frontiers in Neuroinformatics

ORIGINAL RESEARCH
published: xx January 2020
doi: 10.3389/fninf.2020.00003

Magia: Robust Automated Image Processing and Kinetic Modeling Toolbox for PET Neuroinformatics

Tomi Karjalainen[1,2], Jouni Tuisku[2], Severi Santavirta[2], Tatu Kantonen[2], Marco Bucci[2], Lauri Tuominen[3], Jussi Hirvonen[2,4], Jarmo Hietala[2,5], Juha O. Rinne[2,6] and Lauri Nummenmaa[2,7]*

[1]Turku PET Centre, Turku University Hospital, Turku, Finland, [2]Turku PET Centre, University of Turku, Turku, Finland, [3]The Royal's Institute of Mental Health Research, University of Ottawa, Ottawa, ON, Canada, [4]Department of Radiology, University of Turku, Turku, Finland, [5]Department of Psychiatry, Faculty of Medicine, University of Turku and Turku University Hospital, Turku, Finland, [6]Division of Clinical Neurosciences, Turku University Hospital, Turku, Finland, [7]Department of Psychology, University of Turku, Turku, Finland

Processing of positron emission tomography (PET) data typically involves manual work, causing inter-operator variance. Here, we introduce Magia that enables processing of brain PET data with minimal user intervention. We investigated the accuracy of Magia with four tracers: [^{11}C]carfentanil, [^{11}C]raclopride, [^{11}C]MADAM, and [^{11}C]PiB. We used data from 30 control subjects for each tracer. Five persons manually delineated the reference regions for each subject. The data were processed using Magia using the manually and automatically generated reference regions. We first assessed inter-operator variance resulting from the manual delineation of reference regions. We then compared the differences between the manually and automatically produced reference regions and the subsequently obtained metrics. The results show that manually produced reference regions can be remarkably different from each other, leading to substantial differences also in outcome measures. While the Magia-derived reference regions were anatomically different from the manual ones, Magia produced outcome measures highly consistent with the average of the manually obtained estimates. For [^{11}C]carfentanil and [^{11}C]PiB there was no bias, while for [^{11}C]raclopride and [^{11}C]MADAM Magia produced 3–5% higher estimates. Based on these results and considering the high inter-operator variance of the manual method, we conclude that Magia can be reliably used to process brain PET data.

Keywords: PET, neuroinformatics, modeling, reference region

OPEN ACCESS

Edited by:
Ludovico Minati,
Tokyo Institute of Technology, Japan

Reviewed by:
Harumasa Takano,
National Center of Neurology and Psychiatry (Japan), Japan
Frithjof Kruggel,
University of California, Irvine, United States

****Correspondence:***
Tomi Karjalainen
tomi.karjalainen@utu.fi

Received: *18 November 2019*
Accepted: *15 January 2020*
Published: *xx January 2020*

Citation:
Karjalainen T, Tuisku J, Santavirta S, Kantonen T, Bucci M, Tuominen L, Hirvonen J, Hietala J, Rinne JO and Nummenmaa L (2020) Magia: Robust Automated Image Processing and Kinetic Modeling Toolbox for PET Neuroinformatics. Front. Neuroinform. 14:3. doi: 10.3389/fninf.2020.00003

INTRODUCTION

The statistical power of neuroimaging studies has been widely questioned in recent years, leading to calls for significantly larger samples to avoid false-positive and negative findings (Yarkoni, 2009; Button et al., 2013; Cremers et al., 2017). Additionally, the role of researcher degrees of freedom, i.e., the subjective choices made during the process from data collection to its analysis, has been identified as an important reason for poor replicability of many findings (Simmons et al., 2011). Consequently, the focus in neuroimaging has shifted towards standardized,

large-scale neuroinformatics based approaches (Yarkoni et al., 2011; Poldrack and Yarkoni, 2016). Today, several standardized and highly automatized preprocessing pipelines are publicly available for processing functional magnetic resonance images (fMRI; Esteban et al., 2019). Such standardized methods are not, however, currently widely used for the analysis of positron emission tomography (PET) data, although recently some tools have become available (Gunn et al., 2016; Funck et al., 2018).

Compared to fMRI preprocessing, preprocessing of PET data is relatively straightforward because confounding temporal signals are rarely regressed out of the data, and the preprocessing thus only consists of spatial processes, such as frame-realignment and coregistration. Yet, any all-inclusive PET processing pipeline must be able to handle numerous kinetic models to support as many radiotracers as possible. Thus, unlike fMRI preprocessing tools, PET pipelines should handle both the preprocessing as well as the kinetic modeling for numerous tracers, making the development of a comprehensive PET pipeline a challenging task.

A particularly sensitive task in PET analysis is the requirement of the input function. Depending on tracer, the input function can be obtained either from blood samples or directly from the PET images, for example, if a reference region is available for the tracer. The blood samples require manual processing before the input function can be obtained from them. While population-based atlases (Fischl et al., 2002; Tzourio-Mazoyer et al., 2002; Eickhoff et al., 2005) provide an automatic way for defining reference regions (Yasuno et al., 2002; Schain et al., 2014; Tuszynski et al., 2016), they are suboptimal because the process requires warping of either the atlases or the PET images. Ideally, the reference region should be defined separately for each individual before spatial normalization. Consequently, manual delineation is still considered the gold standard for defining the reference regions, thus prohibiting a fully automatic analysis of PET data. Furthermore, manual reference region delineation is time-consuming and relies on numerous subjective choices. To minimize between-study variance resulting from operator-dependent choices (White et al., 1999), a single individual should delineate the reference regions for all studies within a project. Thus, manual delineation is not suited for large-scale projects where hundreds of scans are processed, or neuroinformatics approaches where an even significantly larger number of scans have to be processed.

To resolve these issues, we introduce Magia[1] that enables automatic modeling of brain PET data with minimal user intervention The major advantages of this approach involve:

1. Flexible, parallelizable environment suitable for large-scale standardized analysis.
2. Fully automated processing of brain PET data starting from raw images.
3. Visual quality control of the processing steps.
4. Centralized management and storage of study metadata, image processing methods and outputs for subsequent reanalysis and quality control.

[1]https://github.com/tkkarjal/magia

In this study, we compared Magia-derived input functions and the subsequent outcome measures against those obtained using conventional manual techniques with four tracers binding to different sites: [^{11}C]carfentanil, [^{11}C]raclopride, [^{11}C]MADAM, and [^{11}C]PiB. We also assessed inter-rater agreement in the reference region definition and uptake estimates, and regional and voxel-level outcome measures.

MATERIALS AND METHODS

Overview of Magia

Magia[1] is a fully automatic analysis pipeline for brain PET data. Running on MATLAB (The MathWorks, Inc., Natick, MA, USA), Magia combines methods from SPM12[2] and FreeSurfer[3]–two freely available and widely used tools–with in-house software developed for kinetic modeling. Magia has been developed alongside a centralized database[4] containing metadata about each study, facilitating data storage and neuroinformatics-type large-scale PET analyses. While the implementation of a similar database is highly recommended, Magia can also be installed and used without such database as long as the user can feed in the necessary information about the studies. Magia runs only on Linux/Mac. The Optimization Toolbox for MATLAB is required for fitting some of the models. Magia has been developed using MATLAB R2016b. Magia currently supports the simplified reference tissue model, Logan (Logan, 2000) with both plasma input and reference tissue input, Patlak (Patlak et al., 1983) with both plasma input and reference tissue input, SUV-ratio (Chen and Nasrallah, 2017; standardized uptake value), and fractional uptake ratio (Thie, 1995; FUR) analysis for late scans with plasma input. Also, the two-tissue compartmental model can be fitted to regional-level data.

