Cerebral Serotonin Transporter Measurements with 
across 21 PET Centres

DOI:
10.1177/0271678X18770107

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Journal of Cerebral Blood Flow and Metabolism

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
**Cerebral Serotonin Transporter Measurements with [11C]DASB: A Review on Acquisition and Preprocessing across 21 PET Centres**

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Journal of Cerebral Blood Flow and Metabolism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>JCBFM-0613-17-REV.R2</td>
</tr>
<tr>
<td>Manuscript Type</td>
<td>Review Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>07-Mar-2018</td>
</tr>
</tbody>
</table>

**Complete List of Authors:**

Nørgaard, Martin; Rigshospitalet, Copenhagen University Hospital, Neurobiology Research Unit; University of Copenhagen, Faculty of Health and Medical Sciences
Ganz, Melanie; Rigshospitalet, Copenhagen University Hospital, Neurobiology Research Unit; University of Copenhagen, Faculty of Computer Science
Svarer, Claus; Rigshospitalet, Neurobiology Research Unit
Feng, Ling; Rigshospitalet, Copenhagen University Hospital, Neurobiology Research Unit
Ichise, Masanori; Department of Functional Brain Imaging Research, National Institute of Radiological Sciences, National Institutes for Quantum and Radiological Science and Technology, Chiba, Japan
Lanzenberger, Rupert; Medical University of Vienna, Department of Psychiatry and Psychotherapy
Lubberink, Mark; Uppsala University Hospital, PET Centre
Parsey, Ramin; Stony Brook University, Department of Psychiatry and Behavioral Science
Politis, Marios; Neurodegeneration Imaging Group, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King’s College London, London, United Kingdom
Rabiner, EUGENII; Imanova Limited, Clinical Applications
Slifstein, Mark; Stony Brook University Health Sciences Center School of Medicine, Psychiatry
Sossi, Vesna; University of British Columbia, Physics and Astronomy
Suhara, Tetsuya; National Institutes for Quantum and Radiological Science and Technology
Talbot, Peter; Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
Turkheimer, Federico; Institute of Psychiatry, King's College London, Centre for Neuroimaging
Strother, Stephen; Baycrest, Rotman Research Institute; University of Toronto, Medical Biophysics
Knudsen, Gitte; Univ. Hospital Rigshospitalet, 6931 Neurobiology Research Unit
<table>
<thead>
<tr>
<th>Research Topics:</th>
<th>Brain Imaging, Pharmacokinetics, Receptor imaging, Positron Emission Tomography, Neuroanatomy, MRI, Cortical Mapping, 5-HT, serotonin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Techniques:</td>
<td>Positron Emission Tomography, Imaging, Kinetic modelling, Molecular imaging</td>
</tr>
</tbody>
</table>
Cerebral Serotonin Transporter Measurements with [$^{11}$C]DASB: A Review on Acquisition and Preprocessing across 21 PET Centres

Running Title: The Serotonin Transporter with $^{11}$C DASB across 21 PET Centres

Martin Nørgaard$^{1,2}$, Melanie Ganz$^{1,3}$, Claus Svarer$^1$, Ling Feng$^1$, Masanori Ichise$^6$, Rupert Lanzenberger$^7$, Mark Lubberink$^8$, Ramin V. Parsey$^9$, Marios Politis$^{10}$, Eugenii A. Rabiner$^{8,11}$, Mark Slifstein$^9$, Vesna Sossi$^{12}$, Tetsuya Suhara$^6$, Peter S. Talbot$^{13}$, Federico Turkheimer$^{14}$, Stephen C. Strother$^4$, Gitte M. Knudsen$^{1,2}$*

$^1$Neurobiology Research Unit, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
$^2$University of Copenhagen, Faculty of Health Sciences, Copenhagen, Denmark
$^3$University of Copenhagen, Faculty of Computer Science, Copenhagen, Denmark
$^4$Rotman Research Institute at Baycrest, University of Toronto, Toronto, Canada
$^5$Imanova Limited, London, United Kingdom
$^6$Department of Functional Brain Imaging Research, National Institute of Radiological Sciences, National Institutes for Quantum and Radiological Science and Technology, Chiba, Japan
$^7$Department of Psychiatry and Psychotherapy, Medical University of Vienna, Austria
$^8$Department of Nuclear Medicine and Positron Emission Tomography, Uppsala University, Uppsala, Sweden
$^9$Department of Psychiatry, Stony Brook University School of Medicine, Stony Brook, New York
$^{10}$Neurodegeneration Imaging Group, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King’s College London, London, United Kingdom
$^{11}$Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology and Neuroscience, King’s College London
$^{12}$Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada
$^{13}$Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
$^{14}$Department of Neuroimaging, King’s College London, London, UK

