], Fabian Wenzel [d], Meike W. Vernooij [a,b,*], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands
[b] Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
[c] Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
[d] Digital Imaging, Philips Research Hamburg, Hamburg, Germany
[e] Department of Neurology, Erasmus MC, Rotterdam, The Netherlands
[f] Department of Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 25 January 2019
Received in revised form 11 July 2019
Accepted 16 July 2019
Available online 23 July 2019

Keywords:
Subcortical brain volume
Normative data
MRI
Imaging

ABSTRACT

Brain imaging data are increasingly made publicly accessible, and volumetric imaging measures derived from population-based cohorts may serve as normative data for individual patient diagnostic assessment. Yet, these normative cohorts are usually not a perfect reflection of a patient's base population, nor are imaging parameters such as field strength or scanner type similar. In this proof of principle study, we assessed differences between reference curves of subcortical structure volumes of normal controls derived from two population-based studies and a case-control study. We assessed the impact of any differences on individual assessment of brain structure volumes. Percentile curves were fitted on the three healthy cohorts. Next, percentile values for these subcortical structures for individual patients from these three cohorts, 91 mild cognitive impairment and 95 Alzheimer's disease cases and patients from the Alzheimer Center, were calculated, based on the distributions of each of the three cohorts. Overall, we found that the subcortical volume normative data from these cohorts are highly interchangeable, suggesting more flexibility in clinical implementation.

1. Introduction

Methods to assist (early) diagnosis of neurological diseases and neuropsychiatric disorders in a clinical setting are of great importance. Noninvasive brain imaging, for example, with magnetic resonance imaging (MRI), is an increasingly applied diagnostic tool to detect brain pathology. To detect pathology on brain imaging, an understanding of what is normal is important, especially in diseases with a strong age-related component. A background of "normal aging" should therefore be taken into account, something that is difficult to estimate on a visual assessment alone. Many studies have focused on creating normative values of a broad spectrum of imaging markers of the human brain. By combining small to relatively large imaging data samples of healthy controls from different studies to one large imaging data set, normative values for different brain structure regions in aging were estimated and presented for clinical use (Potvin et al., 2016, 2017; Peterson et al., 2018; Rummel et al., 2018; Tutunji et al., 2018). With more and more brain imaging data from large population cohorts being publicly accessible, simply choosing a single population cohort to use as reference data would be feasible and in many (clinical) settings the most pragmatic

* Corresponding author at: Department of Epidemiology, Erasmus University Medical Center, PO Box 2040, Rotterdam, 3000 CA, The Netherlands. Tel.: +31 10 70 42006; fax: +31 10 70 446 57.
E-mail address: m.vernooij@erasmusmc.nl (M.W. Vernooij).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.neurobiolaging.2019.07.008

option. However, the ideal reference population is the base population from which that individual patient arises, but data from such a population are rarely available in the clinical setting. Although the added value for diagnostic purposes of the use of normative values on top of visual assessment alone in a clinical setting is increasingly recognized (Brewer, 2009; Ross et al., 2013, 2015; Vernooij et al., 2018), it is not known to what extent variations in reference populations may affect the individual patient comparison to reference data. Furthermore, the choice of reference population is accompanied by differences in scanner types, field strength, and acquisition parameters between normative cohorts, which could introduce variation in results obtained from automated brain segmentation methods. Regarding the latter, several studies examined the robustness of automated segmentation methods across field strengths and scanner types, which have shown that reproducible segmentations can be obtained with residual volumetric variability of a few percent (Cavedo et al., 2017; Heinen et al., 2016; Maclaren et al., 2014; Tudorascu et al., 2016; Velasco-Annis et al., 2018). Yet, even with a perfectly robust segmentation method, the question remains whether population differences in structural brain volumes may impact individual patient comparison and whether this would lead to different clinical management. Are reference populations derived from case-control studies, “healthy controls” for example, similar to reference populations derived from population-based cohorts? Or does a reference population need to be similar to the base population from which an individual patient arises? Studies using normative reference data for diagnosis of neurological diseases, such as Alzheimer's disease (AD), commonly focus on volumetric changes in cortical gray matters areas (Potvin et al., 2017; Tondelli et al., 2012). More recently, interest in the role of volume and shape of subcortical brain structures is growing as relevant (early) brain imaging markers (Kälin et al., 2017; Roh et al., 2011; Stepan-Buksakowska et al., 2014). A novel approach for subcortical brain segmentation in T1-weighted MRI brain scans was recently presented, based on a shape-constrained deformable surface model (Wenzel et al., 2018). Experiments on data both 3T and 1.5T for different scanners indicate good agreement with respect to independent ground truth segmentations of the subcortical structures using this model-based brain segmentation (MBS) approach, regardless of the field strength or vendor. In this proof-of-principle study, we assessed differences in normative reference curves for subcortical structure volumes (including hippocampal volume) segmented with the MBS method, between reference populations derived from two population-based studies and normal controls from a large case-control study. Furthermore, we assessed the impact of using these different cohorts on individuals with a higher risk of developing AD (APOE ε4 allele carriers and subjects with mild cognitive impairment (MCI)) and patients with AD.

2. Material

2.1. Reference populations

In this study, cross-sectional samples of three reference populations were used to estimate and compare the subcortical volume percentile curves. The reference populations included the Rotterdam Study, the United Kingdom Biobank (UKBB), and normal controls from the Alzheimer's Disease Neuroimaging Initiative (ADNI). These studies were approved by a research or medical ethical committee, and informed consent was obtained from all subjects. From each study, 3D T1-weighted imaging data were used for subcortical structure segmentation. In Supplemental Figure 1 the age-distribution of the healthy participants of the Rotterdam Study, ADNI and UKBB are shown.

