Comparison of manual and automatic techniques for sub-striatal segmentation in ^{11}C-Raclopride high-resolution PET studies

Jarkko Johansson[1,2], Kati Alakurtti[1,3], Juho Joutsa[2,4], Jussi Tohka[5,8], Ulla Ruotsalainen[6,7], Juha O. Rinne[1,4]

(1) Turku PET Centre, Turku University Hospital, Turku, Finland

(2) Turku PET Centre, University of Turku, Turku, Finland

(3) Department of Diagnostic Radiology, Turku University Hospital, Turku, Finland

(4) Division of Clinical Neurosciences, Turku University Hospital, Turku, Finland

(5) Departamento de Bioingeniería e Ingeniería Aeroespacial. Universidad Carlos III de Madrid, Leganes, Spain

(6) Department of Signal Processing, Tampere University of Technology, Tampere, Finland

(7) BioMediTech, Tampere, Finland

(8) Instituto de Investigación Sanitaria Gregorio Marañon, Madrid, Spain

Correspondence to: Jarkko Johansson, Turku PET Centre, University of Turku, P.O. Box 52, FIN-20521 Turku, Finland. Tel. +358-2-313 30564; fax. +358-2-231 8191; e-mail: jarkko.johansson@tyks.fi

Sources of support: The study was supported by the Hospital District of Southwest Finland (13133, Johansson). JT's work is funded by the Universidad Carlos III de Madrid, the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement nº 600371, el Ministerio de Economía y Competitividad (COFUND2013-40258) and Banco Santander.

Declaration of Interest: Dr Joutsa has received lecturer honoraria from Boehringer-Ingelheim, a research grant from Lundbeck and a travel grant from Abbvie. For the remaining authors none were declared.

ABSTRACT

The striatum is the primary target in regional ^{11}C-Raclopride PET studies, and despite its small volume, it contains several functional and anatomical subregions. The outcome of the quantitative dopamine receptor study using ^{11}C-Raclopride PET depends heavily on the quality of the region-of-interest (ROI) definition of these subregions. The aim of this study was to evaluate subregional analysis techniques because new approaches have emerged but have not yet been directly compared. In this paper, we compared manual ROI delineation to several automatic methods. The automatic methods employed either direct clustering of the PET image or individualization of chosen brain atlases based on MRI or PET image normalization. State-of-the-art normalization methods and atlases were applied, including those provided in the FreeSurfer, SPM8 and FSL software packages. Evaluation of the automatic methods was based on voxel-wise congruity with the manual delineations and the test-retest variability and reliability of the outcome measures using data from seven healthy male subjects who were scanned twice with ^{11}C-Raclopride PET on the same day. The results show that both manual and automatic methods can be used to define striatal subregions. Though most of the methods performed well with regard to the test-retest variability and reliability of binding potential, the smallest average test-retest variability (TRV) and standard error of measurement (SEM) were obtained using a connectivity-based atlas and PET normalization (TRV=4.5%, SEM=0.17). In conclusion, the current state-of-the-art automatic ROI methods can be considered good alternatives for subjective and laborious manual segmentation in ^{11}C-Raclopride PET studies.

KEYWORDS: PET, 11C-Raclopride, striatum, ROI analysis

Introduction

Positron-emission tomography (PET) with ^{11}C-Raclopride provides a widely validated method for assessing baseline levels of dopamine (DA) type 2 receptor (DA2R) availability [1] as well as DA2R occupancy provoked by either pharmacological [2] or non-pharmacological [3;4] stimuli. In humans the DA2R are most abundant in the striatum where they act as modulators of various functions, such as locomotion, and reward-system, as well as many high order cognitive functions such as working memory [5]. The reward system is thought to be linked with addictive behaviors such as substance abuse [6] and pathological gambling [7], and thereby linked to dopaminergic signaling. Thus, ^{11}C-Raclopride-PET can be employed in the research of numerous facets of human behavior and cognition in health and in disease (see [4] for review). As a downside, methodological caveats such as erroneous region-of-interest (ROI) delineation can significantly hamper the reliability and sensitivity of ^{11}C-Raclopride-PET due to small size of striatal substructures relative to the spatial resolution of PET. In the current work we tackled the ROI-delineation issue through investigation of numerous ROI-delineation techniques in the analysis of high-resolution ^{11}C-Raclopride-PET.

Inaccuracies in ROI delineation can hamper the sensitivity of the PET assessment, and in the worst case generate biased inferences (c.f. [4]). In a PET assessment of specific receptor binding, decreased sensitivity can be due to oversized ROIs compared to the receptor population, or due to intra- or inter-subject mismatches in region definitions. Oversized ROIs can yield exaggerated partial-volume effect (PVE) resulting in underestimated PET radioactivity, while direct PET-based segmentation can in the worst case result in biased outcome (c.f. [4]). Furthermore, inaccuracies in the ROI border placement between adjacent regions can result in attenuated stimulus-response in a single assessment, or decreased statistical power in a group-level assessment due to additional methodological variability between subjects.

Concept of sub-striatal ROI-delineation that is based on functional rather than structural territories was first introduced by Mawlawi and colleagues [8] and later reviewed and validated by Martinez and colleagues using d-amphetamine induced striatal activation [9]. Martinez and colleagues [9] defined functional territories on the basis of experimental animal studies that included three striatal subregions:

the limbic striatum (LSTR), involved in drive and motivation; the associative striatum (ASTR), involved in cognition; and the sensorimotor striatum (SMST), involved in locomotion. Furthermore, they showed significant differences in the response to amphetamine-stimulus between the subregions using ^{11}C-Raclopride-PET. The concept of functional subdivision is nowadays established and the ROI-delineations are mostly generated by hand using the guidelines of Mawlawi, Martinez and colleagues. More recently Tziortzi and colleagues [10] have refined the guidelines of Mawlawi and colleagues in the light of current understanding of the striatal functional organization in humans. Nevertheless, manual ROI-delineation inevitably contains a subjective component that hampers the reproducibility of ^{11}C-Raclopride-PET analysis. Thus, automated methods are called upon that provide precise and repeatable striatal ROI-delineation.

The automated, or operator-independent methods for striatal ROI-delineation can be based on a) individualization of a ROI-set defined in a standard space through image deformation (see [11] for review); b) by direct automatic clustering of the PET image (see [12] or c) individual measurement of cortico-striatal connectivity as suggested by Tziortzi and colleagues [13]. In the current work we examined the approaches that did not require DTI-measurements. The first approach based on the individualization of a template ROI-set has been successfully implemented using conventional PET-scanning (see [11] for review). In the study of del Campo and colleagues [11] a probabilistic atlas was generated on the basis of manual ROI-delineations and subsequently individualized using non-rigid transformations. Their conclusion was that automated method based on probabilistic atlas can provide an accurate and efficient alternative to manual ROI-drawing. The second approach employing direct PET image clustering has been previously implemented for high-resolution PET data (see [14] for validation). In the study of Farinha and colleagues the direct PET image clustering was used to generate five striatal subregions in both striata instead of three for the first time [14].

Primary aim of the current study was to explore the characteristics of various ROI-delineation methods employed in the analysis of high-resolution ^{11}C-Raclopride-PET, as new methodologies have emerged but have not yet been evaluated. ^{11}C-Raclopride-PET has widespread usage and recent investigations suggest benefit of connectivity-based ROI-methods over the structure-based manual ROI-delineation

[13]. In particular, the potential advantage of high-resolution PET is not fully exploited if the ROI-methods are inaccurate. We approached the ROI-method optimization task through evaluation of the current state-of-the-art ROI delineation methods with regard to their test-retest characteristics. A test-retest dataset of 7 healthy male subjects scanned twice (during the same day) with the HRRT was previously collected to investigate the short-term repeatability and reliability of the 11C-Raclopride assessment in high-resolution and at resting state [15]. The ultimate aim of test-retest setup is to substitute testing against an unknown ground truth in methodological testing, and it has therefore been recommended for methodological comparison studies (see [16] for review). The repeated scans in the presumably same condition allow estimation of method-wise bias as well as reliability through analysis of variance (ANOVA). It was shown in our previous report [15] that the binding potentials (BPs) within the manually delineated SB ROIs (based on rules by Martinez and colleagues) were not biased crossing the trials for the current dataset, thus, it is reasonable to expect non-biased estimates using any of the tested methods. Furthermore, the test-retest setup can be employed to estimate the dissection between the inherent, true variation between subjects from that due to methodological imprecision.

Methods

Overview

The primary objective of the current work was to compare the striatal ROI delineation approaches in the analysis of high-resolution ^{11}C-Raclopride-PET data to provide practicable guidelines for realistic imaging conditions. Analysis of the ^{11}C-Raclopride-PET data most commonly employs the SRTM [17] with cerebellar cortex as the reference region to obtain the BP_{ND} (see [18] for nomenclature) estimates, thus, the SRTM-BP_{ND} was considered as the primary parameter of interest in the test-retest evaluations.

Subjects

The current study re-uses human ^{11}C-Raclopride-PET measurements acquired originally for different purpose as reported in [15]. The original study protocol was approved by the Ethics Committee of the Hospital District of Southwestern Finland. The study subjects were given written information about all of

the relevant issues involved in the study. Written consent, not limited for single usage of the data, was obtained from each subject. The study was performed according to the Declaration of Helsinki.