A box-diagram describing the main steps in Magia processing is shown in **Figure 1**. Magia starts by preprocessing the PET images. The preprocessing consists of frame-alignment (motion-correction) and coregistration with the MRI. The MRI is processed with FreeSurfer with recon-all to generate anatomical parcellations for defining regions of interest (Schain et al., 2014), and the reference region if one is required for the chosen kinetic model. FreeSurfer assigns an anatomical label to each brain voxel, and the region of interests (ROIs) thus consists of all the voxels with the same anatomical label. Magia performs a two-step correction to the reference tissue mask (see below) before obtaining the input function for modeling; the corrections make the reference regions robust for many scanners and individuals. The MRI is also segmented into gray and white matter probability maps for spatial normalization (Ashburner and Friston, 2000). After modeling, the parametric images are spatially normalized and smoothed. In addition to the parametric images. Magia also calculates region-level parameter estimates for each study. Finally, the results are stored in a centralized archive in a standardized format along with visual quality control metrics, facilitating future population-level analyses.

[2]www.fil.ion.ucl.ac.uk/spm/
[3]https://surfer.nmr.mgh.harvard.edu/
[4]http://aivo.utu.fi

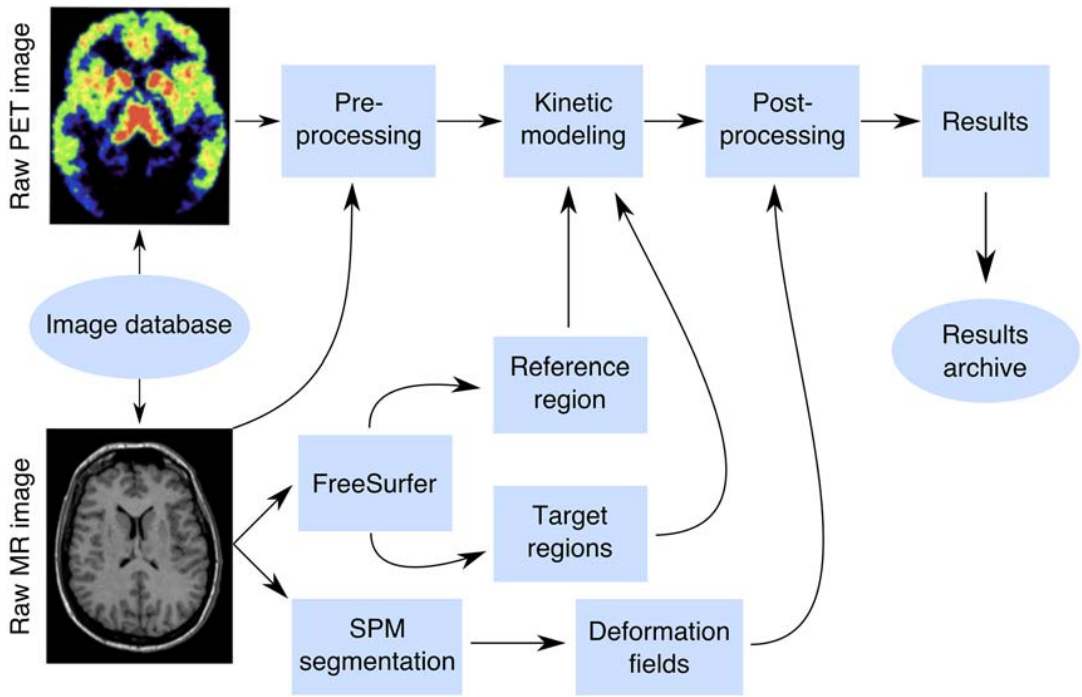

FIGURE 1 | The Magia pipeline combining FreeSurfer cortical mesh generation and parcellation, T1 MRI image segmentation and normalization, automatic reference region and region of interest (ROI) generation, and kinetic modeling.

The above-mentioned steps are only used when applicable. For example, for static images, the frame alignment is skipped, and if there is no related MRI available, then a tracer-specific radioactivity template must be available to normalize the images. For all of the tracers included in this manuscript, such templates can be obtained from https://github.com/tkkarjal/magia/tree/master/templates. Magia also supports tracers that do not have a reference region. For such studies, the preprocessed (e.g., decay-corrected, metabolite-corrected, and possibly extrapolated) plasma input must be available. Magia has default settings for preprocessing, modeling, and post-processing that have worked well during its development. However, Magia is also flexible in the sense that the user can override some of these options if needed.

Validation Data

To assess reliability of Magia we used historical control data using four radioligands with different targets and spatial distribution of binding sites: Dopamine D_2R receptor antagonist [^{11}C]raclopride, μ-opioid receptor agonist [^{11}C]carfentanil, serotonin transporter ligand [^{11}C]MADAM, and beta-amyloid ligand [^{11}C]PIB. For each radioligand we selected 30 studies (**Table 1**). We generated reference regions for all the tracers using traditional manual methods and the new automatic method and compared the results. The study was conducted as a part of a register-based study on brain imaging at Turku PET Centre. Per applicable legislation in Finland, fully anonymized medical register data (including PET and MRI scans) can be analyzed in the context of a register study without obtaining an active informed consent from the individuals included in the register, if information identifying the individuals is not obtained. The study protocol was approved by Turku University Hospital Research Board and the legislative team.

Manual Reference Region Delineation

Five researchers with good knowledge of human neuroanatomy delineated reference regions for every study according to written and visual instructions (**Figure 2A**). Cerebellar cortex was used as a reference region for [^{11}C]raclopride (Gunn et al., 1997), [^{11}C]MADAM (Lundberg et al., 2005) and [^{11}C]PiB (Lopresti et al., 2005). For [^{11}C]carfentanil, the occipital cortex was used (Endres et al., 2003). The regions were drawn using CARIMAS[5] on three consecutive transaxial slices of T1-weighted MR images, which is the current standard manual method at Turku PET Centre. Cerebellar reference was drawn in the cerebellar gray matter within a gray zone in the peripheral part of cerebellum, distal to the bright signal of white matter. The first cranial slice was placed below the occipital cortex to avoid spill-in of radioactivity. Typically, this is a slice where the temporal lobe is clearly separated from the cerebellum by the petrosal part of the temporal bone. The most caudal slice was typically located in the most caudal part of the cerebellum. Laterally, venous sinuses were avoided to avoid spill-in during the early phases of the scans. Posteriorly, there was about a 5 mm distance from the cerebellar surface to avoid spill-out effects. Anteriorly, the border of the

[5]http://turkupetcentre.fi/carimas/

Q5 **TABLE 1** | Summary of the studies.

	[^{11}C]carfentanil	[^{11}C]raclopride	[^{11}C]MADAM	[^{11}C]PiB
N (female)	30 (12)	30 (23)	30 (17)	30 (18)
Age (mean, range)	32 (20–51)	39 (20–60)	42 (25–57)	71 (66–80)
Scanners	HRRT PET/CT PET/MR	GE Advance PET/CT HRRT	HRRT	HRRT
Data range (years)	2007–2016	1998–2014	2008–2015	2014–2016

Scanners: HRRT (HRRT, Siemens Medical Solutions); PET/CT (Discovery 690 PET/CT, GE Healthcare); PET/MR (Ingenuity TF PET/MR, Philips Healthcare); GE Advance (GE Advance, GE Healthcare).

reference region was drawn approximately 2 mm distal to the border of cerebellar white and gray matter, except in the most caudal slice, where the central white matter may no longer be visible.