2.1.1. Rotterdam Study

We included 895 T1-weighted scans (median age = 66.4, interquartile range (IQR) = 22.7, 504 women) from the population-based Rotterdam Study, a prospective longitudinal study among community-dwelling subjects aged 45 years and over (Ikram et al., 2017). Scans were randomly selected from the study such that the age at time of the scan was uniformly distributed within a range of 45-95 years. All brain scans were acquired on a single 1.5-Tesla MRI system (GE Healthcare, US) (Ikram et al., 2017). In total, 225 of the 895 participants were APOE ε4 carriers (25.1%).

2.1.2. Alzheimer's Disease Neuroimaging Initiative

We included 430 (median age = 74.1, IQR = 7.5, 217 women) baseline 3D T1-weighted MRI scans from healthy controls from ADNI. ADNI is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD (adni.loni.usc.edu, for up-to-date information, see www.adni-info.org). With the ADNI data set being the smallest cross-sectional data set of the three cohorts, no further selection based on age was performed, resulting in an age range between 55 and 90 years. Participants were scanned on a 1.5- (n = 231, 53.7%) or 3-Tesla (n = 199, 46.3%) MRI system from GE Medical (n = 162), Philips (n = 71), or Siemens (n = 197). In total, 114 of the 430 participants were APOE ε4 carriers (26.5%).

2.1.3. United Kingdom Biobank

We included 876 (median age = 55.0, IQR = 15.0, 428 women) 3D T1-weighted scans from UKBB, all scanned with a 3-Tesla MRI system (Siemens Healthcare, UK). Scans were randomly selected from the study such that the age at time of the scan was uniformly distributed within a range of 40-70 years. UKBB is a prospective resource gathering extensive questionnaires, physical and cognitive measures, and biological samples in a cohort of 500,000 participants (Sudlow et al., 2015). In total, 238 of the 876 participants were APOE ε4 carriers (27.2%).

2.2. Patient data for subject-specific comparison

We assessed 3D T1-weighted scans from participants with MCI and AD from the Rotterdam Study and ADNI database and a sample of the APOE ε4 allele carriers from the healthy participants from the three reference populations, to evaluate whether subject-specific percentile estimations of different participant groups (APOE ε4 allele carriership, MCI or AD) depend on the chosen reference population. Furthermore, as an independent patient data set, we included MCI and AD cases from the Alzheimer Center Erasmus MC.

2.2.1. AD and MCI cases from the Rotterdam Study and ADNI

From the Rotterdam Study, 3D T1-weighted scans were selected from study participants with MCI (n = 41, age = 72 ± 6.4, 22 women) and prevalent AD (n = 45, age = 81.9 ± 4.6, 25 women) at time of the scan. From the ADNI data set, we selected the baseline 3D T1-weighted scan from patients with MCI (n = 50, age = 75.6 ± 7.0, 19 women) and patients with AD (n = 50, age = 75.1 ± 7.7, 28 women).

2.2.2. AD and MCI cases from the Alzheimer Center Erasmus MC

Scans from patients with MCI and AD from the Alzheimer Center Erasmus MC, Rotterdam, The Netherlands, were used as an independent set. Use of clinical data from the Alzheimer Center for research purposes was approved by the local medical ethical committee. Informed consent was obtained from all patients. We used 19 3D T1-weighted scans from patients with MCI (8 women, age = 69.4 ± 5.6) and 43 3D T1-weighted scans from patients with AD (15 women, age=66.8 ± 9.6) who visited the Alzheimer Center

Erasmus MC between 2011 and 2016. All patient data were acquired on a single 1.5T MRI system (GE Healthcare, US).

2.3. Participant groups

In the rest of the article, the term "participant groups" will be used to describe the different subgroups on which the analyses are performed. The participant groups consist of the following:

- **Healthy**: healthy participants from the three reference populations (N_{total} = 2201, Rotterdam Study: 895, ADNI: 430, UKBB: 876).
- **APOE ε4 carriers**: Healthy participants from the three reference populations who carry one or two APOE ε4 allele(s) (N_{total} = 158, Rotterdam Study: 47, ADNI: 61, UKBB: 50).
- **MCI**: participants from the Rotterdam Study and ADNI data set with MCI (N_{total} = 91, Rotterdam Study: 41, ADNI: 50).
- **AD**: participants from the Rotterdam Study and ADNI data set with AD (N_{total} = 95, Rotterdam Study: 45, ADNI: 50).
- **MCI AC**: patients with MCI who visited the Alzheimer Center (N_{total} = 19).
- **AD AC**: patients with AD who visited the Alzheimer Center (N_{total} = 43).

3. Methods

3.1. Segmentation of subcortical structures on 3D T1-weighted data

This work is based on an MBS as described by Wenzel et al. (2018), utilizing a shape-constrained deformable surface model for segmentation of subcortical brain structures from T1-weighted MRI. Adaptation of subcortical brain surfaces is performed stepwise, starting with global rigid and affine adaptation and followed by multi-affine and fully deformable adaptation. In each step, a weighted sum of internal and external energy is minimized. Here, internal energy relates to deviations from a shape /point distribution model of a training data set. The external energy component is based on the triangle-specific spatial distance to a target point along its normal. Target points are estimated with boundary detector functions that have been trained via a simulated search on the same training data set. For the used version of MBS, the training data set included 96 manually delineated 3T scans, equally distributed between patients with AD and healthy controls between ages 50 and 90 year as well as three device manufacturers (Philips, Siemens, and GE). The segmentation software is optionally available as part of the IntelliSpace Discovery workstation for data analytics in medical imaging.