*Corresponding author:
Professor Gitte Moos Knudsen
Neurobiology Research Unit, Section 6931, Rigshospitalet
9 Blegdamsvej
DK-2100 Copenhagen
+45 3545 6720
gmk@nru.dk
Abstract—Positron Emission Tomography (PET) imaging has become a prominent tool to capture the spatiotemporal distribution of neurotransmitters and receptors in the brain. The outcome of a PET study can, however, potentially be obscured by suboptimal and/or inconsistent choices made in complex processing pipelines required to reach a quantitative estimate of radioligand binding. Variations in subject selection, experimental design, data acquisition, preprocessing, and statistical analysis may lead to different outcomes and neurobiological interpretations. We here review the approaches used in 105 original research articles published by 21 different PET centres, using the tracer $[^{11}\text{C}]$DASB for quantification of cerebral serotonin transporter binding, as an exemplary case. We highlight and quantify the impact of the remarkable variety of ways in which researchers are currently conducting their studies, while implicitly expecting generalizable results across research groups. Our review provides evidence that the foundation for a given choice of a preprocessing pipeline seems to be an overlooked aspect in modern PET neuroscience. Furthermore, we believe that a thorough testing of pipeline performance is necessary to produce reproducible research outcomes, avoiding biased results and allowing for better understanding of human brain function.

Key words: Positron Emission Tomography; preprocessing; kinetic modeling; $[^{11}\text{C}]$DASB; data sharing
INTRODUCTION

Positron Emission Tomography (PET) imaging with selective radiotracers has been extensively used as a tool for novel neuroscience research. PET neuroimaging often utilizes complex workflows, with multiple stages ranging from subject selection, experimental design, data acquisition, preprocessing, statistical analysis to the final neurobiological interpretation. However, while most published articles utilizing molecular neuroimaging have mainly focused on extracting neuroscientifically relevant results, no articles have, to our knowledge, investigated the extent to which these findings may be significantly influenced by different sets of preprocessing steps ("preprocessing pipeline/stage") applied while analyzing the data. A preprocessing pipeline in neuroimaging commonly refers to a set of steps used to denoise and remove artifacts in the data for subsequent statistical analysis (e.g. motion correction and outlier detection), thereby improving the overall quality of the data. However, choices made at any stage of a neuroimaging workflow may significantly affect the chosen steps in the preprocessing pipeline, limiting the generalizability of any preprocessing pipeline studied in isolation from a fixed neuroimaging workflow. For example, medical conditions preventing patients from staying still in the scanner (e.g. Parkinson’s Disease) may require more extensive correction of head movements in the preprocessing stage compared to healthy subjects.

Notably, to date, preprocessing developments in the PET neuroimaging community have often been focusing on an even more limited point of view than examining the overall preprocessing pipeline in isolation. The optimization of preprocessing steps typically entails only limited test data, and is often performed with the aim of optimizing only a single preprocessing step (e.g. kinetic modeling) without explicitly attempting to address potential interactions with other preprocessing steps, or with other stages of a given workflow. Examples of such potential confounds include; subject selection (e.g. of healthy versus diseased cohorts), differences in scanner resolution, duration of a
scanning session, dynamic framing, injected dose/injected mass (data acquisition), differences in image reconstruction, motion correction, different kinetic modeling approaches used to estimate the availability of receptors/transporters (preprocessing), and different statistical model choices used to test for group or longitudinal differences (statistical analysis).

We here review and quantify the impact of the various data acquisition and preprocessing pipeline choices used to quantify the same biological target, using the serotonin transporter and the radioligand [11C]DASB as exemplary case. We chose to specifically focus on the serotonin transporter using [11C]DASB, because this is a well established radioligand in the field and has been used extensively to study various aspects of brain function ranging from schizophrenia to epilepsy (Figure 1).