The occipital reference region was defined on three consecutive transaxial slices, of which the most caudal slice was the second-most caudal slice before the cerebellum. The reference region was drawn J-shaped with medial and posterior parts. The reference region was drawn to roughly follow the shape of the cortical surface, but not individual gyri. The reference region was drawn approximately 1 cm wide with about 2 mm margin to the cortical surface to avoid spill-out effects. The anterior border of the reference region was placed approximately halfway between the posterior cortical surface and the splenium of the corpus callosum. The posterolateral border of the reference region approximated the medial-most part of the posterior horn of the lateral ventricle.

Automatic Reference Region Generation

Figure 2B shows an overview of the automated reference-region-generation process. First, T1-weighted MR images were fed into FreeSurfer to provide subject-specific anatomical masks for cerebellar and occipital cortices. Second, an anatomical correction was applied to the FreeSurfer-generated reference region mask to remove voxels that, based on their anatomical location alone, are likely to suffer from spill-over effects. For the cerebellar cortex, the most important sources of spill-over effects are occipital cortex and venous sinuses. Thus, the most outermost cerebellar voxels were excluded in the anatomical reference region correction. For the occipital cortex, voxels that were lateral to the lateral ventricles were excluded. This is because the most lateral parts of the FreeSurfer-generated occipital cortex extend to areas with specific binding for [^{11}C]carfentanil, and the lateral ventricles provide a reliable anatomical cut-off point for thresholding. Finally, the radioactivity concentration distribution within the anatomically corrected reference region was estimated, and the tails of the distribution were excluded. The lower and upper boundaries for the signal intensities were defined by calculating the full width at half maximum (FWHM) of the mean PET signal intensity distribution. This step ensures that the reference region will not contain voxels with atypically high or low radioactivity (e.g., signal from outside the brain). The automatic reference region generation process thus combines information from anatomical brain scans and the PET images to get a reliable estimate of nonspecific binding.

Quantifying Operator-Dependent Variability

We first investigated how subjective choices in manual reference-region delineation translate into differences in reference region masks, reference-region time-activity curves (TACs), and outcome measures. Anatomical differences in reference region masks were assessed in two ways: first, we calculated within-study spatial overlap between the manual reference regions. The spatial overlap was calculated in two stages: it was first calculated separately for all different manual reference region pairs, and those numbers were then averaged over to obtain a summary statistic for each study. Second, we investigated the differences in volumes of the manually delineated reference regions using the intra-class correlation coefficient (ICC). To estimate ICC, we first estimated a random-effects model $y \sim 1 + (1 \mid \text{operator}) + (1 \mid \text{study})$, where, y is the variable of interest, and then calculated the proportion of variance explained by the variance of the random-effect-components (Nakagawa et al., 2017). Calculated this way, ICC is restricted to between 0 and 1. The R package brms[6] was used to estimate the models, and the R package performance[7] was used to estimate ICC.

Differences in reference region TACs were assessed by calculating area under the curve (AUC) of them. Prior to the ICC analysis, we standardized all the AUCs with the mean radioactivity within the union of all manually delineated reference regions. This standardization removes uninteresting between-study variance resulting from different scanners, body masses and injected doses. The operator-caused variation in outcome measures was also assessed using ICC.

The Volumetric Similarity of the Manual and Automatic Reference Regions

We compared the volumes of reference regions to assess whether the two techniques generate reference regions of systematically different sizes. For each study, we calculated the mean volume from all manually delineated reference regions and compared it to the volume of the Magia-derived reference region. We also quantified the anatomical overlap between the manually and the automatically derived reference regions. The overlap was defined as the ratio between the number of common voxels and the number of manual voxels. For each study, the overlap was first calculated separately for every manually delineated reference region after which the mean overlap was calculated.

[6]https://cran.r-project.org/package=brms
[7]https://easystats.github.io/performance/index.html

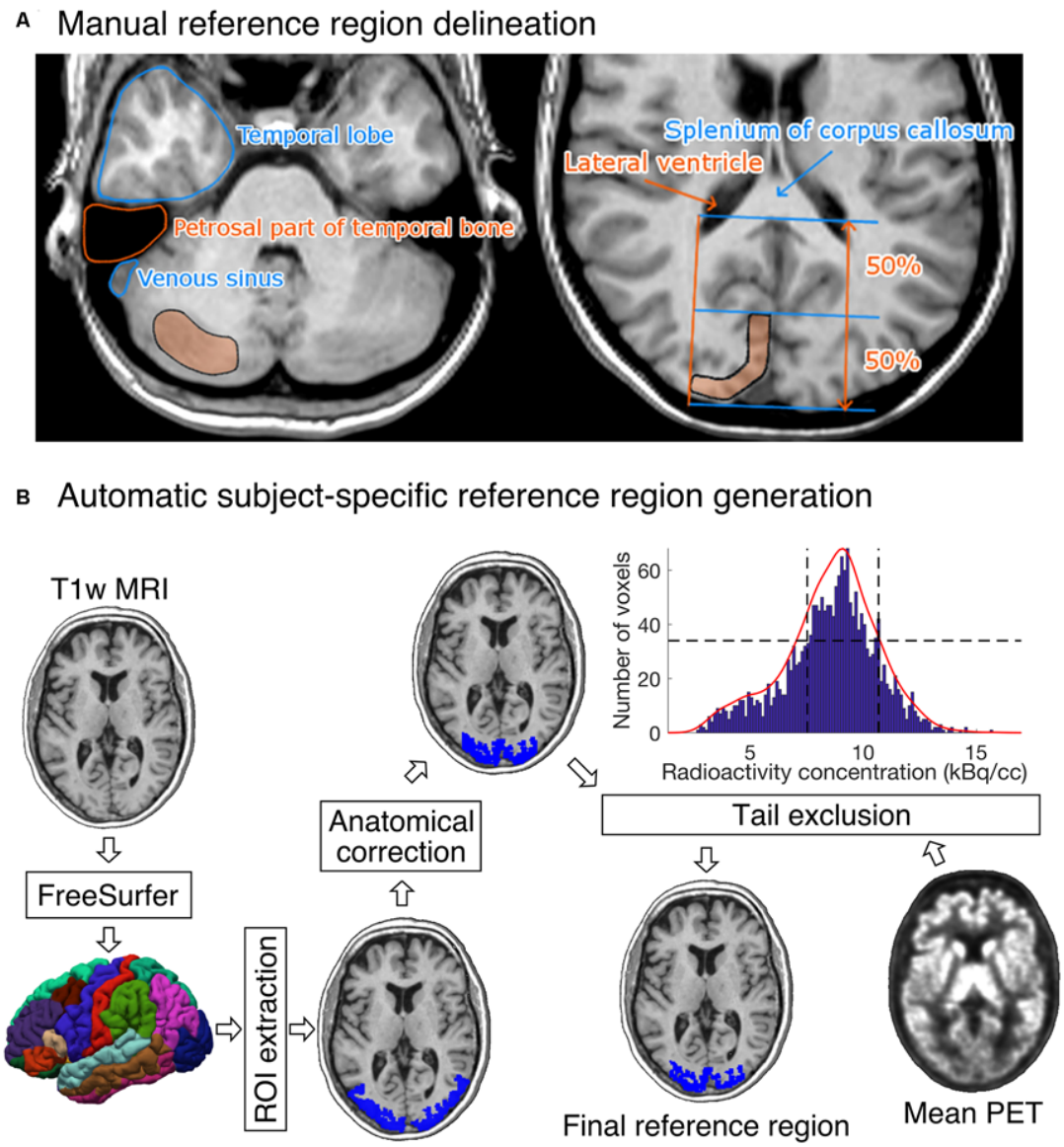

FIGURE 2 | (A) Visual instructions of the most cranial slice of manually delineated cerebellar (left) and occipital (right) reference regions. The reference regions were delineated on three consecutive transaxial T1-weighted MR images. Cerebellar reference region is shown on the left and occipital reference region on the right. **(B)** The diagram shows how a T1-weighted MR image of an individual's brain is processed to produce the final reference region. The shown example is from the [^{11}C]carfentanil data set. The rectangles represent processing steps between inputs and outputs. The FreeSurfer step assigns an anatomical label to each voxel of the subject's T1 weighted MR image. The ROI extraction step extracts a prespecified ROI from FreeSurfer's output. The anatomical correction removes voxels that are most likely to suffer from spillover effects; for [^{11}C]carfentanil data this means voxels lateral to the lateral ventricles. In the tail-exclusion step, radioactivity distribution within the anatomically corrected reference region is estimated, and the voxels whose intensities are on the tail-ends of the distribution are excluded.