3.2. Percentile curve fitting

For fitting of percentile curves for each subcortical volume in each of the three normative cohorts, we used the lambda-mu-sigma (LMS) method (Cole and Green, 1992).The LMS method can deal with skewed distributions and results in smooth percentile curves. The assumption of the LMS method is that the data are standard normally distributed after applying the Yeo-Johnson transformation, which is an extension of the Box-Cox transformation (Cole and Green, 1992). This method estimates the λ-parameter of the Yeo-Johnson transformation (Yeo and Johnson, 2000) (L), the median (M), and coefficient of variation (S) for the appropriate subcortical structure volume at each age. With the parameters L, M, and S, percentiles can be computed at each age to obtain a smooth curve. The smoothness of the fitted curves is influenced by the degrees of freedom δ, a user-defined parameter. In our experiments, we set the smoothness parameter δ to a value of 2 and we utilized the R-package VGAM (Yee, 2010) for the percentile curve fitting. The volume of a brain region may also be influenced by other covariates than age, for example, sex and head size. Including a covariate in the LMS model results in an age-dependent correction for the confounder. We therefore included sex in the LMS model as an confounder, which allows different percentile curves for men and women. To ensure an head size correction independent of age, head size was regressed out before fitting the LMS models. The precision of the estimated percentile curves depends on the number of data points in the appropriate age range. If the data are nonuniformly distributed over age, it could be that the curve estimation is not precise in the part where there are very few data points. To assess the precision of the fitted curves, we used a bootstrapping procedure, by random sampling subjects with replacement and re-estimating the percentile curves. A distribution of possible curves was collected, from which confidence intervals were estimated (Carpenter and Bithell, 2000).

Percentile curves were fit on the Rotterdam Study, UKBB, and ADNI reference populations separately for the subcortical volumes of the hippocampus, amygdala, putamen, thalamus, caudate and nucleus accumbens, and globus pallidus. With the MBS, the volume of the caudate and the nucleus accumbens are combined into one volume. Furthermore, for the analysis, the subcortical volumes were the sum of the left and right volume. The MBS method does not segment the extraventricular cerebrospinal fluid (CSF); therefore, the exact intracranial volume (i.e., the sum of brain tissue and all CSF) was not available. To correct for head size, the "estimated intracranial volume" was constructed as the sum of total brain volume and the intraventricular CSF volume. An explorative comparison of the estimated intracranial volume and the intracranial volume segmented previously in the Rotterdam Study for other purposes with FreeSurfer 5.1 showed a good correlation (0.93); therefore, the estimated intracranial volume was used to correct for head size.

3.3. Subject-specific comparison

To assess the influence of using a specific reference population on subject-specific percentile values, scans from all three cohorts served as a joint test set to reduce a cohort-specific bias caused by the different age range covered by each cohort. We estimated the percentile value for every subcortical structure, for all participant groups based on each of the three reference cohorts. This results in three percentile values per subcortical volume for each participant. To assess differences in these percentile values, the distributions of the percentile values based on the three reference populations within the different participant groups are compared using a Welch's two-sample t-test. In addition, the shift function, as described by Rousselet and Wilcox et al. (Rousselet and Wilcox, 2017), was used to describe the differences between the percentile distributions based on the three different populations, to account for non-normally distributed percentile distributions within the participant groups.

4. Results

Table 1 shows the characteristics of the different participant groups. Characteristics of the participant groups per cohort are shown in Supplementary Table 1.

4.1. Normative percentile curves

In Fig. 1, the normative percentile curves based on the Rotterdam Study, ADNI, and UKBB data sets are shown for the subcortical structure volumes: (A) hippocampus, (B) amygdala, (C) thalamus,

Table 1
Characteristics of the participant groups

Characteristic	Healthy	APOE ε4 carriers	MCI	AD	MCI AC	AD AC
Age (y)[a]	63.6 (20.5)	68.8 (18.7)	74.5 (10.7)	79.2 (10.4)	70.1 (6.9)	66.0 (11.9)
Sex, women	1149 (0.52)	74 (0.47)	41 (0.45)	53 (0.56)	8 (0.42)	15 (0.35)
Hippocampus volume (mL)	6.3 (0.8)	6.3 (0.8)	5.7 (0.9)	4.7 (0.8)	6.0 (0.7)	5.3 (0.9)
Amygdala volume (mL)	1.9 (0.3)	1.9 (0.3)	1.7 (0.3)	1.4 (0.3)	1.9 (0.3)	1.6 (0.3)
Putamen volume (mL)	8.2 (1.0)	8.3 (1.0)	7.7 (1.0)	7.2 (0.8)	7.9 (1.0)	7.3 (0.8)
Thalamus volume (mL)	13.0 (1.5)	13.1 (1.4)	12.2 (1.5)	11.2 (1.0)	12.2 (1.3)	12.2 (1.7)
Caudate and accumbens[b] volume (mL)	7.5 (0.9)	7.6 (0.9)	7.2 (0.9)	6.6 (0.8)	7.2 (1)	6.8 (1.4)
Globus pallidus volume (mL)	2.8 (0.4)	2.8 (0.3)	2.7 (0.4)	2.5 (0.3)	2.8 (0.3)	2.7 (0.3)
Estimated intracranial volume (mL)	1229.4 (127.2)	1245.0 (126.6)	1229.4 (144.8)	1137.7 (123.6)	1239.4 (137.9)	1212.2 (140.9)

Continuous variables are presented as means (standard deviations), and categorical variables as numbers (percentages).
Key: APOE ε4 carriers, healthy participants from the three reference populations who carry one or two APOE ε4 allele(s) ($N_{total} = 158$); AD AC, patients with AD who visited the Alzheimer Center ($N_{total} = 43$); AD, participants from the Rotterdam Study and ADNI data set with AD ($N_{total} = 95$); Healthy, healthy participants from the three reference populations ($N_{total} = 2201$); MCI, participants from the Rotterdam Study and ADNI data set with MCI ($N_{total} = 91$); MCI AC, patients with MCI who visited the Alzheimer Center ($N_{total} = 19$).
[a] Age is presented as the median and interquartile range because of the non-normal distribution of age.
[b] Combined caudate and nucleus accumbens volume (mL).