All of the subjects were right handed and nonsmokers. The age, height, and weight of the subjects were 24.5±3.5 years, 185.5±12.5 cm, and 74±14 kg, respectively (mean±SD). To exclude any structural brain abnormalities and obtain anatomical references, all subjects underwent 1.5T magnetic resonance imaging. Each subject underwent two ^{11}C-Raclopride PET scans in a resting condition on the same day between 10.00 a.m. and 6.00 p.m., the tracer injections being at least 2.5 hours apart.

The binding potentials of these subjects have been reported previously [15]. However, in the current study, the data set was exploited for the ROI method comparison, rather than for research of the binding potential per se. None of the results in the current study are identical to the previously reported results due to small differences in the data preprocessing and manual ROI delineation. However, there was a strong correlation between the results from these two studies (data not shown); and the current results should not be considered as individual observations in a meta-analysis.

MR imaging

The MRI was performed with a 1.5-T MRI system (Gyroscan Intera CV Nova Dual; Philips Medical Systems, Best, The Netherlands) with a SENSE head coil. Transversal T1-FFE 3D images were acquired with isotropic 1 mm voxel size, repetition time (TR) 25 ms, echo time (TE) 4.6 ms and flip angle (deg) 30. Parallel imaging was used with factor 2.

PET imaging

The ^{11}C-Raclopride preparation and PET imaging have been previously described [15]. PET measurements were performed using the HRRT scanner (Siemens Medical Solutions, Knoxville, TN, USA), a brain-dedicated high-resolution PET research tomograph. The HRRT system is a dual-crystal-layer scanner capable of depth-of-interaction (DOI) measurement of the coincidence photons. The intrinsic resolution of the HRRT system is approximately 2.5 mm, while the measured spatial resolution (full-width-at-half-maximum, FWHM, of a point-source) varies between 2.5 mm and 3.5 mm within a field-of-view (FOV) covering most of the brain [19]. In the current work a transmission scan was

acquired prior to ^{11}C-Raclopride injection for attenuation-map calculation. PET scanning in listmode was initiated at the time of tracer injection and continued until 55 minutes from injection time. A dynamic serie was generated from the listmode data using the following frame sequence: 2x0.5, 9x1, 3x2, 3x3, and 6x5 minutes. Image reconstructions were performed using the ordinary-Poisson ordered-subsets expectation maximization (OP-OSEM) algorithm [20], with resolution-modeling (RM-OP-OSEM) [21] based on external measurement of the scanner point-spread function (PSF). Number of iterations in RM-OP-OSEM reconstruction was 10 while the number of subsets was 16, and the image voxel-size was approximately 1.22x1.22x1.22 mm^3. Tissue attenuation maps were reconstructed using the maximum-a-posteriori for transmission data (MAP-TR) algorithm with the standard human brain priors for bone, soft tissue, noise, water, or air [22]. Scattered events were estimated using the single scatter simulation algorithm [23], while randoms were estimated from the block singles using a variance reduction algorithm [24]. In the RM-OP-OSEM algorithm the emission data was not pre-corrected but the correction terms were included in the update equation in order to avoid bias due to zero-truncation of the pre-corrected data [20].

Image Preprocessing

Image preprocessing steps are illustrated in a diagram in Figure 1. Image preprocessing was conducted in Statistical Parametric Mapping (SPM) software (version 8, Wellcome Trust Centre for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm/). Firstly, the SPM8 realign-function was used for correcting the dynamic PET imaging data for misalignments between scans as well as between frames using normalized mutual-information (NMI) algorithm. The first frame with a decent signal (frame 5) was chosen as a reference in both sessions. Initially, each frame in both PET sessions were registered within session (to frame 5 position), followed by registration of the second PET session data to the first PET session orientation (frame 5 of session two to the frame 5 of session one). After the first pass, a mean image of all frames (from both sessions) was formed and each frame was re-registered to the mean image. The two pass procedure was employed in order to minimize misalignment between PET sessions that might hamper the test-retest characteristics in ROI-based analysis. Registrations were visually confirmed. The subsequent analysis were performed with the assumption that the data from

the two PET-sessions were spatially aligned, that is, the ROI-methods that do not directly employ PET-data were not repeated for each PET, but the same ROIs were employed for both sessions. The two pass registration process should help to protect against any systematic biases in the registration between the two PET sessions.

Furthermore, a MRI-based procedure was employed to standardize the head posture in PET data. First, the MRI-data was registered with the individual mean PET image (c.f. previous paragraph) using the SPM8 coreg-function and NMI optimization. Registrations were visually confirmed. Secondly, the MRI-data was registered with a MNI152 template image using rigid registration (coreg-function). The coreg-function allowed simultaneous registration of pre-registered PET-data through transformation matrix manipulation. Both the MRI- and PET-data were resliced into 1.5 mm x 1.5 mm x 1.5 mm voxel size of the MNI152 template image. The rigid registration does not provide exact spatial matching between individuals but works as a means to standardize the orientation of the anterior commissure (AC) – posterior commissure (PC) line. Standardization of the AC-PC line orientation may facilitate the manual ROI delineation as the structures appear in the same orientation as in the typical brain atlases, while the head posture during PET imaging can be somewhat tilted.

Non-rigid Image Deformation Procedures

Spatial normalization parameters (deformation fields; non-rigid mappings) were obtained using the Unified Segmentation algorithm in SPM8 [25] and using combination of the FLIRT and FNIRT functions in FSL5 (FMRIB, Oxford, UK, http://fsl.fmrib.ox.ac.uk/) [26]. First, the spatial normalization mappings were applied to the individual mean PET images. After visual inspection the FSL-based normalized PET-data was found slightly more consistent as compared to the SPM-based, thus FSL-normalization was chosen for PET-template image generation. A Raclopride-specific template image was formed from the seven normalized ^{11}C-Raclopride mean images as an average (see Figure, Supplemental Digital Content 6, for visualization). PET-based image normalization was performed using SPM8 Normalise-function [27], and the novel Raclopride-specific template image as a target. Both the source and target image data were pre-smoothed using a 3D Gaussian filter with 4 mm (FWHM) kernel size in each direction.

In the current work the goal of non-rigid spatial matching was to form an individualizing spatial mapping from the atlas-defining space to the individual coordinates. Atlas-based ROI-delineation can be obtained through either mapping the PET-data into atlas-defining space or mapping the atlas into individual space. In the current work the latter approach was chosen to minimize the need for interpolation of the PET-data, and to more readily obtain quality assurance of the spatial deformations through visual inspection of the ROIs in the individual space. The inverse deformation (that is, deformation from standard space to the individual space) was obtained directly in the Unified Segmentation algorithm in SPM, while the FSL-based deformation fields were explicitly inversed using invwarp-function of FSL. The PET-based SPM normalizations were inversed using the Deformations-function and the Inverse-component in SPM8. The three individualizing mappings are hereafter denoted as MRIF, MRIS and PETS for MRI-based mappings using FSL and SPM and PET-based mapping using SPM, respectively.

ROI Methods

The manual and automatic ROI methods for sub-striatal segmentation examined in this study are summarized in Table 1 and illustrated in a diagram in Figure 1. The methods examined in the current study had differences in the naming and definitions of the striatal subregions. To allow for a comparison between methods, common regions consisting of the whole striatum (STR), the limbic striatum (LSTR), the associative striatum (ASTR) and the sensorimotor striatum (SMST) were formed as primary targets. Table 2 summarizes the initial regions and combinations for each method.

Manual Segmentation

Manual segmentation of the striatum was performed by two operators, one with more experience (MANSEG1) and other with less experience but instructed by the first (MANSEG2). Individual T1-weighted MRI fused to PET sum image (both sessions together) was used as a reference. Thus, the ROI delineation was performed only once by both operators and the same ROIs (within operator) were employed to extract data from both sessions using co-registered PET-images (c.f. previous paragraphs). The anatomical landmarks described by Mawlawi, Martinez and colleagues [8;9] were applied in the ROI delineation, and all ROIs were drawn on coronal slices. The PET sum image was used as additional

guidance in the striatum border search to minimize PVE. Manual segmentation included five anatomical substructures per hemisphere, which were combined into the three functional substructures: the ventral striatum as such formed the LSTR, the pre-commissural portions of the caudate (CAUA) and putamen (PUTA) formed the ASTR, and the post-commissural portions of the caudate (CAUP) and putamen (PUTP) formed the SMST (see Table 2).

Atlas-based Segmentation

The atlas individualization process is illustrated in a diagram in Fig. 1. Three atlases were considered in the atlas-based striatal segmentation: the probabilistic structural atlases by Fischl and colleagues [28] and by Tziortzi and colleagues [10] and a connectivity based atlas by Tziortzi and colleagues [13]. The latter two are included in the FSL package (striatum-structural-2mm and striatum-con-label-thr50-7sub-2mm), whereas the first is a FreeSurfer built-in atlas. The structural atlas in FreeSurfer is based on automated labeling of several brains, whereas the structural FSL atlas was obtained by directly segmenting the MNI152 template according to the guidelines in [10]. The SB atlas segmentations yielded caudate (CAU), putamen (PUT) and nucleus accumbens (NACC) or ventral striatum (VST) ROIs, which were transformed into STR (all regions) and LSTR (NACC or VST) for comparison (see Table 2). The structural atlases were individualized within the host software packages only and are denoted as FSSEG for FreeSurfer and as STRUCTMRIF for FSL.