Similarity of the Reference Region Radioactivity Concentrations

A functionally homogenous region should have approximately Gaussian distribution of radioactivity measured with PET (Teymurazyan et al., 2013). Functional homogeneousness was assessed using radioactivity distributions within the reference regions. The automatically and manually derived reference region masks were used to extract radioactivity concentration distribution within the reference regions. The study-specific manual distributions were averaged over the manual drawers to provide a single manual distribution for each study. The radioactivity concentrations were converted into SUV, after which the distributions were averaged over studies to provide tracer-specific distributions. Mean, standard deviations, mode, and skewness of the distributions were used to quantify the differences in the distributions.

Similarity of the Reference Region Time-Activity Curves

We compared the similarity of the automatically and manually delineated reference region TACs. For each study, the manual reference region TAC was defined as the average across the manual TACs to minimize the subjective bias in adhering to the instructions for manual reference region delineation. Activities were expressed as standardized uptake values (SUV, g/ml) which were obtained by normalizing tissue radioactivity concentration (kBq/ml) by total injected dose (MBq) and body mass (kg), thus making the different images more comparable to each other. To assess the similarity of the shapes of reference region TACs, we calculated Pearson correlations between the manually and automatically delineated TACs for each tracer. Bias was assessed using the area under the curve (AUC).

Assessing the Similarity of the Outcome Measures

We used nondisplaceable binding potential (BP_{ND}) to quantify uptakes of [^{11}C]carfentanil, [^{11}C]raclopride and [^{11}C]MADAM. It reflects the ratio between specific and nondisplaceable binding in the brain. The binding potentials were calculated using a simplified reference tissue model whose use has been validated for these tracers (Gunn et al., 1997; Endres et al., 2003; Lundberg et al., 2005). SUV-ratio between 60 and 90 min was used to quantify [^{11}C]PiB uptake (Lopresti et al., 2005). All the studies were first processed using Magia. To obtain the outcome measures resulting from manually delineated reference regions the procedure was repeated with the only exception of replacing the automatically generated reference regions with a manually generated reference region. Thus, the only differences observed in the uptake estimates originate from differences in the reference regions. We estimated the outcome measures in one representative ROI for each tracer, and also calculated parametric images. The ROIs were extracted from the FreeSurfer parcellations.

RESULTS

Operator-Dependent Variation

The influence of different operators on reference region volumes, reference region time-activity AUCs, and outcome measures are presented for each tracer in **Table 2**. The spatial overlap between the manually delineated masks was modest, as the maximum overlap was 41% for [^{11}C]raclopride studies, while the overlap for the other tracers was 14–22%. The ICC for reference region volumes were moderate to good (0.74. . .0.83) for all tracers except [^{11}C]MADAM (ICC = 0.46). The reference region TAC AUCs varied substantially especially for [^{11}C]carfentanil and [^{11}C]MADAM, while for [^{11}C]PiB operator had little influence on the AUCs (ICC = 0.95). The operator had the most influence on outcome measures for [^{11}C]carfentanil and [^{11}C]MADAM. For [^{11}C]raclopride and [^{11}C]PiB operators had little influence on outcome measures (ICC $\geq$ 0.95).

Differences Between Manually and Automatically Produced Reference Regions

Differences in Reference Region Masks

We first compared the anatomical similarities between the automatically and manually delineated reference regions. For each tracer, automatic reference regions were consistently larger than manually derived reference regions (**Figure 3** and **Supplementary Figure S1**). In four [^{11}C]carfentanil studies at least one of the manually drawn reference region was larger than the automatic occipital reference region. Magia-generated cerebellar reference regions were always larger than mean manual cerebellar reference regions. The automatically produced reference regions are naturally larger than the manually delineated ones because manual delineation requires mechanic work from highly trained individuals, thus providing a cost to the size of the regions.

Next, we determined whether the Magia-derived reference regions overlap with the manually drawn reference regions. The automatic occipital reference region for [^{11}C]carfentanil overlapped only 14% with a manual occipital reference region. The low overlap is explained by the substantial difference between the sizes of the manually and automatically generated occipital ROIs. Automatic cerebellar reference regions overlapped with manual reference regions by 55%, 59% and 61% for [^{11}C]raclopride, [^{11}C]MADAM and [^{11}C]PiB, respectively.

Differences in Reference Region SUV Distributions

The overlap between the manual and automatic radioactivity distributions was approximately 90% for all tracers (**Supplementary Figure S2**). All distributions were unimodal and highly symmetric for all tracers. The means of the distributions were practically equal (maximum difference of 0.07%). The standard deviations of the distributions differed

Q5

TABLE 2 | Operator-caused variation in basic characteristics derived from the reference region masks.

	Intra-class correlation coefficient			
Tracer	**Spatial overlap (%)**	**Reference region volume**	**Reference TAC AUC**	**Outcome measure**
[^{11}C]carfentanil	22	83	61	75
[^{11}C]raclopride	41	79	80	97
[^{11}C]MADAM	18	46	58	76
[^{11}C]PiB	14	74	95	95

TAC, time-activity curve; AUC, area under curve.