(D) putamen, (E) caudate and nucleus accumbens, and (F) globus pallidus. Considering the percentile curves and the corresponding confidence intervals around each curve, the percentile curves of hippocampus volume and caudate and nucleus accumbens of the three normative cohorts largely overlap, with a slightly higher volume in ADNI compared with Rotterdam Study and UKBB. For the amygdala, the percentile curves show small differences with higher volumes for UKBB, followed by ADNI and the lowest volumes for Rotterdam Study. For the putamen, thalamus, and globus pallidus, the ADNI and UKBB curves largely overlap, but the Rotterdam Study percentile curves show a lower volume. Furthermore, for almost all subcortical structures, the Rotterdam Study percentile curves show a larger decrease in volume over age than ADNI; however, the steepness of the curves between Rotterdam Study and UKBB seems comparable.

4.2. Subject-specific comparison

In Table 2, the average percentile values and standard deviations are shown for the different participant groups when based on each of the three normative cohorts. In general, differences shown in the percentile curves in Fig. 1 result in significant differences in the percentile distributions. For hippocampus and caudate volume, there are no significant differences in using the percentile curves from the Rotterdam Study, ADNI, or UKBB for any of the different participant groups. For the volumes of the putamen, there were significant differences in the percentiles within the healthy participants. Yet, for the APOE ε4 carriers, patients with MCI and AD, these differences were not significant. For the volumes of the thalamus, both the percentiles of the healthy participants and the patients with AD were statistically significant.

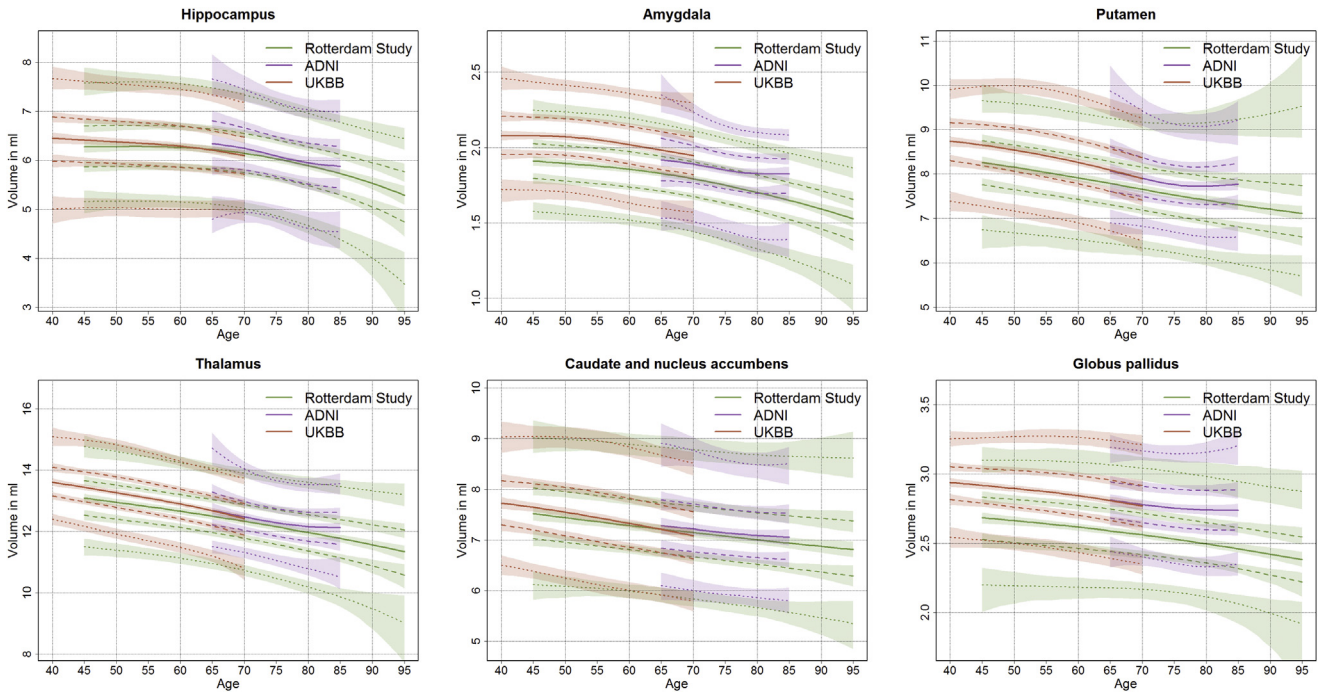

Fig. 1. Percentile curves of the subcortical structure volumes based on Rotterdam Study, UK Biobank (UKBB), and Alzheimer's Disease Neuroimaging Initiative (ADNI) data. The lower and upper dotted lines represent the 2.5% and 97.5% percentile curves, the lower and upper dashed lines represent the 25% and 75% percentile lines, and finally, the solid line represents the 50% percentile line, respectively. Around each percentile line the confidence interval is shown.