The connectivity-based atlas was generated using diffusion tensor imaging (DTI) and a tractography technique to find the projection territories in the striatum that innervate certain cortical regions [13]. Tziortzi and colleagues employed probabilistic tractography in 12 healthy male volunteers to investigate the projections between the cortex and striatum yielding connectivity-based (CB) subdivision of the striatum in each subject. As a result of the connectivity analysis, Tziortzi and colleagues generated atlases at different probability thresholds. For the current study we chose a 7 substructure atlas with a 2 mm voxel-size that was obtained at a 50% probability threshold (striatum-con-label-thr50-7sub-2mm). The 7 substructures were named by the cortical regions that they innervate: limbic, executive, rostral and caudal motor, parietal, temporal and occipital. For the CB atlas ROIs, the motor, parietal, temporal and occipital components were combined to form the SMST, whereas the executive component was

considered the ASTR and limbic was considered the LSRT (see Table 2). The connectivity atlas was individualized using the individualizing mappings as described above and are denoted as CONNMRIF, CONNMRIS and CONNPET for MRI-FSL, MRI-SPM and PET-SPM normalizations, respectively.

Direct PET Image Segmentation

Principle of the direct PET image segmentation process is illustrated in a diagram in Figure 1. An automatic PET image clustering algorithm employs the known heterogeneity of the dopamine receptor distribution [29] to find clusters of distinct ^{11}C-Raclopride binding in the striatum. Similar to the methods of Tohka et al. [30] and Farinha et al. [14], the algorithm consists of two main steps: 1) initial striatum extraction and 2) weighted kernel k-means clustering [31]. In the current study the initial step was implemented using a Markov Random Field (MRF) model-based extraction similar to that of other authors [30;32](see Document, Supplemental Digital Content 1, that describes the method MRF-based method), instead of a possibly more error prone deformable surface model approach we have used in [14]. Benefit of using MRF based striatum extraction lies in the inclusion of the PVE kernel in the image model inspired by MRI segmentation approaches [32]. Prior to the clustering step the BP_{ND} image was smoothed using an edge-preserving 3D Gaussian filter with FWHM=2.5mm and the result of the striatum extraction step as a striatal mask (see Fig. 1). In the second step, the weighted kernel k-means algorithm was applied to partition the connection graph into five segments using a geometrical division of the striatum into pre- and post-commissural caudate and putamen and ventral striatum as initialization, instead of a random initialization approach [14]. The automatic PET image segmentation results are denoted as PETSEG throughout the report. PETSEG yielded the same substructures as the manual segmentation; thus, the same combination strategy was used to obtain LSTR, ASTR and SMST (see Table 2).

Pharmacokinetic modeling methods

Pharmacokinetic modeling was performed for ROI time-activity course (TAC) data using SRTM [17]. STRM yields a binding potential of the tracer that is relative to the non-displaceable binding as measured in the reference tissue and is therefore denoted as BP_{ND} [18]. The ROI-TAC modeling was

conducted using in-house software fit_srtm version 3.0.7 for non-linear SRTM (Turku PET Centre, Turku, Finland). An atlas-based automated method for cerebellar ROI generation was chosen for the reference region TAC extraction. The reference region method was chosen on the basis of cerebellar distribution volume test-retest characteristics (see Table, Supplemental Digital Content 2, for cerebellar test-retest characteristics).

Statistical methods

Spatial agreement between MANSEG1 and each of the segmentation methods was determined at the ROI voxel set level, that is, the MANSEG1 ROIs were considered as a reference, although not necessarily ground truth. The ROI voxel set similarity was measured using the Jaccard coefficient expressed as the ratio between the size of the intersection and the size of the union of the voxel sets

$$J(A,B)=(|A\cap B|)/(|A\cup B|)$$

The Jaccard coefficient takes values between 0 (no agreement) and 1 (perfect match) and is a widely used performance measure for evaluating image segmentation algorithms.

Furthermore, agreement of the binding potentials (BP_{ND} as obtained through ROI-TAC SRTM-analysis) between MANSEG1 and each of the segmentation methods was determined using the limits of agreement (LOA), expressed as the mean of pairwise differences and 95% confidence interval. Moreover, the Pearson's product moment correlation coefficient was calculated between MANSEG1-BP_{ND} and each of the methods to assess the similarity of the BP_{ND} rank order.

The test-retest setting allowed method-independent estimation of the repeatability and reliability of each method. That is, instead of choosing one of the possibly erroneous methods as a reference, each method can be assessed independently. Therefore test-retest setting has been suggested for method comparison studies (see [16] for review). In the current study the test-retest characteristics were considered as the primary figure of merit in the evaluation of BP_{ND}-estimates (as obtained through various ROI-delineations), and the test-retest characteristics were calculated also for the ROI-volumes when applicable. Using MRI-based atlas-individualization methods there was only one ROI-set for each individual, thus no replication characteristics for ROI-volume could be formulated. While using the direct

PET-image segmentation, PET-based atlas individualization and manual segmentation there were two independent ROI-sets for each individual and formulation of the ROI volume test-retest characteristics was possible.

The repeatability of BP_{ND} and ROI-volume (when applicable) were calculated using the test-retest variability (TRV) [16], which relates the difference in the two measurement outcomes to that of their mean. The per cent TRV can be expressed as

$$TRV(\%)=100\% \times (2|X_1-X_2|)/(X_1+X_2)$$

where X_1 is the measurement outcome in PET-session 1 and X_2 is the measurement outcome in PET-session 2. The TRV(%) is reported as the mean±SD over the seven subjects for each method.

A common measure of test-retest reliability is intra-class correlation coefficient (ICC) that is based on the one-way random effects analysis of variance model, denoted as ICC(3,2) in Shrout and Fleiss [33]:

$$ICC=(MS_B - MS_w)/(MS_B +(k-1)MS_w)$$

where MS_B indicates the between subjects and MS_w is the within subjects mean of the sum of squared differences, and k=2. It is apparent that ICC is a relative measure of reliability and depends heavily on the current samples variability. Generally large between-subject variability (large MS_B) results in ICC closer to one (good reliability) although the within-subject repeatability might be poor. In method comparison studies this feature of the ICC contains a particular caveat due to the possibly method-dependent outcome of MS_B. That is, one of the methods can exaggerate the between-subject variability and thereby wrongly benefit from that as high ICC. In the current study the manual approach embodies a considerable subjective component that can result in exaggerated between-subject variability in the ROI-delineation, thus a more robust measure of reliability is required.

It has been noted that an absolute measure of reliability would be more appropriate in method comparison studies [16]. Moreover, an absolute measure would allow for the calculation of the confidence interval of the minimal detectable change (single subject). In the current study the standard error of measurement (SEM) was applied as an absolute measure of reliability, as expressed in [16]:

$SEM = SD\sqrt{(1-ICC)}$

where SD indicates the standard deviation of binding parameters from all subjects. The minimal detectable change (MD) was calculated on the basis of SEM for a confidence interval of 95% [16]

$MD = SEM * 1.96 * \sqrt{2}$

MD depicts how much two measurements of the same individual must differ from each other to be considered as true change, thus decrease in MD could be interpreted as improved sensitivity. MD(%) was calculated relative to the mean BP_{ND}.

The repeatability of the ROI volumes was calculated when applicable; for instance, PETSEG and CONNPET were made independently on PET1 and PET2, and manual segmentation was repeated by two operators. Furthermore, coefficients of variation were calculated from regional distribution volumes. Descriptive statistics include the arithmetic mean±SD and the coefficient of variation ($COV(\%) = 100\% S/\bar{X}$) for binding parameters.

Results

Agreement of ROI Volumes

Figure 2 illustrates the ROI delineations generated by the various methods (of the combined regions). There was reasonable agreement between the methods with regard to the location of the striatum. However, manual delineation (particularly MANSEG1) yielded systematically smaller volume for the whole striatum ROI as compared to the automated methods (see Figure, Supplemental Digital Content 3, which illustrates the method-wise volumes). Average volume of the whole striatum was 13 cm^3 for the manual segmentation (MANSEG1) while average volumes generated with automated methods were in the range of 16-21 cm^3. Direct PET segmentation was associated with smaller average striatal volumes as compared to other automated methods. Size (volume) of the whole striatum ROI may be associated with the impact of PVE, in particular, too liberate ROI delineation may yield exaggerated PVE. Indeed, it was notified that a strong negative linear correlation ($R=-0.93$; see Figure, Supplemental Digital Content 4, which shows the linear regression lines) existed between the average striatal volume (by method) and

average BP_{ND} (by method), indicating increased PVE with automated methods, in particular when using atlas-based methods.

Variability in the striatal volume between subjects was assessed using CoV(%) as presented in Table 3. Manual segmentation and direct PET segmentation yielded larger between subject variability as compared to atlas-based automated methods, with approximately two-fold CoV(%) for MANSEG1 as compared to that of atlas-based methods. Furthermore, the within subject variability in ROI volumes was calculated when applicable (see Table 3 TRV(%) Volume). The inter-operator variability in manual ROI volume was in the range of 16%-44% (average TRV(%)), while the direct PET image segmentation showed TRV(%) of 7%-17% and the CB atlas individualization using PET-normalization (CONNPET) as small as 2%-3% (average TRV(%)).