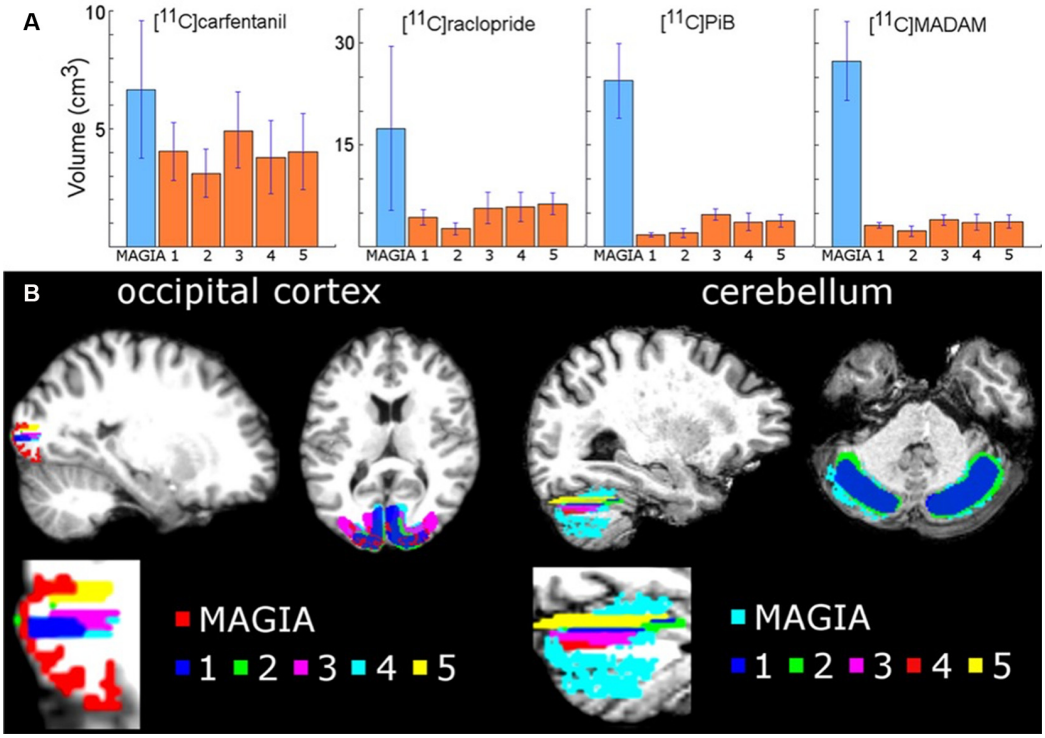

FIGURE 3 | (A) Mean volumes of Magia-generated reference regions compared to mean volumes of manually delineated reference regions. **(B)** Visual examples of Magia-generated and manual reference regions for one study.

by 14%, 11%, 12% and 18% for [^{11}C]carfentanil, [^{11}C]MADAM, [^{11}C]PIB and [^{11}C]raclopride, respectively. The modes of the automatically and manually derived distributions were 1.5 and 1.55 for [^{11}C]carfentanil, 1.95 and 2.05 for [^{11}C]MADAM, 1.65 and 1.70 for [^{11}C]PIB, and 1.35 and 1.35 for [^{11}C]raclopride. Thus, the maximum difference was less than 5%. The skewnesses of the Magia-derived and manually derived distributions were 1.2 and 0.9 for [^{11}C]carfentanil, 1.3 and 1.2 for [^{11}C]MADAM, 2.0 and 1.6 for [^{11}C]PIB, and 2.4 and 2.0 for [^{11}C]raclopride.

Differences in Reference Region Time-Activity Curves

The Magia-produced TACs were on average very similar to the average TACs calculated based on the manually delineated reference regions (**Figure 4**). The Pearson correlation coefficients were above 0.99 for all tracers. **Supplementary Figure S3** shows how the Magia-derived reference region time-activity curve AUCs compare against the manually obtained results. For [^{11}C]carfentanil, the between-study AUC means were practically identical (<1%). The Magia-produced reference regions had 2.6%, 1.1%, and 1.8% lower AUCs than the manual reference regions for [^{11}C]raclopride, [^{11}C]MADAM, and [^{11}C]PiB, respectively.

Differences in Outcome Measures

Pearson correlation coefficients between the mean of manual outcome measures and the Magia-derived outcome measures were 0.79, 0.98, 0.84, and 0.99 for [^{11}C]carfentanil, [^{11}C]raclopride, [^{11}C]MADAM, and [^{11}C]PiB, respectively. The outcome measures derived using automatic and manual methods are visualized in **Figure 5** in one representative ROI, and the relative bias in the whole brain between them is visualized in **Figure 6B**. For [^{11}C]carfentanil and [^{11}C]PiB Magia produced basically no bias (less than 1%). For [^{11}C]MADAM, Magia produced up to 3–5% higher binding potential estimates in regions with high specific binding. In cortical regions with low specific binding, the bias was over 10%. For [^{11}C]raclopride, Magia produced approximately 4–5% higher binding potential estimates in striatum. In the thalamus, the bias was 8–10%. Elsewhere in the brain the bias varied considerably between 13–20%. For both [^{11}C]MADAM and [^{11}C]raclopride, the relative bias decreased significantly with increasing binding potential (**Figure 6C**).

DISCUSSION

We established that the automated Magia pipeline produces consistent estimates of radiotracer uptake for all the tested ligands, with very little or even no bias in the outcome measures. As expected, the manual delineation method suffered from significant operator-dependent variability, highlighting the importance of standardization of the process. The consistency coupled with significant gains in processing speed suggests that Magia is well suited for automated analysis of brain-PET data for large-scale neuroimaging projects.

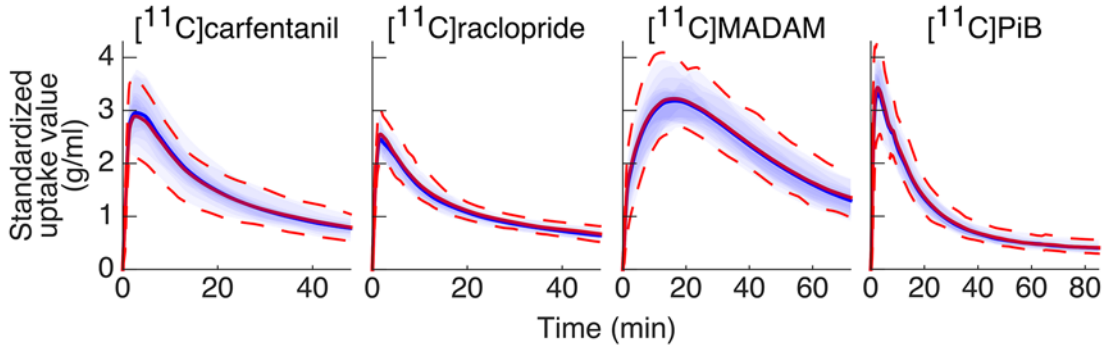

FIGURE 4 | Between-subject mean time-activity curves (TACs). Blue = Magia; red = manual.

Outcome Measures Can Substantially Depend on Who Delineated the Reference Region

We estimated the amount of operator-dependent variation in outcome measures. Despite all operators drawing the ROIs using the same instructions (presented both verbally and as visual/written instructions available for reference while working) the ICC analyses show that for [^{11}C]carfentanil and [^{11}C]MADAM, the variation produced by different operators is significant, indicating that for these two tracers the subjective variation in manual ROI delineation (e.g., which transaxial slices to use, how to define ROI boundaries etc.) significantly influences the magnitude of binding potential estimates. Out of the tracers using the cerebellar cortex as the reference region, [^{11}C]MADAM had the lowest ICC with 76%. For [^{11}C]raclopride and [^{11}C]PiB the ICCs were over 95%, indicating that for these tracers manual delineation of reference regions may not be as crucial source of variation.

These differences between tracers likely reflect differences in the uniformity of the PET signal within the reference region. If the reference region were perfectly homogenous with respect to the PET signal, it would not matter at all which voxels to choose. In reality, however, the PET signal is highly heterogenous. For example, the PET signal depends on the transaxial slices used. Presumably, these heterogeneities are substantial for [^{11}C]carfentanil and, to a lesser extent, for [^{11}C]MADAM, while the PET signal from cerebellar cortex using [^{11}C]raclopride and [^{11}C]PiB is significantly more homogenous. Indeed, the spatial overlap between the manually delineated reference region was higher for [^{11}C]carfentanil (22%) than for [^{11}C]PiB (14%), suggesting that even small differences in spatial overlap translate into substantial differences in binding potential for [^{11}C]carfentanil.

The influence of the operator on reference TAC AUCs was even larger. For all the tracers, the ICC of outcome measures was higher than the ICC for reference TAC AUCs. For example, while [^{11}C]raclopride BP_{ND} was barely influenced by the individual manually delineating the reference region, the ICC for [^{11}C]raclopride reference TAC AUC was only 80%, almost 20%-units less than for BP_{ND}. Thus, even the reference region TACs for [^{11}C]raclopride was not remarkably consistent between the operators, further highlighting the sensitivity of the delineation process despite detailed written and visual instructions. These results highlight the need for reference-region generation processes that do not suffer from subjectivity.