Table 2
Comparison of percentile values of the participant groups based on each of the three cohorts as reference curves

Subcortical structure	Participant group	RS_{ref}	$ADNI_{ref}$	$UKBB_{ref}$	Difference (p-value)	N_{total} (RS; ADNI; UKBB)
Hippocampus	Healthy	0.51 (0.29)	0.51 (0.30)	0.51 (0.30)	1	2201 (895; 430; 876)
	APOE ε4 carriers	0.51 (0.31)	0.51 (0.32)	0.51 (0.32)	1	158 (47; 61; 50)
	MCI	0.34 (0.30)	0.33 (0.30)	0.34 (0.32)	1	91 (41; 50; 0)
	AD	0.14 (0.21)	0.14 (0.21)	0.13 (0.21)	1	95 (45; 50; 0)
Amygdala	Healthy	0.62 (0.30)	0.56 (0.31)	0.38 (0.29)	<0.001[a]	2201 (895; 430; 876)
	APOE ε4 carriers	0.60 (0.32)	0.55 (0.33)	0.38 (0.29)	<0.001[b]	158 (47; 61; 50)
	MCI	0.40 (0.31)	0.34 (0.30)	0.20 (0.24)	<0.001[b]	91 (41; 50; 0)
	AD	0.22 (0.27)	0.18 (0.24)	0.09 (0.16)	<0.001[c]	95 (45; 50; 0)
Putamen	Healthy	0.57 (0.28)	0.55 (0.30)	0.49 (0.30)	<0.001[b]	2201 (895; 430; 876)
	APOE ε4 carriers	0.60 (0.26)	0.58 (0.29)	0.52 (0.29)	0.39	158 (47; 61; 50)
	MCI	0.51 (0.25)	0.47 (0.28)	0.46 (0.27)	1	91 (41; 50; 0)
	AD	0.48 (0.29)	0.43 (0.31)	0.50 (0.30)	1	95 (45; 50; 0)
Thalamus	Healthy	0.56 (0.28)	0.61 (0.31)	0.52 (0.29)	<0.001[a]	2201 (895; 430; 876)
	APOE ε4 carriers	0.59 (0.27)	0.64 (0.29)	0.56 (0.28)	0.17	158 (47; 61; 50)
	MCI	0.41 (0.29)	0.43 (0.32)	0.46 (0.29)	1	91 (41; 50; 0)
	AD	0.42 (0.27)	0.39 (0.29)	0.51 (0.26)	0.028[d]	95 (45; 50; 0)
Caudate and nucleus accumbens	Healthy	0.51 (0.28)	0.53 (0.30)	0.53 (0.29)	1	2201 (895; 430; 876)
	APOE ε4 carriers	0.55 (0.26)	0.57 (0.28)	0.57 (0.28)	1	158 (47; 61; 50)
	MCI	0.45 (0.28)	0.45 (0.30)	0.50 (0.30)	1	91 (41; 50; 0)
	AD	0.38 (0.27)	0.38 (0.29)	0.47 (0.30)	0.54	95 (45; 50; 0)
Globus pallidus	Healthy	0.66 (0.28)	0.42 (0.32)	0.40 (0.30)	<0.001[e]	2201 (895; 430; 876)
	APOE ε4 carriers	0.7 (0.26)	0.47 (0.32)	0.46 (0.31)	<0.001[e]	158 (47; 61; 50)
	MCI	0.67 (0.25)	0.40 (0.30)	0.41 (0.28)	<0.001[e]	91 (41; 50; 0)
	AD	0.61 (0.30)	0.36 (0.31)	0.39 (0.30)	<0.001[e]	95 (45; 50; 0)

Mean and standard deviation of the percentiles of the different participant groups (healthy participants, APOE ε4 carriers, participants with MCI, and participants with AD) based on the reference curves of each of the three normative cohorts (RS ref, ADNI ref, and UKBB ref).
Difference: smallest p-value of the paired t-tests; N_{total}: sample size of the participant groups.
Key: RS, Rotterdam Study; UKBB, United Kingdom Biobank; ADNI, Alzheimer's Disease Neuroimaging Initiative; APOE ε4 carriers, healthy participants from the three reference populations who carry one or two APOE ε4 allele(s); AD, participants from the Rotterdam Study and ADNI data set with AD; AD AC, patients with AD who visited the Alzheimer Center; Healthy, healthy participants from the three reference populations; MCI, participants from the Rotterdam Study and ADNI data set with MCI; MCI AC, patients with MCI who visited the Alzheimer Center.
[a] percentile values based on the three normative cohorts are all significantly different from each other.
[b] percentile values based on UKBB data are significantly different from those based on the other cohorts.
[c] percentile values based on Rotterdam Study data are significantly different from those based on the UKBB data.
[d] percentile values based on ADNI data are significantly different from those based on the UKBB data.
[e] percentile values based on Rotterdam Study data are significantly different from those based on the other cohorts.

For the amygdala volume, the percentile values based on the three cohorts were all significantly different. However, for the APOE ε4 carriers, MCI and AD cases, only the percentile values based on UKBB were significantly lower than the other reference cohorts. For the globus pallidus, there was a significantly higher percentile value based on the Rotterdam Study data versus the other two cohorts, which is a reflection of the significantly lower percentile curves for the Rotterdam Study, as shown in Fig. 1.

In Supplemental Fig. 2, the results from the shift-function analyses are shown, for the four different participant groups. The results show overall a straight line for all participant groups, indicating a fixed percentile difference when comparing the percentiles based on two different populations, which is independent of the percentile value itself. The exceptions are the comparison of amygdala percentiles based on UKBB compared with those based on the Rotterdam Study and ADNI. Here, a higher percentile value is related to a larger percentile difference. The same holds for the globus pallidus percentile based on the Rotterdam Study compared with ADNI and UKBB.