Illustration of the ROI contours in Figure 2 showed systematic inter-method differences in the definition of the limbic striatum ROIs that resulted in poor spatial agreement as measured using the Jaccard coefficient and Pearson's correlation coefficient (see Table 3) relative to the manual segmentation (MANSEG1) , while the agreement was somewhat better in the associative and sensorimotor striatum. In particular, the direct PET image segmentation and CB atlas methods yielded large deviation from the SB limbic striatum ROIs generated either manually or using atlas individualization.

Agreement of the BP_{ND} Estimates

Inter-method variability in ROI volumes was manifested in BP_{ND} estimates. As was noted earlier the average BP_{ND} estimate was strongly associated with the striatal volume in the whole striatum, and also in the associate striatum, but not in the limbic or sensorimotor striatum (see Figure, Supplemental Digital Content 4, which shows the linear regression lines). Thus the differences in BP_{ND} estimates in the limbic and sensorimotor striatum were not mainly driven by the differences in PVE. On the other hand, the Pearson's correlation coefficient showed strong inter-method correlation between the BP_{ND} estimates (except for CONNMRIF in SMST, see Table 3), whereas, the LOA range (see Table 3) showed both negative and positive systematic biases relative to MANSEG1. Even between manual segmentations significant differences were seen in the associative and sensorimotor striata.

Repeatability and Reliability of the BP_{ND} estimates

Intra-method within subject variability of the BP_{ND} was smallest using CONNPET (see TRV(%) (BP_{ND}) in Table 3) followed by MANSEG1 and MANSEG2. However, differences in between subject variability (repeatability) were mostly small, while that of PETSEG was somewhat larger. The repeatability of PETSEG and, subsequently, its reliability was ruined by a single outlier scan with a markedly lower dose (3.3 MBq/kg compared with the group mean±SD of 5.6±1.6 MBq/kg) in the second scan. Although the outlying subject had exceptionally poor quality PET data, the sensitivity of PETSEG with regard to noise must be acknowledged.

Intra-method between subject variability of the BP_{ND} was smallest using CONNMRIF (see CoV(%) (BP_{ND}) in Table 3) followed by CONNPET. There was up to 61% increase in CoV(%) between the smallest (CONNMRIF) and largest variability (MANSEG1). On the other hand, the large between subject variability resulted in high ICC, as expected. The highest ICC was obtained using MANSEG1 followed by MANSEG2 and CONNPET. The SEM, however, indicated a different order in performance – the smallest SEM (best reliability) was obtained using CONNPET followed by CONNMRIF and then manual segmentation, although the differences were not very large. Only PETSEG showed markedly higher SEM as compared to the other methods, due to one outlier scan with significantly poorer image quality. The average minimal detectable change (MD) was close to 10% (smallest using CONNPET, see Table 3) except for PETSEG (15%).

Test-retest characteristics within the initial subregions are presented in the Table, Supplemental Digital Content 5. For the manual delineation and direct PET image segmentation the five structural striatal subregions showed mostly similar test-retest characteristics as the combined functional subregions, except for the posterior caudate which showed somewhat poorer performance for MANSEG1 and PETSEG as compared to the combined regions. For the CB atlas ROIs the five subregions that constituted the sensorimotor striatum ROI showed varying test-retest performance that depended on the ROI size. The largest subregions (caudal and rostral motor, and parietal) showed comparable test-retest performance with the combined regions, while the smallest subregions (occipital and temporal) showed poorer performance.

Discussion

This study compared several ROI-delineation methods for high-resolution ^{11}C-Raclopride-PET. Manual ROI-delineation based on rules by Mawlawi and colleagues [8] and Martinez and colleagues [9] can be considered as a gold standard due to its widespread usage. However, the manual ROI delineation approaches are associated with considerable operator-dependent variability and high cost, and alternative methods are constantly developed. Popularity of automated methods for functional striatal subdivision has been limited due to complexity of the segmentation task. Nevertheless, few previous studies have shown feasibility of automated segmentation within striatum. For the current study atlas-based approach and direct PET image clustering algorithms were implemented based on promising results in previous studies [11;14]. In the current study the automated methods were not only compared against the manual ROI-delineation as is commonly done, but the methodological performance was primarily assessed using test-retest protocol. The test-retest protocol has been suggested for method comparison studies (see [16] for review) for its method-independent performance evaluation. In addition, secondary analysis was made through comparison with the manual ROI-method to allow descriptive analysis of the method performance.

Manual segmentation

For the data presented in this study, the inter-operator variability in manual ROI-delineation was considerable, yielding statistically significant differences between BP_{ND} estimates (Table 3). Although the linear regression analysis showed good agreement of the BP_{ND} estimates the LOA range indicated statistically significant inter-operator differences in associative and sensorimotor striatum but not in the limbic striatum (Table 3). Inter-operator differences may be attributed to subjective striatum-border search, that is, the fused PET and MRI data can be interpreted differently according to operators experience and habits. The rules of Mawlawi and Martinez and colleagues describe the structural cues based on T1W MRI-data only, but it has been thought that using PET-image fusion as a guideline would help to protect against unwanted PVE. Apparently, for the data presented in this study the more experienced operator (MANSEG1) was more conservative with regard to the striatum border placement than the novice operator (MANSEG2). This was indicated by the systematically smaller ROI-volumes of

the experienced operator as compared to the beginner (see Figure, Supplemental Digital Content 3, which illustrates the striatal volumes). Smaller ROIs of the experienced operator yielded systematically higher BP_{ND} estimates, thus implying lesser PVE-contamination in MANSEG1 ROIs. In principle, lesser PVE-contamination is a desirable property, but often PVE-compensation methods are associated with increased methodological variation. For the data presented in this study the between subject variability for both the ROI volume and BP_{ND} estimates was highest using MANSEG1 ROIs followed by MANSEG2. In the absence of ground truth data it can not be unequivocally inferred whether the increased variability is due to true variability between subjects that is more accurately measured using smaller ROIs or whether it is more due to erroneous variability in ROI-delineations. As was discussed before, the commonly employed test-retest reliability measure of ICC strongly favors large variability between subjects, be it erroneous or true variation. Consequently the ICC measures for manual segmentation were superior to those using automated methods, and superior using MANSEG1 ROIs as compared to using MANSEG2 ROIs. However, the performance rank order was rather different according to the SEM, which is an absolute measure of reliability taking into account the standard deviation of the BP_{ND} estimates, that is the between subject variability. The SEM was marginally superior using MANSEG2 ROIs and more so using the atlas-based automatic ROIs in comparison with MANSEG1. SEM has been suggested for measuring the reliability instead of ICC as it is less vulnerable to increased methodological variation (see [16] for review). Thus, according to the SEM the more conservative border placement of the more experienced operator may have been associated with increased methodological variation albeit the impact of PVE was likely smaller.

Atlas-based methods

The atlas-based methods showed good overall performance; there were no complete failures and general agreement with manual delineation was reasonable. Albeit the ROIs did not match exactly with the manual segmentation (see Jaccard and Pearson's correlation coefficient in Table 3), the BP_{ND} estimates showed strong correlation in the linear regression analysis (except for CONNMRIF in SMST). Thus, for the data presented in this study the atlas-based automatic methods were able to replicate the between subject rank order of BP_{ND} estimates of the manual ROI-delineation, whereas, the LOA range

showed statistically significant differences between the BP_{ND} estimates. In the limbic striatum the SB atlas ROIs showed small underestimation (FSSEG) or no bias (STRUCTMRIF) while the CB atlas ROIs showed overestimation (CONNMRIF) or non-significant bias (CONNPET) as compared to MANSEG1, whereas, in the associative and sensorimotor striata the CB atlas ROIs showed systematic underestimations as compared to MANSEG1. As was discussed earlier, differences in the associative and sensorimotor striata may be mostly attributed to the ROI size rather than gross differences in the ROI placement, whereas, there was a clear difference in the limbic striatum ROI placement between SB and CB ROIs. That is, the placement of limbic striatum ROI was similar using manual ROI-delineation and SB atlas ROIs but the CB atlas ROIs were placed more towards caudal ventral putamen and less towards rostral dorsal caudate (see Figure 2 for ROI contours). Thus a visible difference in the LSTR ROI placement was demonstrated using CB atlas as compared to structure-based approaches. The data presented here can not be used to make inferences whether the SB or CB approach would yield limbic striatum ROIs that better delineates the functional organization. However, the data presented in this study shows feasibility of the novel CB atlas in automated ROI-analysis, while the analysis made by Tziortzi and colleagues [13] suggests improved regional selectivity of CB ROIs over the SB ROIs for measuring the dopamine response.