Reliability of Magia's Uptake Estimates

Importantly, Magia produced parameter estimates consistent with the *averaged* manual estimates (Pearson correlation coefficients >0.78 for all tracers). This suggests that: (i) even though individual operators yield different output metrics these are sampled from the same true parameter space; which (ii) is in turn accurately reflected by the Magia output. There was no systematic bias for [^{11}C]PiB SUVR and [^{11}C]carfentanil BP_{ND}. For [^{11}C]PiB, the difference between the manual and automatic SUVR estimates fluctuated randomly around zero. Because SUVR was used to quantify [^{11}C]PiB uptake, the random fluctuation was independent of the brain region. For [^{11}C]carfentanil, the random fluctuation was slightly greater in low-binding regions (but still within ±5%). In contrast to [^{11}C]PiB and [^{11}C]carfentanil, there were systematic differences between the manual and automatic binding potential estimates for [^{11}C]raclopride and [^{11}C]MADAM. For both tracers the bias decreased as a function of specific binding, and in high-binding regions ($BP_{ND} > 1.5$) the bias was less than 5%. Even if the bias increased sharply with decreasing binding potential, the problematic regions are not typically considered very interesting because of their poor signal-to-noise ratio.

The systematic bias for [^{11}C]MADAM and [^{11}C]raclopride is also reflected in the small differences in reference to tissue TACs. For the tracers using cerebellar reference region, Magia-derived reference tissue TACs had 2–3% lower AUCs. The peaks of the TACs were also slightly lower. For [^{11}C]PiB, the bias did not propagate into outcome measures because the SUV-ratio was calculated between 60 and 90 min when there was no bias in TACs. Because binding potential reflects the ratio between specific binding and unspecific binding (obtained from reference tissue), the reference TAC AUCs directly propagate into biases in binding potentials. Thus, these data indicate that Magia may produce slightly higher binding potential estimates than

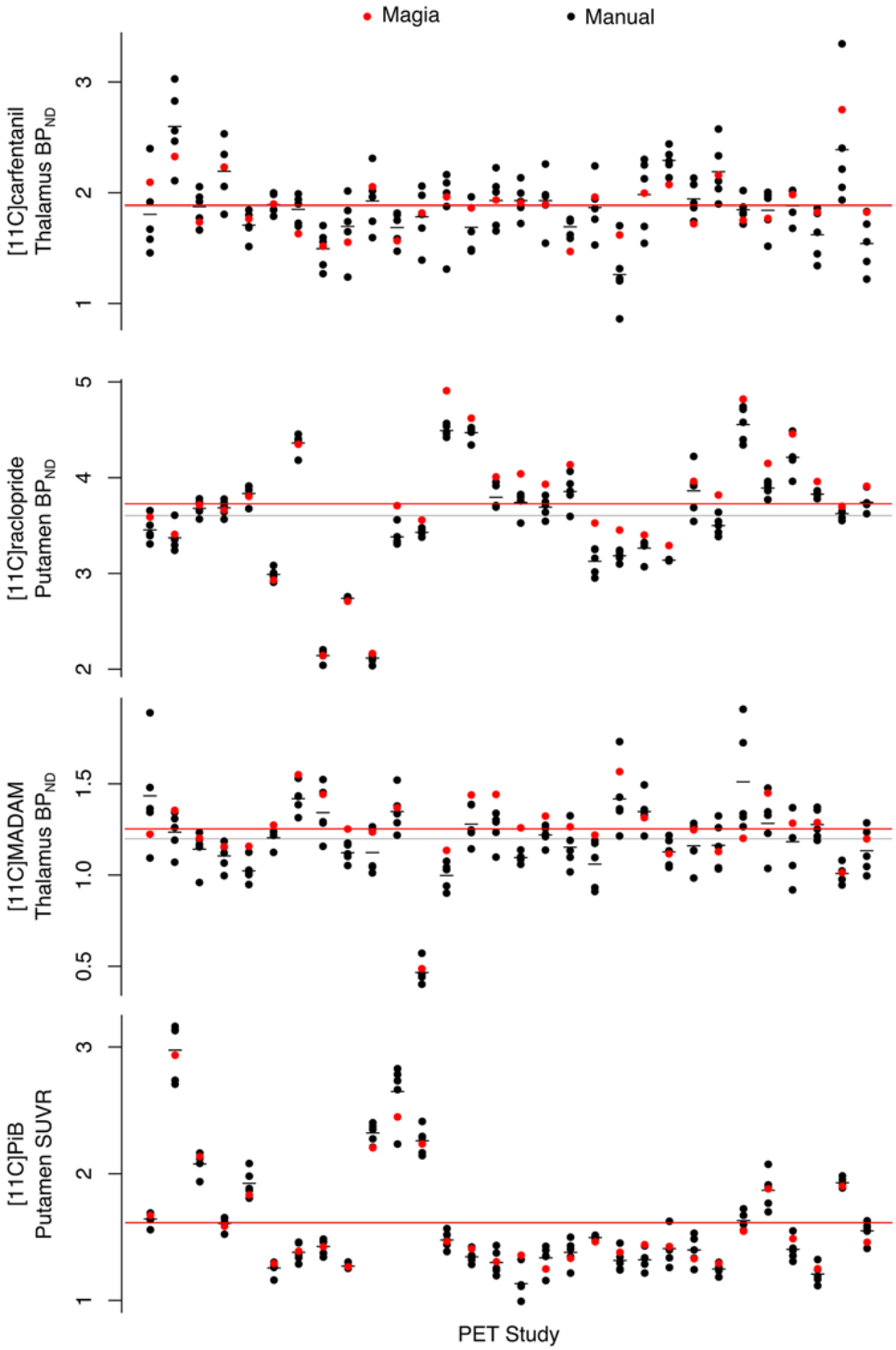

FIGURE 5 | Comparison of Magia-derived outcome measures against manually obtained ones.

traditional methods at least if the cerebellar cortex is used as the reference region. These data do not, however, imply that the bias should be regarded as error: in fact, Magia produces significantly larger reference regions, and consequently the reference tissue TACs are less noisy. This is desirable because the noise in the input function influences model fitting. However, the bias also means that Magia-produced estimates should not be combined with estimates produced with other methods.