Since [11C]DASB was first described in 2000 through the end of March 2017, nearly 170 [11C]DASB PET papers have been published, and this number is growing by 1-2 articles per month. We systematically searched PubMed for studies using “[11C]DASB and PET” in the time period between September 2000 to March 2017, and found a total of 169 publications. Non-human studies (N = 49), reviews (N = 4), and methodological papers (N = 12) were excluded due to substantial differences in acquisition and preprocessing, leaving 104 publications eligible for scrutiny. One paper not identified by the search, was subsequently added, summing up to a total of 105 original research articles. We catalogued the different sample sizes and patient cohorts investigated in the published [11C]DASB studies, the various data acquisition techniques used, and the preprocessing steps applied to the data. We systematically outline and quantify the impact of the remarkable variety of ways in which researchers are currently performing these studies, while implicitly expecting generalizable results across research groups. Although this review specifically focuses on the radioligand [11C]DASB, the underlying considerations apply to any given PET or SPECT.
radiotracer, as optimal neuroimaging workflows are highly dependent on the inherent characteristics of the radioligand of interest.

DATA ACQUISITION WORKFLOW AND OUTCOME

In order to investigate the variability in data acquisition and preprocessing, we provide an overview of the different acquisition and preprocessing choices that have been made in previous studies. We also examine how differences in reported findings might be influenced by differences in methodologies. For this purpose, we extract the $[^{11}C]$DASB PET binding potentials ($B_{ND}$) in striatum and Anterior Cingulate Cortex (ACC) as well as other relevant information from 90 studies with healthy controls encompassing a total of 1856 healthy controls. We chose to examine the healthy controls only because they serve as null data, achieved with different experimental designs. The available $B_{ND}$'s and standard deviations from the published studies were used as the dependent variable in separate linear models, correcting for the number of healthy controls included in the study, age, age standard deviation, choice of MRI hardware, choice of PET hardware, number of frames, injected dose, motion correction, choice of volumes-of-interest (VOI), and choice of kinetic modeling (Table S1). All covariates were standardized columnwise to have mean 0 and standard deviation 1. To limit the degrees of freedom we did not specify any interactions in the linear model, despite their obvious existence (e.g. PET scanner x injected dose).

The omission of potential interactions is a limitation of the current analysis, but is driven by limited data.

Development of $[^{11}C]$DASB in PET Neuroimaging and Subject Selection

N,N-dimethyl-2-(2-amino-4-cyanophenylthio)benzyl-amine, or more commonly referred to as DASB, was developed by Wilson and colleagues at the Center for Addiction and Mental Health,
Toronto Canada, and their first-in-human study was published in 2000\textsuperscript{1,3}. Their preliminary analyses indicated that DASB radiolabeled with carbon-11 effectively penetrated the blood-brain barrier, and displayed retention characteristics in accordance with the known anatomical distribution of cerebral serotonin reuptake sites. In any aspect, \([^{11}\text{C}]\text{DASB}\) turned out to be a highly suitable radiotracer to map the serotonin transporter using dynamic 4D PET imaging. Since 2000, \([^{11}\text{C}]\text{DASB}\) has been used extensively, so far by 21 PET centres, investigating various aspects of brain function. In Figure 1, we provide a timeline of the number of healthy controls and patient cohorts that have been investigated and published using \([^{11}\text{C}]\text{DASB}\). Whenever possible, we have attempted to correct the data in Figure 1 for duplicates, to encounter only the net number of included healthy volunteers from the \([^{11}\text{C}]\text{DASB}\) PET studies. Our analysis of the reported values from the literature suggests no statistical evidence for an impact of the number of subjects included in the study on BP\textsuperscript{ND} or between-subject variation. We found a trend for an association between age and between-subject variation of ACC BP\textsubscript{ND} (P = 0.075), suggesting that between-subject ACC BP\textsubscript{ND} is more variable in elderly than in young controls. While this may be caused by cortical atrophy or other age-related disorders, it warrants further examination of how the impact of acquisition and preprocessing choices may vary as a function of age.