Finally, in Table 3, the mean and standard deviation of the estimated percentiles for the participants with AD and MCI from the Alzheimer Center are shown. Within this sample, there was

Table 3
Comparison of percentile values of the participants with MCI and AD from the Alzheimer Center, based on the reference curves from each of the three normative cohorts (RS_{ref}, $ADNI_{ref}$, and $UKBB_{ref}$)

Subcortical structure	Participants	Average percentile (SD)			Difference	N
		RS_{ref}	$ADNI_{ref}$	$UKBB_{ref}$		
Hippocampus	MCI	0.36 (0.32)	0.36 (0.33)	0.38 (0.33)	1	19
	AD	0.16 (0.23)	0.16 (0.22)	0.16 (0.22)	1	43
Amygdala	MCI	0.58 (0.36)	0.53 (0.36)	0.39 (0.33)	1	19
	AD	0.27 (0.3)	0.22 (0.28)	0.14 (0.24)	0.5	43
Putamen	MCI	0.49 (0.3)	0.46 (0.32)	0.42 (0.31)	1	19
	AD	0.3 (0.23)	0.26 (0.24)	0.23 (0.23)	1	43
Thalamus	MCI	0.33 (0.23)	0.34 (0.26)	0.31 (0.22)	1	19
	AD	0.34 (0.29)	0.35 (0.32)	0.31 (0.3)	1	43
Caudate and nucleus accumbens	MCI	0.41 (0.3)	0.41 (0.31)	0.43 (0.3)	1	19
	AD	0.33 (0.33)	0.34 (0.34)	0.35 (0.34)	1	43
Globus pallidus	MCI	0.67 (0.24)	0.4 (0.32)	0.4 (0.3)	0.08	19
	AD	0.66 (0.31)	0.44 (0.33)	0.43 (0.32)	0.023[a]	43

Difference: smallest p-value of the paired t-tests; N_{total}: sample size of the participant groups; N: sample size of the Alzheimer Center set.
Key: AD, Alzheimer's disease; MCI, mild cognitive impairment; RS, Rotterdam Study; UKBB, United Kingdom Biobank; ADNI, Alzheimer's Disease Neuroimaging Initiative.
[a] percentile values based on Rotterdam Study data are significantly different from those based on other cohorts.

only a significant difference for the globus pallidus volume in the patients with AD. Other percentile estimations in these groups did not differ depending on the reference curves applied.

5. Discussion

In this study, we calculated normative reference curves for subcortical structure volumes from reference populations that were either derived from population-based studies or from normal controls of a case-control study. We used a segmentation method for which previous experiments on data from both 3T and 1.5T for different scanners indicate good agreement with respect to independent ground truth segmentations, regardless of the field strength or vendor. We found that for most subcortical structures, the percentile curves of the subcortical structures largely overlap. This indicates only small differences between the subcortical volumes of these reference populations, regardless of differences in vendors, field strength, acquisition, and population differences. When estimating the percentile values for various participant groups that may be evaluated in a clinical setting (APOE ε4 carriers, and patients with MCI and AD), the choice of reference population did not influence the percentile distribution significantly, except for the smallest subcortical structures: amygdala and globus pallidus. In particular, the hippocampus percentile curve was very robust across the participant groups. This indicates that individual diagnostic assessment in a clinical setting, based on subcortical volume information, may not be biased by the use of a specific reference population.

5.1. Strengths and limitations

A major strength of this study is the use of a single segmentation tool on MRI scans from various different large reference populations, giving a comprehensive overview of subcortical volumes in aging in these populations. Another strength of the study is the availability of scans from patient groups (MCI and AD) from the Rotterdam Study and ADNI, as well as a patient population independent from the reference populations, that is, the Alzheimer Center data. There are a number of limitations associated with this study. First, a limitation concerning the volume segmentation method used in this study is the lack of segmentation of extraventricular cerebrospinal fluid because of which the intracranial volume estimated in this study gives an underestimation of the true intracranial volume (or head size). This underestimation may lead to an underestimation of the atrophy effect in aging, when the changes in ventricular cerebrospinal fluid are not representative of the extraventricular cerebrospinal fluid changes in aging. Yet, a sensitivity analysis within the Rotterdam Study population in whom both estimated intracranial volume and exact intracranial volume were available showed these effects to be negligible. Second, the LMS method used in this study for the estimation of the percentile curves results in smooth percentile curves in aging, which can deal with skewed distributions. Within this study, other methods to estimate percentile curves would also have been suitable, assuming that the subcortical volumes over age are normally distributed. Within the context of this study, we believe that the impact of the choice of the percentile curve estimation method on the differences between populations is minimal, as long as the same percentile curve fitting method is the same for the different reference populations. Third, a limitation concerning the generalizability of these percentile curves is the fact that the vast majority of the healthy study participants and the participant groups are Caucasian. Therefore, differences in percentile curves which could result from differences in ethnicity of study populations are not assessed in this study. Fourth, in this study, we are not able to determine the exact source of differences in the subcortical volumes between reference populations because of the variation in vendor, field strength, and acquisition used in the different populations. However, we are able to demonstrate the magnitude of these differences, indicating the impact of these differences on individual patient assessment in an everyday clinical setting. Fifth, the lack of overlap of the complete age range of all three reference populations is a limitation of this study, making comparison of the reference curves more difficult. Given the important differences in age ranges, comparison of percentile volumes of healthy participants was performed on the combined healthy participants of the Rotterdam Study, ADNI, and UKBB, although the percentile curves itself were fitted on these same reference populations. Ideally, separate healthy participants from the same reference populations, which are not included in the percentile curve fitting, would be used to test percentile value differences for the different reference populations. Furthermore, this study was limited to the subcortical structure volumes including hippocampus volume, whereas in a clinical setting, other (cortical) brain volumes would be also of importance. The current choice for subcortical volumes was driven on the one hand by an increasing scientific interest into subcortical volumes in neurological diseases (including neurodegenerative diseases in older age) and on the other hand because of the availability of a proven robust segmentation algorithm, which performs population- and vendor-independent, eliminating potential sources of noise. Yet, a logical next step would be to explore the dependence of cortical segmentation algorithms on the choice of reference population. Finally, in this study, the patient population on which the effect of different reference populations were estimated consisted of only patients with AD or participants at higher risk of AD. Next step would be to evaluate reference curves of a broad spectrum of brain structures based on different reference populations and the effect of these differences on diagnostic assessments in different neurological diseases and neuropsychiatric disorders.