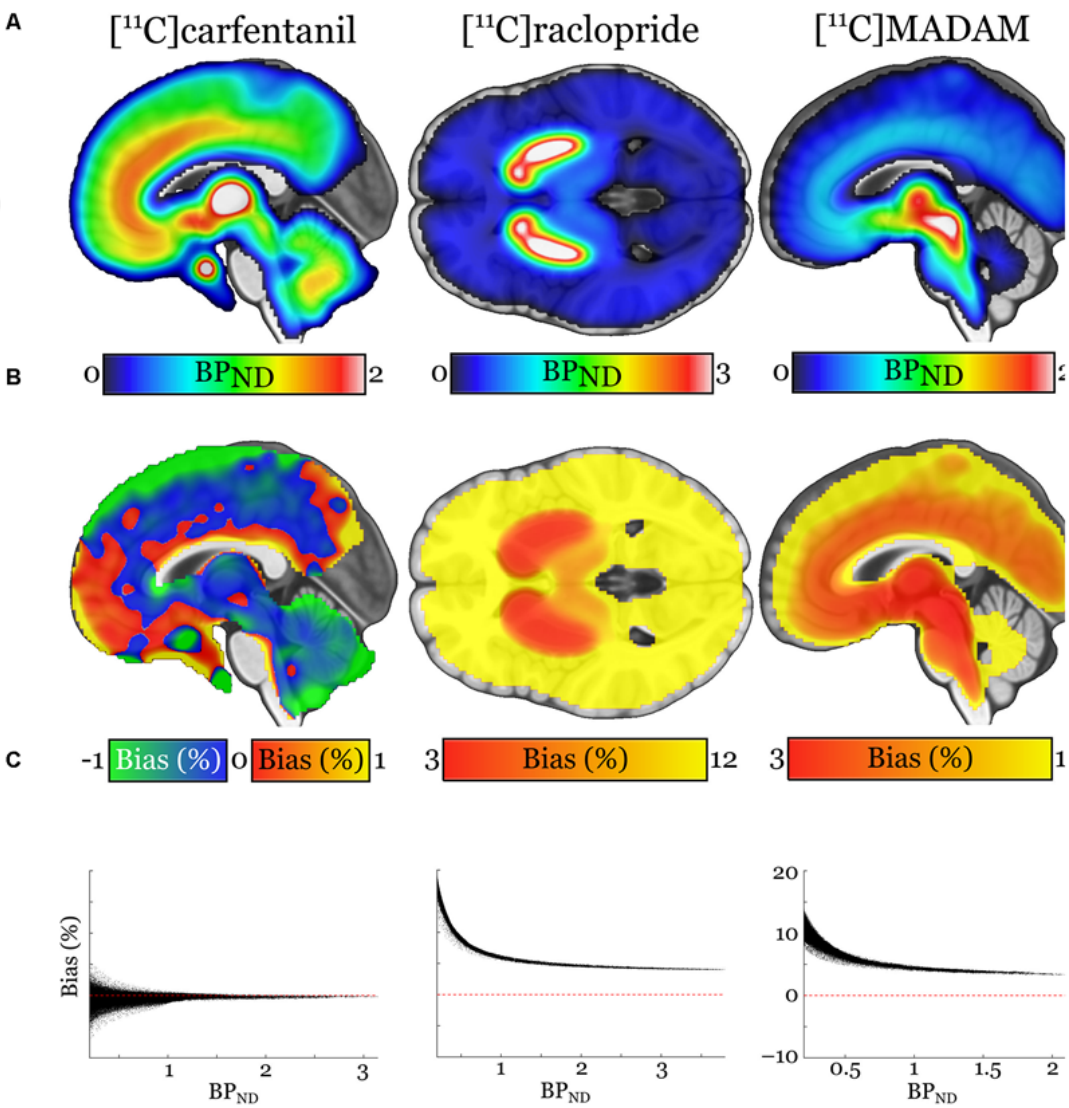

FIGURE 6 | (A) Visualization of the outcome measure distributions for each tracer. **(B)** Maps visualizing the relative biases of the Magia-derived outcome measures compared to the averages obtained by manual reference region delineation. The manual method is here presented as the ground truth, because the manual outcome for each scan is an average over five individual estimates, while the Magia result relies on a single estimate. **(C)** Associations between the outcome measure magnitude and relative bias.

Functional Homogeneity of the Reference Regions

We tested whether the assumption of homogenous binding within the reference regions holds for both automatic and manual reference regions. A homogenous source region should produce unimodal and approximately symmetric radioactivity distributions 21. Between-study average distributions were unimodal and symmetric for all tracers for both the manual and automatic methods. The distribution means were practically identical, but the modes were 1–2% higher for Magia. The manual distributions were slightly wider (the standard deviations were approximately 15% larger) because Magia cuts the distribution tails. The manual distributions were also slightly less skewed. Because averaging distributions tends to make them more Gaussian, this difference probably arises from the fact that the manual distributions that were used in the comparison were defined as an average over the five distributions delineated by the independent operators. The distribution overlaps were approximately 90% for all tracers. In sum, these results show that the Magia-generated reference region radioactivity distributions satisfy the requirement of functional uniformity.

Reference Tissue Time-Activity Curves

Despite their topographical differences, the automatically and manually produced reference regions yielded very similar TACs. For all tracers, the Pearson correlation coefficient between

average automatic and manual reference tissue TACs was above 0.99. The TAC shapes were thus in excellent agreement. For [^{11}C]carfentanil, also the AUC of reference region TACs were highly similar. The AUCs of cerebellar TACs were 2–3% lower for Magia, indicating that the cerebellar automatic TACs were slightly negatively biased compared to their manual counterparts. The source of this difference unknown but it could result e.g., from heterogenous nonspecific binding within cerebellar cortex or from spill-in or spill-over effects. Whatever explains the small difference, these data do not directly indicate which method produced more realistic TACs. However, because the Magia-generated cerebellar reference regions were without exception substantially larger than their manual counterparts, the TACs of Magia presumably have a higher signal-to-noise ratio, suggesting that the Magia-derived metrics may compare favorably against the manually obtained metrics.

Solving Time Constraints in the Processing of PET Data

On average, drawing the reference region for a single subject took around 15 min, and without any automatization the modeling and spatial processing of the images standard tools (e.g., PMOD or Turku PET Centre modeling software) take on average 45 min. In contrast, it takes less than 5 min to set Magia running for a single study. Although the time advantage—roughly an hour per study—gained from automatization is still modest in small-scale studies (e.g., three 8-h working days for a study with 24 subjects) the effect scales up quickly, and manual modeling of a database of just 400 studies would take already 50 days. This is a significant investment of human resources, in particular, if the analyses have to be redone later with, for example, different modeling parameters requiring repeating of at least some parts of the process.

Comparison of Magia to Existing Tools

Several tools already exist for processing brain PET data. MIAKAT (Gunn et al., 2016) is another MATLAB-based tool that combines preprocessing and kinetic modeling. Compared to Magia, MIAKAT is missing support for the two-tissue compartmental model, SUV-ratio, as well as FUR-analyses. APPIAN (Funck et al., 2018) is another recent development that, unlike Magia, includes partial volume correction. However, APPIAN lacks motion-correction and also supports fewer kinetic models than Magia, and like MIAKAT, APPIAN also uses neuroanatomical atlases for ROI definition. Both of these tools, as well as all the other existing tools, are restricted in the sense that they require both MRI and PET data. Magia, in contrast, can also process brain PET data without MRI if a tracer-specific template is available. Magia also comes with default modeling options for several tracers. Accordingly, Magia is currently the most flexible open-source tool available for automated processing of brain PET data.

Limitations

Magia is currently fully automatic only for tracers for which a reference region exists. However, even for blood-based inputs, Magia requires minimal user intervention, as Magia can read in the input function from the appropriate location. Magia was originally developed with the assumption that a T1-weighted MR image is available for each subject (for reference region delineation and spatial normalization). Because this assumption limited the applicability of the approach for reanalysis of some historical data, Magia can now also use neuroanatomical atlases for ROI definition and tracer-specific radioactivity templates for spatial normalization. Templates for each of the tracers used in this manuscript are available in https://github.com/tkkarjal/magia/tree/master/templates, and Magia can use whatever templates the user may have available. Thus, the availability of MRI is not necessary, but it is strongly recommended because most of the testing has been done with MRI-based processing, and because the ROIs as well as reference regions can then be generated in the native space. The drawback of FreeSurfer-based ROI-generation is that it is relatively slow (∼ 10 h). Partial volume correction is not currently implemented in Magia, yet this feature will be added in future releases. Finally, Magia processes the studies independently of each other. Within-subject designs would benefit from consideration of multiple images per participant, but this is currently not possible.