**Positron Emission Tomography Scanners and Reconstructions**

We found that in the 21 centres, 9 different scanners have been used (Figure 2). The first paper published by Houle et al. 2000\textsuperscript{3} (Center for Addiction and Mental Health, Toronto, Canada) presented data acquired with a Scanditronix/GEMS PC2048-15B 2D brain PET scanner, a state-of-the-art scanner from the late 80’s. The data were attenuation corrected and reconstructed using filtered back-projection (FBP). The performance of the Scanditronix/GEMS PC2048-15B scanner
was evaluated in 1989 by Holte and colleagues\textsuperscript{4}, reporting the in-plane axial full-width half maximum (FWHM) to be 5.9 mm for direct planes, and 5 mm for cross planes in the central area of the field-of-view (FOV). In addition, with a coincidence timing window of 12.5 ns and a lower energy threshold of 300 keV, the average sensitivity (including 16\% scatter) was 251 cps·MBq\textsuperscript{-1}·mL\textsuperscript{-1} for the direct planes, whereas the average sensitivity was 351 cps·MBq\textsuperscript{-1}·mL\textsuperscript{-1} for the cross planes. After the study by Houle et al. \textsuperscript{2000}\textsuperscript{3} and until 2008, a total of six DASB studies were conducted with the GEMS scanner, all published by the Center for Addiction and Mental Health, Toronto, Canada, including the first \textsuperscript{11}C DASB study discussing quantification strategies by Ginovart et al. \textsuperscript{2001}\textsuperscript{5}.

After the first publication of \textsuperscript{11}C DASB, attention increased substantially around the World, motivating researchers to investigate new hypotheses related to the serotonin transporter. Consequently, resulting in a large number of different scanners used to map \textsuperscript{11}C DASB binding. Ogawa and colleagues from Japan\textsuperscript{6} used an Eminence SET-3000GCT/X PET scanner (performance evaluated in 2006\textsuperscript{7}) to investigate the effects of Tramodol for pain treatment; this is currently the only published \textsuperscript{11}C DASB study using this scanner. Another Japanese group\textsuperscript{8} used a SHR12000 tomograph from Hamamatsu Photonics (performance evaluated in 2002\textsuperscript{9}) to study the serotonin transporter in Alzheimer’s Disease; this is the only \textsuperscript{11}C DASB study published using this scanner. Both of these scanners operate in 3D-mode, providing an excellent in-plane spatial resolution ranging from approximately 3 mm FWHM in the center of the FOV to 5 mm FWHM at 10 cm off center. This makes them somewhat ideal PET scanners to capture cortical features of the serotonin transporter, as on average, cortex is only 3 mm thick\textsuperscript{10}.

The National Institute of Radiological Sciences in Chiba Japan, published two \textsuperscript{11}C DASB studies in 2006\textsuperscript{11} and 2010\textsuperscript{12}; these were the only studies using an ECAT47 PET scanner. This PET scanner also operates in 3D-mode, but unlike the Eminence and Hamamatsu...
scanners, which have an axial resolution of 3-5 mm FWHM, this scanner has an in-plane axial spatial resolution of 6.2 mm in the center of the FOV, and 7.2 mm at 10 cm off center. This means that the spatial resolution is almost half as good in the center of the FOV, and more severe partial volume effects (PVE) are to be expected. Several integrated PET/CT systems have also been used to map the serotonin transporter, including the Biograph HiRez and the Biograph TruePoint, both manufactured by Siemens, having a spatial resolution of approximately 4.5 mm. A total of 7 published $[^{11}\text{C}]$DASB PET papers have used this scanner. The most commonly used PET scanners for measuring $[^{11}\text{C}]$DASB are the ECAT EXACT HR+ PET scanner from Siemens (performance evaluated in 1997), the GE Advance PET scanner from General Electric (performance evaluated in 2002) and the High Resolution Research Tomography (HRRT) PET scanner from Siemens (performance evaluated in 2002). Van Velden and colleagues directly compared the performances of the HRRT and HR+ scanner in 2009. The in-plane spatial resolution of the HRRT is 2.3-3.4 mm FWHM, whereas the in-plane spatial resolution of the HR+ scanner is 4.3-8.3 mm FWHM.

Furthermore, the sensitivity of the HRRT is higher than that of the HR+ scanner, 39.8 kcps·kBq$^{-1}$·mL$^{-1}$ compared to 21.9 kcps·kBq$^{-1}$·mL$^{-1}$, respectively.

Finally, the GE Advance PET scanner from General Electrics, has an in-plane spatial resolution of 4.4 mm FWHM in the center of the FOV, and 6 mm FWHM in the outer FOV. The GE Advance scanner sensitivity is approximately 27.6 kcps·kBq$^{-1}$·mL$^{-1}$, placing it as the 2nd best performing scanner regarding scanner sensitivity, with the HRRT in 1st place, and the ECAT EXACT HR+ in 3rd place.

Our analysis of the reported values from the papers revealed that $\text{BP}_{\text{ND}}$ in both striatum and ACC was associated with the PET-scanner used in the study. This was also true for the between-subject standard deviation of $\text{BP}_{\text{ND}}$ (Figure S1). A higher scanner resolution was associated with
higher BP_{ND}’s and higher between-subject standard deviations (P = 0.027).