5.2. Differences between reference populations

In general, we found slightly lower reference volumes based on the scans from the Rotterdam Study compared with the ADNI and UKBB reference curves. A possible explanation for these differences is that the Rotterdam Study population is a population-based cohort, whereas ADNI has a case-control design. Within the healthy set of the Rotterdam Study, participants with MCI or AD at time of the scan were excluded, whereas the ADNI controls are cognitively normal with no memory complaints and no significant cognitive impairment. Therefore, the control subjects from ADNI are expected to be healthier than the Rotterdam Study population. On the other hand, Rotterdam Study percentile curves also show a slightly lower value than the UKBB percentile curves, which is a population-based cohort as well. This may be due to a lower response rate in the UKBB, with the possibility of healthy selection bias. Furthermore, a slight increase in subcortical volume has been seen in the ADNI reference curves from age 80 to 85. A possible explanation might be that with increasing age (especially from age 80 and older), the ADNI controls become proportionately healthier than control subjects from a population-based study because of the fact that the higher a subject's age, the more likely he or she is to have memory complaints. This could be interpreted as an increasing healthy selection bias with age. Another possible explanation for differences between reference populations could be the fact that scans from both the Rotterdam Study and UKBB have each been acquired on the same vendor (with only a single scanner for the Rotterdam Study), whereas scans in the ADNI database were collected from different scanners, field strengths, and scanner types. Therefore, characteristics of a single

scanner or vendor with an impact on volumetric segmentation might be more dominant in Rotterdam Study and UKBB curves. This effect might have a larger impact on small structures as well as such with subtle contrast boundaries like amygdala and globus pallidus, explaining more pronounced differences between their corresponding percentile curves and additionally the larger percentile differences with higher percentile volumes in these small structures. The study by Potvin et al. (2016) that created normative curves for subcortical structures and evaluated the effects of scanner characteristics also showed that for the amygdala structure volume, the effects of scanner characteristics were modest, whereas in the most other structures, the effect was minor compared to age, sex, and intracranial volume.

6. Summary and conclusion

Overall, we found that the percentile curves of the subcortical structure based on three different reference populations largely overlap, indicating only small differences between the subcortical volumes of these populations, regardless of differences in vendors, field strength, acquisition, and population differences. Therefore, we conclude that the subcortical volume data of these three cohorts are interchangeable, suggesting more flexibility in clinical implementation.

Disclosure

Martin Bergtholdt and Fabian Wenzel are employees of Philips Research. Wiro J. Niessen is a cofounder, chief scientific officer, and shareholder of Quantib BV. None of the authors has potential conflicts of interest related to this manuscript.

Acknowledgments

This project has received funding from the European Union's Seventh Framework Programme for Research, Technological Development, and Demonstration under grant agreement no. 601055 (VPH-DARE@IT). This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 666992. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (project: ORACLE, grant agreement No: 678543).

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam; Netherlands Organization for the Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc; Cogstate; Eisai Inc; Elan Pharmaceuticals, Inc; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc; Fujirebio; GE Healthcare; IXICO Ltd; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co, Inc; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This research has been conducted using the UK Biobank Resource under Application Number 23509.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neurobiolaging.2019.07.008.

References

Brewer, J.B., 2009. Fully-automated volumetric MRI with normative ranges: Translation to clinical practice. Behav. Neurol. 21, 21–28.

Carpenter, J., Bithell, J., 2000. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat. Med. 19, 1141–1164.

Cavedo, E., Suppa, P., Lange, C., Opfer, R., Lista, S., Galluzzi, S., Schwarz, A.J., Spies, L., Buchert, R., Hampel, H., 2017. Fully automatic MRI-based hippocampus volumetry using FSL-FIRST: Intra-scanner test-retest stability, inter-field strength variability, and performance as enrichment biomarker for clinical trials using prodromal target populations at risk for Alzheimer's. J. Alzheimer's Dis. 60, 151–164.

Cole, T.J., Green, P.J., 1992. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat. Med. 11, 1305–1319.

Heinen, R., Bouvy, W.H., Mendrik, A.M., Viergever, M.A., Biessels, G.J., De Bresser, J., 2016. Robustness of automated methods for brain volume measurements across different MRI field strengths. PLoS One 11, e0165719.

Ikram, M.A., Brusselle, G.G., Murad, S.D., van Duijn, C.M., Franco, O.H., Goedegebure, A., Klaver, C.C., Nijsten, T.E., Peeters, R.P., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vernooij, M.W., Hofman, A., 2017. The Rotterdam Study: 2018 update on objectives, design and main results. Eur. J. Epidemiol. 32, 807–850.