CONCLUSION

Magia is a standardized and fully automatic analysis pipeline for processing brain PET data. By standardizing the reference region generation process, Magia eliminates operator-dependency in producing outcome-measures. For [^{11}C]carfentanil that uses the occipital cortex as the reference region, the reduced variance comes with no cost for bias in BP_{ND}. The SUVR estimates were also unbiased for [^{11}C]PiB. [^{11}C]raclopride and [^{11}C]MADAM BP_{ND} was slightly overestimated. However, compared to the variance resulting from operator dependency, this bias was negligible and may actually favor Magia. In any case, bias is meaningless in most population-level analyses. Magia enables standardized analysis of brain PET data, facilitating shift towards larger samples and more convenient data sharing across research sites.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available. The current data-sharing guidelines prohibit publishing the data. Q7

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Q8

AUTHOR CONTRIBUTIONS

ToK developed Magia, analyzed the data, and wrote the manuscript. JT contributed to the development of Magia and

edited the manuscript. SS manually delineated reference regions, contributed to data analysis and edited the manuscript. TaK contributed to data analysis and edited the manuscript. MB contributed to late development of Magia and edited the manuscript. LT contributed to the early development of Magia and edited the manuscript. JuH planned statistical analyses and edited the manuscript. JaH and JR provided data and edited the manuscript. LN provided data, contributed to the development of Magia and edited the manuscript.

FUNDING

Q9 This work was supported by the Academy of Finland grants #265915 and #294897 to LN, and Sigrid Juselius Foundation separate grants to LN and JR, Päivikki and Sakari Sohlberg Foundation to ToK, Finnish Cultural Foundation Varsinais-Suomi Regional Fund to ToK, State Research Funding (VTR) for Turku University Hospital to ToK, State Research Funding (VTR) for Turku University Hospital to JR.

ACKNOWLEDGMENTS

This manuscript includes work in-part from a Thesis "Magia: Robust Automated Modeling And Image Processing Platform For Pet Neuroinformatics" by Santavirta (Santavirta, 2018). The manuscript has also been released as a preprint at biorXiv (Karjalainen et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2020.00003/full#supplementary-material. Q10

REFERENCES

Ashburner, J., and Friston, K. J. (2000). Voxel-based morphometry—the methods. *Neuroimage* 11, 805–821. doi: 10.1006/nimg.2000.0582

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475

Chen, Y. J., and Nasrallah, I. M. (2017). Brain amyloid PET interpretation approaches: from visual assessment in the clinic to quantitative pharmacokinetic modeling. *Clin. Transl. Imaging* 5, 561–573. doi: 10.1007/s40336-017-0257-4

Cremers, H. R., Wager, T. D., and Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLoS One* 12:e0184923. doi: 10.1371/journal.pone.0184923

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., and Amunts, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Endres, C. J., Bencherif, B., Hilton, J., Madar, I., and Frost, J. J. (2003). Quantification of brain mu-opioid receptors with [^{11}C]carfentanil: reference-tissue methods. *Nucl. Med. Biol.* 30, 177–186. doi: 10.1016/s0969-8051(02)00411-0

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* 16, 111–116. doi: 10.1038/s41592-018-0235-4

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355. doi: 10.1016/s0896-6273(02)00569-x

Q15 Funck, T., Larcher, K., Toussaint, P.-J., Evans, A. C., and Thiel, A. (2018). APPIAN: automated pipeline for PET image analysis. *Front. Neuroinform.* 12:64. doi: 10.3389/fninf.2018.00064

Q16 Gunn, R., Coello, C., and Searle, G. (2016). Molecular imaging and kinetic analysis toolbox (MIAKAT)—a quantitative software package for the analysis of PET neuroimaging data. *J. Nucl. Med.* 57, 1928–1928.

Gunn, R. N., Lammertsma, A. A., Hume, S. P., and Cunningham, V. J. (1997). Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. *NeuroImage* 6, 279–287. doi: 10.1006/nimg.1997.0303

Karjalainen, T., Santavirta, S., Kantonen, T., Tuisku, J., Tuominen, L., Hirvonen, J., et al. (2019). Magia: robust automated modeling and image processing toolbox for PET neuroinformatics. *bioRxiv* [Preprint]. doi: 10.1101/604835

Logan, J. (2000). Graphical analysis of PET data applied to reversible and irreversible tracers. *Nucl. Med. Biol.* 27, 661–670. doi: 10.1016/s0969-8051(00)00137-2

Lopresti, B. J., Klunk, W. E., Mathis, C. A., Hoge, J. A., Ziolko, S. K., Lu, X., et al. (2005). Simplified quantification of Pittsburgh Compound B amyloid imaging PET studies: a comparative analysis. *J. Nucl. Med.* 46, 1959–1972. Q16

Lundberg, J., Odano, I., Olsson, H., Halldin, C., and Farde, L. (2005). Quantification of ^{11}C-MADAM binding to the serotonin transporter in the human brain. *J. Nucl. Med.* 46, 1505–1515. Q16

Nakagawa, S., Johnson, P. C. D., and Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. R. Soc. Interface* 14:20170213. doi: 10.1098/rsif.2017.0213

Patlak, C. S., Blasberg, R. G., and Fenstermacher, J. D. (1983). Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. *J. Cereb. Blood Flow Metab.* 3, 1–7. doi: 10.1038/jcbfm.1983.1

Poldrack, R. A., and Yarkoni, T. (2016). From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annu. Rev. Psychol.* 67, 587–612. doi: 10.1146/annurev-psych-122414-033729

Santavirta, S. (2018). *Magia: Robust Automated Modeling and Image Processing Platform for PET Neuroinformatics.* Turku: University of Turku.

Schain, M., Varnäs, K., Cselényi, Z., Halldin, C., Farde, L., and Varrone, A. (2014). Evaluation of two automated methods for PET region of interest analysis. *Neuroinformatics* 12, 551–562. doi: 10.1007/s12021-014-9233-6

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366. doi: 10.1177/0956797611417632

Teymurazyan, A., Riauka, T., Jans, H. S., and Robinson, D. (2013). Properties of noise in positron emission tomography images reconstructed with filtered-backprojection and row-action maximum likelihood algorithm. *J. Digit. Imaging* 26, 447–456. doi: 10.1007/s10278-012-9511-5

Thie, J. A. (1995). Clarification of a fractional uptake concept. *J. Nucl. Med.* 36, 711–712.

Tuszynski, T., Rullmann, M., Luthardt, J., Butzke, D., Tiepolt, S., Gertz, H.-J., et al. (2016). Evaluation of software tools for automated identification of neuroanatomical structures in quantitative β-amyloid PET imaging to diagnose Alzheimer's disease. *Eur. J. Nucl. Med. Mol. Imaging* 43, 1077–1087. doi: 10.1007/s00259-015-3300-6

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., and Delcroix, N. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15, 273–289. doi: 10.1006/nimg.2001.0978

White, D. R. R., Houston, A. S., Sampson, W. F. D., and Wilkins, G. P. (1999). Intra- and interoperator variations in region-of-interest drawing and their effect on the measurement of glomerular filtration rates. *Clin. Nucl. Med.* 24, 177–181. doi: 10.1097/00003072-199903000-00008

Yarkoni, T. (2009). Big correlations in little studies: inflated fMRI correlations reflect low statistical power—commentary on Vul et al. (2009). *Perspectives on Psychological Science* 4, 294–298. doi: 10.1111/j.1745-6924.2009.01127.x

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670. doi: 10.1038/nmeth.1635

Yasuno, F., Hasnine, A. H., Suhara, T., Ichimiya, T., Sudo, Y., Inoue, M., et al. (2002). Template-based method for multiple volumes of interest of human brain PET images. *NeuroImage* 16, 577–586. doi: 10.1006/nimg.2002.1120

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