Impact of Normalization Method in Atlas-based Approach

The data presented in this study suggests flexibility with regard to the choice of normalization procedure. In the current study we employed individual T1W MRI-data as well as PET-data to find a mapping from the atlas defining space to the individual space where the ROIs were analyzed. The individualizing mappings based on MRI-data were generated using state-of-the-art algorithms implemented in SPM8 and FSL5 for non-rigid image matching. Differences between the ROIs from SPM8 and FSL5 were small (see Table, Supplemental Digital Content 5, for sub-striatal BP_{ND} in initial regions). Furthermore, the normalizing (inverse of individualizing) mappings from FSL5 MRI-matching were exploited to generate an ^{11}C-Raclopride specific PET-template. The PET-template was generated using the PET sum-images from the current study, and the PET-based individualizing mappings were estimated using an algorithm implemented in SPM8. Albeit the differences were not large, the PET-based

individualization of the CB-atlas performed better than MRI-based in the test-retest evaluation. Both the within subject variability (TRV(%) see Table 3) and SEM were slightly superior using PET-based mappings. It can be discussed whether the PET-based normalization would have benefited from that the PET-template was made using the same data as was used to find the individualizing mappings. The concern is however alleviated by the fact that fourteen PET sum-images were averaged in the template formation, thus comparable performance could be expected for independent datasets. It has been noted also by others that the striatum is a challenging target for MRI-based normalization, and PET-based normalization has been suggested instead [34].

Direct PET image segmentation

Direct PET image segmentation algorithm performed well in general but in a single occasion precision of the initial striatum extraction was not adequate. Closer analysis of the data revealed visible difference in image quality of the outlier scan as compared to the others, attributed to significantly lower radioactive dose as compared to average dose (3.3 MBq/kg compared with the group mean±SD of 5.6±1.6 MBq/kg). Various parameter combinations were tested to improve the segmentation outcome, however without significant improvement. Applying larger kernel size (FWHM>5 mm) in the pre-segmentation smoothing step (c.f. Figure 1) would have allowed extraction of the striatum more robustly, but at the cost of inaccurate segmentation particularly in the thin structure of rostral caudate nucleus. Therefore, it was deemed necessary not to increase the smoothing kernel size in order to maintain the benefit from high-resolution PET imaging. Otherwise the initial striatum extraction and subsequent clustering performed well, but sensitivity of the current method with regard to image noise must be acknowledged. Conservative smoothing combined with the MRF-based resolution modeling in the striatum extraction yielded small average striatal volume as compared to other automated methods. The average striatal volume from PETSEG ROIs was second smallest after MANSEG1 suggesting excellent performance with regard to protection against PVE. Unfortunately, for the data presented here the outlier scan ruined the overall test-retest performance of PETSEG, but the method showed great potential that might be exploited in future studies. For instance, combining (averaging) the BP_{ND} images of the same subject (if several scans made) would enhance the image SNR considerably and the outcome of direct PET image

segmentation would be more reliable. In the current study, however, the purpose was to investigate the replicability of the segmentations. A remaining question is whether the direct image clustering is sensitive to changes in BP_{ND} distribution. It has been shown that differences in the response to amphetamine-induced stimulation exist between striatal subregions [9;13], but the direct PET image segmentation has not been applied to such data. The study of Egerton and colleagues [4] showed potential risk of using direct PET image segmentation in a comparison between baseline and task-induced stimulation when the ROIs were generated independently. The potential hazard in changing uptake distribution assessments is related to changing ROI definitions over sessions and thereby under- or overestimation of the activation effect. Thus, it might be advisable to combine the BP_{ND} data prior to image clustering not only for improved SNR but also for bias-free ROI-analysis. Another remaining question is whether the direct PET image clustering yields subregions that are relevant with regard to functional organization of the striatum. It is often assumed that brain anatomy follows function, or in this case the receptor populations follow function, but the correspondence is seldom one-to-one. The data presented in this study showed visible difference in the PETSEG limbic striatum ROIs as compared to both SB and CB delineations (see Figure 2 for substriatal contours). Different initializations were tested (including manual ROIs) but the clustering outcome was nearly identical, implying that there is a strong gradient in the BP_{ND} estimates within ventral portions of the rostral caudate and putamen that defined the LSTR cluster. The PETSEG LSTR ROIs extended more dorsally within caudatus as compared to other methods, and less caudally within putamen as compared to CB ROIs, whereas, the associative and sensorimotor striatum ROIs matched well with the CB ROIs (except for the rostral caudatus). It will remain for the future studies to show whether and how well the gradients in BP_{ND} distribution correspond with response to dopamine-stimulation. Furthermore, it might be worthwhile to pursue the possibilities of direct PET image clustering in the investigations perplexed by changes in the striatal volume per se. It has been shown using 11C-Raclopride that there is an age-related decline in D2-receptor density on the one hand [35], and decrease of the striatal volume as measured using MRI on the other hand [36]. Consequently, interplay between the striatal volume loss and decline of the D2-receptor binding has been acknowledged as a potential confound due to increased PVE [Morris1999]. In

the study by Morris and colleagues [37] PVE compensation was employed for the PET data and clearly slower decline in D2-receptor density as compared to earlier studies was shown, implying a confound due to inadequate PET analysis in some earlier studies [37]. Thus, the direct PET image segmentation might be a viable option for improving the PET analysis in aging research.

Choice between ROI-methods in the striatum

As was discussed earlier, the outcome from manual ROI-delineations was possibly less contaminated by PVE as compared to the atlas-based methods and to some degree compared to direct PET image segmentation, but likely at the cost of higher methodological variation. In addition, there was significant inter-operator variability that would effectively prevent multi-operator ROI-delineation in high-resolution ^{11}C-Raclopride studies. Albeit the intra-operator variability was not directly measured in the current study the elevated between-subject variability suggests variation in manual ROI-delineations within operator. Be it true or erroneous, a slightly elevated variation of BP_{ND} was observed also between repetitions using manual ROIs. Test-retest within-subject variability is commonly thought to reflect the relative (BP_{ND}) change that can be detected using a given method. On the other hand, small within-subject variability can be obtained without high specificity, curbing the sensitivity of the assessment. Thus, it is important to consider the interplay of the within- and between-subject variability in the method performance evaluation. For the data presented here the macro-parameter MD(%) (see Table 3) was calculated on the basis of SEM to designate the per cent change in individual BP_{ND} estimates that can be regarded as true change. Thus, smaller MD(%) could be interpreted as improved sensitivity. For the data presented here the average MD(%) was close to 10% in combined regions for all methods except PETSEG (15%), with smallest average using CB-atlas and PET-based normalization (see Table 3), thus implying that at individual level the sensitivity of other methods except PETSEG were very similar, albeit the impact of PVE was likely different. Furthermore, in group-level analysis the BP_{ND} between-subject variability (standard deviation of BP_{ND} estimates over subjects) is often used in the statistical power calculations in place of true cohort variability. That is, the BP_{ND} SD and mean BP_{ND} (see Table 3) are commonly employed in the sample size calculations. We used G*Power (version 3.1.9.2, Universität Kiel, Germany) to calculate the effect size on the basis of mean and SD of BP_{ND} in Table 3, and

subsequently the total sample size for detection of 10% decrease in group mean BP_{ND} with one-tailed paired t-test (with α=0.05, and power=0.95). For the data presented here the sample size was 8-10 subjects using CONNPET, while for the manual segmentation the minimum sample sizes were 13-25 (MANSEG1) and 9-23 (MANSEG2) subjects. Thus, the CB-atlas method showed superior sensitivity over manual ROI-delineation at both the individual- and group-level analysis. However the data presented here does not provide direct validation of the methods, the relative improved consistency shown here for the CB-atlas method suggests potential benefit over manual ROI-delineation.

As was discussed earlier the PET-based normalization might have offered some improvement in the atlas-individualization as compared to more commonly employed MRI-based normalization. Striatum is a demanding target for MRI-based normalization due to mixture of gray and white matter cells (thus the name striatum) and PET-based normalization has been acknowledged as a viable alternative [34]. In the current study, however, the PET data was acquired using a high-resolution PET scanner and the PET-template was generated using the same dataset, rendering the generalization of the current results rather difficult. That is, poor spatial resolution of the PET data from conventional scanners might hamper the PET-based normalization to a degree that would favor the usage of high-resolution MRI-based normalization instead. On the other hand, the source and target data in PET-normalization were pre-smoothed using a 4 mm (FWHM) kernel size that can be regarded as a means of resolution matching (c.f. [38]). The ^{11}C-Raclopride-template generated in the current study can be requested from the corresponding author, in case its applicability with data from other scanners needs to be tested, or otherwise interested. In certain circumstances also the direct PET image segmentation algorithm may be an appealing alternative, although validity of the (BP_{ND}) gradient-based clustering may require further validation. Implementation of the direct PET image clustering algorithm (in Matlab) can be requested from the corresponding author.

Reference Region ROI Generation

In addition to the striatal segmentation the cerebellar ROI-delineation was evaluated in the current study because of its broad usage in the analysis of ^{11}C-Raclopride-PET and reference-tissue-based modeling (see Table, Supplemental Digital Content 2, for cerebellar test-retest characteristics).