Kälin, A.M., Park, M.T.M., Chakravarty, M.M., Lerch, J.P., Michels, L., Schroeder, C., Broicher, S.D., Kollias, S., Nitsch, R.M., Gietl, A.F., Unschuld, P.G., Hock, C., Leh, S.E., 2017. Subcortical shape changes, hippocampal atrophy and cortical Thinning in Future Alzheimer ' s disease patients. Front. Aging Neurosci. 9, 1–17.

Maclaren, J., Han, Z., Vos, S.B., Fischbein, N., Bammer, R., 2014. Reliability of brain volume measurements: a test-retest dataset. Scientific Data 1, 1–9.

Peterson, M., Warf, B.C., Schiff, S.J., State, T.P., Medicine, S., 2018. Normative human brain volume growth. J. Neurosurg. Pediatr. 21, 478–485.

Potvin, O., Mouiha, A., Dieumegarde, L., Duchesne, S., 2016. NeuroImage Normative data for subcortical regional volumes over the lifetime of the adult human brain. NeuroImage 137, 9–20.

Potvin, O., Dieumegarde, L., Duchesne, S., Initiative, N., 2017. NeuroImage Normative morphometric data for cerebral cortical areas over the lifetime of the adult human brain. NeuroImage 156, 315–339.

Roh, J.H., Anqi, Q., Seo, S.W., Soon, H.W., Kim, J.H., Kim, G.H., Kim, M.-J., Lee, J.-M., Na, D.L., 2011. Volume reduction in subcortical regions according to severity of Alzheimer' s disease. J. Neurol. 258, 1013–1020.

Ross, D.E., Ochs, A.L., Seabaugh, J.M., Shrader, C.R., Initiative, t.A.D. N., 2013. Man versus Machine: comparison of Radiologists' interpretations and NeuroQuant volumetric analyses of brain MRIs in patients with Traumatic brain Injury. J. Neuropsychiatry Clin. Neurosci. 25, 32–39.

Ross, D.E., Ochs, A.L., Desmit, M.E., Seabaugh, J.M., Havranek, M.D., Initiative, T.A.D.N., 2015. Man versus Machine Part 2: comparison of Radiologists ' interpretations and NeuroQuant measures of brain Asymmetry and Progressive atrophy in patients with Traumatic brain Injury. J. Neuropsychiatry Clin. Neurosci. 27, 147–152.

Rousselet, G.A., Wilcox, R.R., 2017. Beyond differences in means: robust graphical methods to compare two groups in neuroscience. Eur. J. Neurosci. 46, 1738–1748.

Rummel, C., Aschwanden, F., McKinley, R., Wagner, F., Salmen, A., Chan, A., Wiest, R., 2018. A fully automated Pipeline for normative atrophy in patients with neurodegenerative disease. Front. Neurol. 8, 1–16.

Stepan-Buksakowska, I., Szabo, N., Horinek, D., Toth, E., Hort, J., Warner, J., Charvat, F., Vecsei, L., Rocek, M., Kincses, Z.T., 2014. Cortical and subcortical atrophy in Alzheimer disease: parallel atrophy of thalamus and hippocampus. Alzheimer Dis. Assoc. Disord. 28, 65–72.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A.,

Young, A., Sprosen, T., Peakman, T., Collins, R., 2015. UK Biobank: an open access resource for Identifying the Causes of a Wide range of Complex diseases of Middle and old age. PLoS Med. 12, 1–10.

Tondelli, M., Wilcock, G.K., Nichelli, P., De Jager, C.A., Jenkinson, M., Zamboni, G., 2012. Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. Neurobiol. Aging 33, 825 e25–e36.

Tudorascu, D.L., Karim, H.T., Maronge, J.M., Alhilali, L., Fakhran, S., Aizenstein, H.J., Muschelli, J., Crainiceanu, C.M., 2016. Reproducibility and bias in healthy brain segmentation: comparison of two popular neuroimaging platforms. Front. Neurosci. 10, 1–8.

Tutunji, R., El, M., Saaybi, S., Al, N., Tamim, H., 2018. Thalamic volume and dimensions on MRI in the pediatric population: normative values and correlations (A cross sectional study). Eur. J. Radiol. 109, 27–32.

Velasco-Annis, C., Akhondi-Asl, A., Stamm, A., Warfield, S.K., 2018. Reproducibility of brain MRI segmentation algorithms: Empirical comparison of local MAP PSTAPLE, FreeSurfer, and FSL-FIRST. J. Neuroimaging 28, 162–172.

Vernooij, M.W., Jasperse, B., Steketee, R., Koek, M., Vrooman, H., Ikram, M.A., Papma, J., Lugt, A.V.D., Smits, M., Niessen, W.J., 2018. NeuroImage: clinical Automatic normative quantification of brain tissue volume to support the diagnosis of dementia: a clinical evaluation of diagnostic accuracy. NeuroImage 20, 374–379. https://doi.org/10.1016/j.nicl.2018.08.004.

Wenzel, F., Meyer, C., Stehle, T., Peters, J., Siemonsen, S., Thaler, C., Zagorchev, L., 2018. Rapid fully automatic segmentation of subcortical brain structures by shape-constrained surface adaptation. Med. Image Anal. 46, 146–161.

Yee, T.W., 2010. The VGAM package for categorical data analysis. J. Stat. Softw. 32, 1–34.

Yeo, I.-K., Johnson, R.A., 2000. Biometrika Trust A New Family of Power Transformations to Improve Normality or Symmetry. Oxford University Press on behalf of Biometrika Trust Stable Biometrika 87 (4), 954–959.