Currently, the manual ROI-delineation can be considered as a gold standard for its widespread usage, but as for the striatal ROI-drawing there can be substantial differences between operators' opinions. On the other hand, differences in the ROI-delineations per se may have lesser significance within the cerebellum where the ^{11}C-Raclopride uptake is non-specific, but the noise-characteristics of the reference-tissue-TAC become likely more important due to the model fitting process. Manual ROI-delineation employs T1W MRI-data in the search of cerebellar cortex border typically from transaxial images, frequently with the help of fused PET data to avoid blood-rich territories such as veins. The process of manual cerebellar cortex delineation is laborious and it is therefore often limited to very few transaxial slices resulting in small ROIs and noisy TACs. Thus, robust automated method for cerebellar ROI-delineation would be highly beneficial for improved reference-tissue-based modeling. Algorithms implemented in the FreeSurfer software have become a popular choice in automated ROI-generation, but in a recent study by Schain and colleagues [39] the cerebellar FreeSurfer ROI was deemed suboptimal for reference-tissue-based modeling. In their study the comparison was however made only to manual segmentation and not using method-independent measure such as the test-retest characteristics in the current study. Furthermore, their suggestion was that the performance of the auto-ROI generation would be ligand-dependent [39]. For the data presented here the average cerebellar FreeSurfer ROI covered over ten-fold larger volume as compared to MANSEG1, and approximately six-fold larger volume as compared to atlas-based ROIs (see Table, Supplemental Digital Content 2, for cerebellar ROI characteristics). The atlas-based ROIs were deliberately limited to a specific cerebellar subregion explaining the vast difference with FreeSurfer ROIs. Albeit the ROI volumes varied extensively the outcome of full pharmacokinetic modeling with arterial input (distribution volume) showed strong correlation between the methods (see Table, Supplemental Digital Content 2, for cerebellar ROI characteristics), implying small role of the ROI delineation in the quantitation of non-specific uptake TAC. However, the repeatability of the outcome was superior using the larger ROIs of automated methods as compared to those from manual segmentation (see Table, Supplemental Digital Content 2, for TRV(%)), suggesting a benefit from improved SNR. In the current study we did not notice any bias from additional spill-out signal originating from blood-rich areas using the automated methods,

likely due to different characteristics of 11C-Raclopride uptake as compared to ligands tested by Schain and colleagues [39].

Conclusions

The data presented in this study supports implementation of fully automated ROI-generation for ^{11}C-Raclopride-PET analysis. It was demonstrated that ROIs generated using connectivity-based atlas individualization can yield BP_{ND} estimates in substriatal domains that highly correlate with those obtained using manual ROIs. Furthermore, the ROIs generated automatically from CB atlas showed superior test-retest characteristics for both within- and between-subject variability, and consequently enhanced sensitivity for BP_{ND} alterations both at individual- and group-level as compared to manual ROIs. In addition, automated reference-region ROI generation was superior to that by manual ROI-delineation according to improved test-retest characteristics.

The CB substriatal atlas and cerebellar atlas individualization were successful using either MRI- or PET-based normalization. The PET-based normalization was slightly superior to that by MRI-normalization in the striatum possibly partly due to high-resolution of the PET-data presented here. However, the data presented here supports experimentation with fully PET-based automated ROI-generation using other scanners as well. While the direct PET image clustering algorithm showed great potential it should be applied with caution – robustness of the current implementation was questioned here by a failure to segment a single poor quality image. If the image quality can be guaranteed through for instance BP_{ND} averaging the direct PET image segmentation might provide an interesting alternative.

It will remain for the future studies to show whether the CB definition of the striatal subdomains, or those based on direct PET image segmentation can provide similar or possibly improved ROI-delineations with regard to true functional organization of the striatum, as compared to those obtained using conventional manual ROI-drawing. Furthermore, similar comparison is in place using different subject groups that might show larger variability in the striatal anatomy. The data presented here is representative for young healthy male subjects and the performance may be slightly different for subjects with anomalies or for instance enlarged ventricles.

Reference List

[1] Farde L, Hall H, Ehrin E, Sedvall G. Quantitative analysis of D2 dopamine receptor binding in the living human brain by PET. Science 1986 Jan 17;231(4735):258-61.

[2] Farde L, Nordstrom AL. PET analysis indicates atypical central dopamine receptor occupancy in clozapine-treated patients. Br J Psychiatry Suppl 1992 May;(17):30-3.

[3] Koepp MJ, Gunn RN, Lawrence AD, Cunningham VJ, Dagher A, Jones T, et al. Evidence for striatal dopamine release during a video game. Nature 1998 May 21;393(6682):266-8.

[4] Egerton A, Mehta MA, Montgomery AJ, Lappin JM, Howes OD, Reeves SJ, et al. The dopaminergic basis of human behaviors: A review of molecular imaging studies. Neurosci Biobehav Rev 2009 Jul;33(7):1109-32.

[5] Birren JE, Schaie KW. Handbook of the psychology of aging. 5. ed ed. San Diego, Calif: Academic Press; 2001.

[6] Koob GF, Bloom FE. Cellular and molecular mechanisms of drug dependence. Science 1988 Nov 4;242(4879):715-23.

[7] Joutsa J, Johansson J, Niemela S, Ollikainen A, Hirvonen MM, Piepponen P, et al. Mesolimbic dopamine release is linked to symptom severity in pathological gambling. Neuroimage 2012 May 1;60(4):1992-9.

[8] Mawlawi O, Martinez D, Slifstein M, Broft A, Chatterjee R, Hwang DR, et al. Imaging human mesolimbic dopamine transmission with positron emission tomography: I. Accuracy and precision of D(2) receptor parameter measurements in ventral striatum. J Cereb Blood Flow Metab 2001 Sep;21(9):1034-57.

[9] Martinez D, Slifstein M, Broft A, Mawlawi O, Hwang DR, Huang Y, et al. Imaging human mesolimbic dopamine transmission with positron emission tomography. Part II: amphetamine-induced dopamine release in the functional subdivisions of the striatum. J Cereb Blood Flow Metab 2003 Mar;23(3):285-300.

[10] Tziortzi AC, Searle GE, Tzimopoulou S, Salinas C, Beaver JD, Jenkinson M, et al. Imaging dopamine receptors in humans with [11C]-(+)-PHNO: dissection of D3 signal and anatomy. Neuroimage 2011 Jan 1;54(1):264-77.

[11] del CN, Tait RJ, Acosta-Cabronero J, Hong YT, Izquierdo-Garcia D, Smith R, et al. Quantification of receptor-ligand binding potential in sub-striatal domains using probabilistic and template regions of interest. Neuroimage 2011 Mar 1;55(1):101-12.

[12] Tohka J, Wallius E, Hirvonen J, Hietala J, Ruotsalainen U. Automatic Extraction of Caudate and Putamen in 11C-Raclopride PET Using Deformable Surface Models and Normalized Cuts. Nuclear Science, IEEE Transactions on 2006 Feb;53(1):220-7.

[13] Tziortzi AC, Haber SN, Searle GE, Tsoumpas C, Long CJ, Shotbolt P, et al. Connectivity-based functional analysis of dopamine release in the striatum using diffusion-weighted MRI and positron emission tomography. Cereb Cortex 2014 May;24(5):1165-77.

[14] Farinha RJ, Ruotsalainen U, Hirvonen J, Tuominen L, Hietala J, Fonseca JM, et al. Segmentation of striatal brain structures from high resolution PET images. Int J Biomed Imaging 2009;2009:156234.

[15] Alakurtti K, Aalto S, Johansson JJ, Nagren K, Tuokkola T, Oikonen V, et al. Reproducibility of striatal and thalamic dopamine D2 receptor binding using [11C]raclopride with high-resolution positron emission tomography. J Cereb Blood Flow Metab 2011 Jan;31(1):155-65.

[16] Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005 Feb;19(1):231-40.

[17] Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. Neuroimage 1996 Dec;4(3 Pt 1):153-8.

[18] Innis RB, Cunningham VJ, Delforge J, Fujita M, Gjedde A, Gunn RN, et al. Consensus nomenclature for in vivo imaging of reversibly binding radioligands. J Cereb Blood Flow Metab 2007 Sep;27(9):1533-9.

[19] de Jong HW, van Velden FH, Kloet RW, Buijs FL, Boellaard R, Lammertsma AA. Performance evaluation of the ECAT HRRT: an LSO-LYSO double layer high resolution, high sensitivity scanner. Phys Med Biol 2007 Mar 7;52(5):1505-26.

[20] Comtat C, Bataille F, Michel C, Jones JP, Sibomana M, Janeiro L, et al. OSEM-3D reconstruction strategies for the ECAT HRRT. Nuclear Science Symposium Conference Record, 2004 IEEE 2004 Oct 16;6:3492-6.

[21] Comtat C, Sureau FC, Sibomana M, Hong IK, Sjoholm N, Trebossen R. Image based resolution modeling for the HRRT OSEM reconstructions software. Nuclear Science Symposium Conference Record, 2008 NSS '08 IEEE 2008 Oct 19;4120-3.

[22] Nuyts J, Dupont P, Stroobants S, Maes A, Mortelmans L, Suetens P. Evaluation of maximum-likelihood based attenuation correction in positron emission tomography. Nuclear Science, IEEE Transactions on 1999 Aug;46(4):1136-41.

[23] Watson CC. New, faster, image-based scatter correction for 3D PET. Nuclear Science, IEEE Transactions on 2000 Aug;47(4):1587-94.

[24] Byars LG, Sibomana M, Burbar Z, Jones J, Panin V, Barker WC, et al. Variance reduction on randoms from coincidence histograms for the HRRT. Nuclear Science Symposium Conference Record, 2005 IEEE 2005 Oct 23;5:2622-6.

[25] Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005 Jul 1;26(3):839-51.

[26] Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. Neuroimage 2012 Aug 15;62(2):782-90.

[27] Ashburner J, Friston KJ. Nonlinear spatial normalization using basis functions. Hum Brain Mapp 1999;7(4):254-66.

[28] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002 Jan 31;33(3):341-55.

[29] Alakurtti K, Johansson JJ, Tuokkola T, Nagren K, Rinne JO. Rostrocaudal gradients of dopamine D2/3 receptor binding in striatal subregions measured with [(11)C]raclopride and high-resolution positron emission tomography. Neuroimage 2013 Nov 15;82:252-9.

[30] Juslin A, Tohka J. Unsupervised Segmentation of Cardiac PET Transmission Images for Automatic Heart Volume Extraction. Engineering in Medicine and Biology Society, 2006 EMBS '06 28th Annual International Conference of the IEEE 2006 Aug 30;1077-80.

[31] Dhillon IS, Guan Y, Kulis B. Weighted graph cuts without eigenvectors a multilevel approach. IEEE Trans Pattern Anal Mach Intell 2007 Nov;29(11):1944-57.

[32] Tohka J, Kivimaki A, Reilhac A, Mykkanen J, Ruotsalainen U. Assessment of brain surface extraction from PET images using Monte Carlo Simulations. Nuclear Science, IEEE Transactions on 2004 Oct;51(5):2641-8.

[33] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-8.

[34] Kuhn FP, Warnock GI, Burger C, Ledermann K, Martin-Soelch C, Buck A. Comparison of PET template-based and MRI-based image processing in the quantitative analysis of C11-raclopride PET. EJNMMI Res 2014;4(1):7.

[35] Rinne JO, Hietala J, Ruotsalainen U, Sako E, Laihinen A, Nagren K, et al. Decrease in human striatal dopamine D2 receptor density with age: a PET study with [11C]raclopride. J Cereb Blood Flow Metab 1993 Mar;13(2):310-4.

[36] McDonald WM, Husain M, Doraiswamy PM, Figiel G, Boyko O, Krishnan KR. A magnetic resonance image study of age-related changes in human putamen nuclei. Neuroreport 1991 Jan;2(1):57-60.

[37] Morris ED, Chefer SI, Lane MA, Muzic RF, Jr., Wong DF, Dannals RF, et al. Loss of D2 receptor binding with age in rhesus monkeys: importance of correction for differences in striatal size. J Cereb Blood Flow Metab 1999 Feb;19(2):218-29.

[38] van Velden FH, Kloet RW, van Berckel BN, Buijs FL, Luurtsema G, Lammertsma AA, et al. HRRT versus HR+ human brain PET studies: an interscanner test-retest study. J Nucl Med 2009 May;50(5):693-702.

[39] Schain M, Varnas K, Cselenyi Z, Halldin C, Farde L, Varrone A. Evaluation of two automated methods for PET region of interest analysis. Neuroinformatics 2014 Oct;12(4):551-62.

List of Supplemental Digital Content

- Supplemental Digital Content 1. Document that describes the MRF-based striatum extraction.
- Supplemental Digital Content 2. Table that present cerebellar ROI-volumes, and distribution volume test-retest characteristics.
- Supplemental Digital Content 3. Figure of bar-graphs of striatal ROI-volumes, and corresponding BP_{ND} estimates.
- Supplemental Digital Content 4. Figure that presents the linear correlation analysis between method-wise average striatal volumes and corresponding BP_{ND} estimates.
- Supplemental Digital Content 5. Table that presents the BP_{ND} estimates and test-retest characteristics in the initial (non-combined) ROIs.
- Supplemental Digital Content 6. Figure that illustrates the PET-template.

List of Tables and Figures

and transaxial projections represent combination from multiple cutting-planes, to show all structures at once.

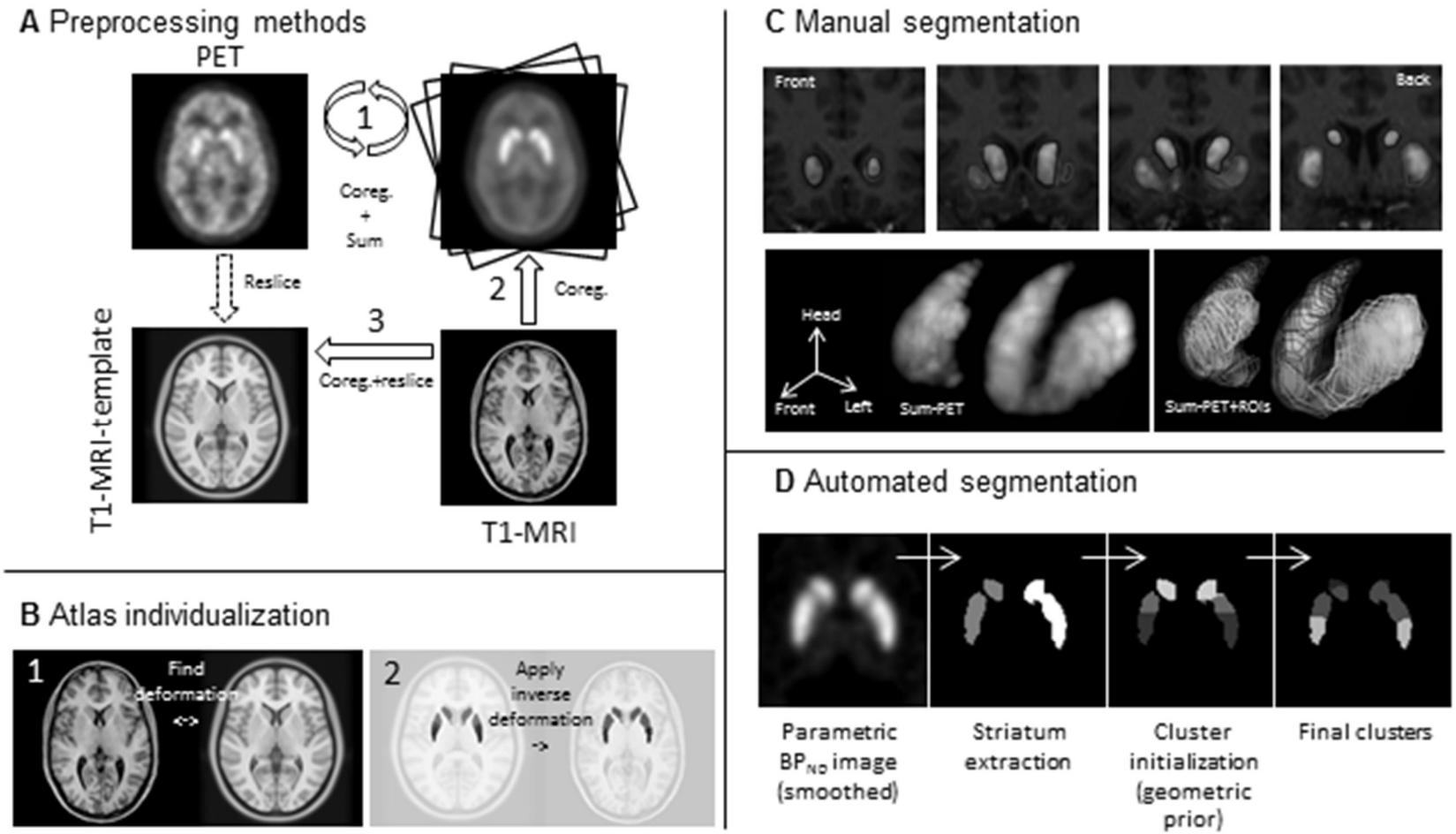
Method	Source	Summary	Atlas/guidelines
Manual segmentation (MANSEG1/2)	PET+MRI	Manual ROI delineation on fused PET+MRI	No atlas. Guidelines in (Mawlawi et al., 2003) were applied
Automatic striatum clustering (PETSEG)	PET	Two-step direct segmentation: 1) striatum extraction using MRF 2) clustering using k-means	No atlas. Initialization of k-means clustering based on location.
Structure-based atlas individualization with FreeSurfer (FSSEG)	MRI	Atlas-based segmentation provided in FreeSurfer	Probabilistic structural atlas (Fischl et al., 2002)
Structure-based atlas individualization with FSL (STRUCTMRIF)	MRI	Structural atlas individualization based on FSL MRI normalization	Segmentation of MNI152 template (Tziortzi et al., 2011)
Connectivity-based atlas individualization/FSL (CONNMRIF)	MRI	Connectivity atlas individualization based on FSL MRI normalization	DTI tractography analysis between cortical regions and striatum provided 7 substructures (Tziortzi et al., 2014)
Connectivity-based atlas individualization with SPM8 and MRI (CONNMRIS)	MRI	Connectivity atlas individualization based on SPM8 MRI normalization	
Connectivity atlas individualization with SPM8 and PET (CONNPET)	PET	Connectivity atlas individualization based on SPM8 PET normalization	

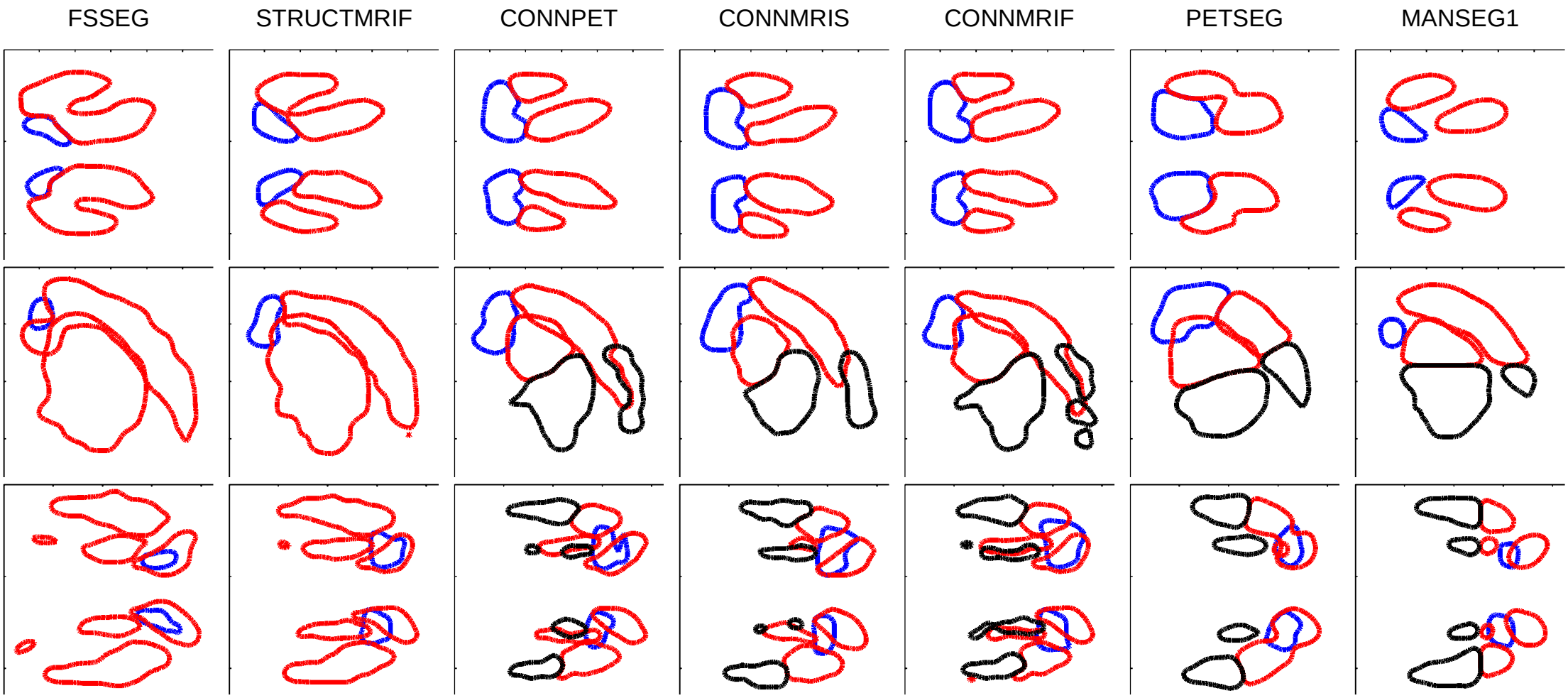
Method Region	MANSEG1, MANSEG2, PETSEG*	FSSEG, STRUCTMRIF*	CONNMRIF, CONNMRIS, CONNPET*
Limbic striatum (LSTR)	VST	NACC/VST	LSTR
Associative striatum (ASTR)	CAUA+PUTA	-	EXE
Sensorimotor striatum (SMST)	CAUP+PUTP	-	CAM+ROM+PAR+TEM+OCC
Caudate (CAU)	CAUA+CAUP	CAU	-
Putamen (PUT)	PUTA+PUTP	PUT	-
Anterior caudate (CAUA)	CAUA	-	-
Posterior caudate (CAUP)	CAUP	-	-
Anterior putamen (PUTA)	PUTA	-	-
Posterior putamen (PUTP)	PUTP	-	-
Ventral striatum (VST)	VST	VST (FSL)	-
Nucleus accumbens (NACC)	-	NACC (FS)	-
Executive cortex ter. (EXE)	-	-	EXE
Caudal motor cortex ter. (CAM)	-	-	CAM
Rostral motor cortex ter. (ROM)	-	-	ROM
Parietal cortex ter. (PAR)	-	-	PAR
Temporal cortex ter. (TEM)	-	-	TEM
Occipital cortex ter. (OCC)	-	-	OCC

* MANSEG1/2=manual segmentation, PETSEG=automatic PET image clustering, FSSEG=FreeSurfer segmentation, STRUCTMRIF=structural atlas segmentation in FSL, CONNMRIF/MRIS/PET=connectivity-based atlas individualized using FSL-MRI, SPM-MRI or SPM-PET image normalization.

		Volume		Agreement (Volume)		BP_{ND}		TRV(%) (BP_{ND})		Reliability (BP_{ND})			Agreement (BP_{ND})	
	Method	CoV(%)	TRV(%)	Jaccard	R	Mean±SD	CoV(%)	Mean±SD	Range	ICC	SEM	MD(%)	LOA (range)	R
Limbic Striatum	MANSEG1	18				3.76±0.55	14.6	4.55±3.19	-1.8-9.1	0.94	0.14	10.2		
	MANSEG2	16.3	21.1±12.8	0.46	0.26	3.65±0.51	13.9	4.66±5.32	-0.6-15.7	0.9	0.16	12.4	[-0.28,0.06]	0.95
	PETSEG	17.7	17.4±17.6	0.16	-0.03	4.49±0.47	10.6	7.28±3.50	-11.6-11.2	0.73	0.24	15.1	[0.46,1.00]	0.87
	STRUCTMRIF	8.3		0.33	-0.48	3.62±0.32	8.8	4.36±2.73	-4.9-7.6	0.86	0.12	9.2	[-0.41,0.13]	0.95
	CONNMRIF	8.2		0.21	-0.54	4.05±0.41	10.2	4.30±2.50	-6-9.3	0.9	0.13	8.9	[0.09,0.49]	0.96
	CONNPET	10.6	2.6±1.7	0.2	-0.34	3.96±0.35	8.9	3.42±1.81	-5.4-5.6	0.92	0.1	7	[-0.04,0.44]	0.95
	FSSEG	9.4		0.18	0.08	3.24±0.30	9.2	5.33±5.44	-10.2-15.6	0.74	0.15	13.1	[-0.87,-0.17]	0.81
Associative Striatum	MANSEG1	14.3				5.07±0.50	9.8	5.42±3.51	-9-9	0.81	0.22	11.9		
	MANSEG2	14.2	16.3±11.1	0.61	0.7	4.65±0.38	8.1	4.92±3.83	-8.5-9.1	0.74	0.19	11.5	[-0.58,-0.24]	0.96
	PETSEG	13.2	7.2±3.5	0.51	0.67	4.98±0.50	10	7.74±5.05	-14.1-12.2	0.63	0.31	17	[-0.22,0.04]	0.97
	CONNMRIF	7.2		0.47	0.74	4.53±0.25	5.6	5.26±3.69	-8.8-10	0.41	0.2	12	[-0.84,-0.24]	0.88
	CONNPET	9.8	2.2±1.4	0.47	0.73	4.56±0.35	7.7	5.08±3.64	-9.5-9.4	0.71	0.19	11.6	[-0.70,-0.32]	0.96
Sensorimotor Striatum	MANSEG1	21.9				5.47±0.59	10.8	4.34±2.66	-5.8-8.1	0.91	0.18	9		
	MANSEG2	9.7	43.7±16.6	0.56	0.75	4.82±0.39	8.1	4.84±2.97	-5.8-10	0.78	0.18	10.5	[-0.94,-0.36]	0.91
	PETSEG	18.5	16.1±13.3	0.53	0.68	5.20±0.51	9.8	7.01±4.50	-13.6-11.6	0.7	0.28	15	[-0.46,-0.08]	0.95
	CONNMRIF	8		0.33	0.19	4.23±0.25	6	5.16±2.47	-8.3-7.1	0.58	0.17	10.9	[-1.86,-0.62]	0.1
	CONNPET	8.3	2.2±1.0	0.36	0.18	4.38±0.34	7.8	4.90±2.52	-8.2-6.5	0.78	0.16	10.1	[-1.36,-0.82]	0.98

* MANSEG1/2=manual segmentation, PETSEG=automatic PET image clustering, FSSEG=Freesurfer segmentation, STRUCTMRIF=structural atlas segmentation in FSL, CONNMRIF/MRIS/PET=connectivity atlas segmentation based on FSL-MRI, SPM-MRI or SPM-PET image normalization. BP_{ND} (mean±s.d. from both scans)=SRTM-based binding potential using PETS reference region TAC. CoV(%)=coefficient of variation. TRV(%)=test-retest variability. ICC=intra-class correlation coefficient. SEM=standard error of measurement. MD(%)=minimal detectable change. LOA=limits of agreement (relative to MANSEG1). R=Pearson's correlation coefficient (relative to MANSEG1).


A Preprocessing methods
PET
1
Coreg.
+
Sum
Reslice
2
Coreg.
3
Coreg.+reslice
T1-MRI-template
T1-MRI
B Atlas individualization
1
Find
deformation
<->
2
Apply
inverse
deformation
->
C Manual segmentation
Front
Back
Head
Front
Left
Sum-PET
Sum-PET+ROIs
D Automated segmentation
Parametric
BP_{ND} image
(smoothed)
Striatum
extraction
Cluster
initialization
(geometric
prior)
Final clusters


FSSEG
STRUCTMRIF
CONNPET
CONNMRIS
CONNMRIF
PETSEG
MANSEG1