This means that more subjects are needed to detect a statistical difference in a group analysis.

On the other hand, the larger between-subject variability may also be caused by increased ability to
detect subject-specific binding, as reflected by a higher resolution scanner.

The HRRT scanner has high sensitivity, but is limited by relatively small detector elements which
means, that the number of acquired counts is lower than in other scanners, potentially resulting in
more noisy data.

Moreover, the spatial resolution differs significantly between scanners with the HR+ being
nearly isotropic, whereas the GE Advance has a much better axial resolution than transaxial
resolution (non-isotropic voxels). This means that the resolution is dependent on the orientation of
the image, resulting in different spill-over effects of the tracer in different directions. This makes it
difficult to correct for PVEs, and may consequently interact with subsequent preprocessing steps
such as motion correction, co-registration and normalization to a standard space.

Instituting a more standardized policy for the reporting/usage of PET scanner performances
should ensure that future readers are better able to effectively evaluate and understand the potential
biases and uncertainties of the data. We note, that researchers often only report the FWHM in the
center of the FOV when publishing papers, creating a limited/biased interpretation if cortical
regions are the primary region of interest.

In addition, reviewers should pay special attention to the use of 2D reconstruction over 3D
reconstruction, non-isotropic over isotropic resolution, and if any additional smoothing steps are
applied to the data (e.g. Strecker et al. 2011^{20}), as these steps significantly degrade the spatial
resolution.

Anatomical information from Magnetic Resonance Imaging
Several different techniques have been used to provide the anatomical information needed to guide the functional information provided by the PET data. The most common procedure is to acquire an anatomical T1-weighted Magnetic Resonance Image (MRI) as a reference image and spatially align the two images (co-registration). However, the field-strength of the MRI scanner will have an impact on the reconstructed MRI image, affecting both the subsequent parcellation of the brain into anatomical subregions and the co-registration to the PET data. Tradeoffs between spatial and temporal resolution and signal/noise also matter, but this topic is considered beyond the scope of this paper.

In the reported \(^{11}\text{C}\)DASB studies, the field-strength used for the MRI scanners includes 0.3T\(^8\), 0.5T\(^21\), 1.5T\(^22\), 3.0T\(^23\) to 7.0T\(^24\). When no MRI is acquired, the PET image is most often either normalized to a common atlas space (e.g. Lanzenberger et al. 2012\(^25\)), in which generic regions have been predefined, or manual delineations are applied directly on the PET image. This requires additional smoothing or resampling steps, and interacts with nearly all data acquisition steps such as the resolution of the PET scanner, duration, framing and injected dose (Figure 2, described in detail below). In addition, because a standard MRI atlas does not follow the subject-specific anatomy (i.e. cortical folding patterns), it is likely that this procedure will exacerbate the PVE when evaluating regional PET distributions compared to when a subject-specific MRI is available.

For example, if the PET-MRI co-registration is inaccurate, the PET signal might seem to originate from white-matter signal instead of gray matter (GM), or vice versa. In total, we found six different methods whereby anatomical information has been extracted. In our analysis, we found no evidence for an impact of choice of MRI-scanner on BP\(_{\text{ND}}\) or between-subject variation.

The two to date most widely published methods are acquisitions of either a 1.5T (43%) or 3.0T (32%) T1-weighted MRI (Figure 2). Not unexpectedly, the more recent publications tend to use 3T MRI scanners because most institutes regularly update their MR scanners, with newer ones having
higher field strength. However, the extent to which differences in MRI acquisition might affect the final outcome of a complex workflow is largely unknown.

**Data Acquisition (Duration, Framing, Injected dose, Reconstruction)**

Variations in $[^{11}\text{C}]\text{DASB}$ PET data acquisition are distributed across a parameter space containing the (1) time duration of the scanning session, (2) dynamic framing (time-sequence), (3) injected dose and (4) PET reconstruction. Unless list mode acquisitions are available, dynamic PET studies are mostly acquired for a fixed time duration, with multiple 3D-frame acquisitions distributed over a pre-defined time period. However, the chosen framing varies substantially from study to study, and a total of 17 different sequences have been used so far, i.e., framing $\in \{17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 30, 33, 35, 36, 38, 50\}$ frames. This choice will affect the signal distribution within the acquisition space (FOV), since reduced frame length will result in a reduction in true counts per frame, especially in late frames where the radioactive tracer has decayed due to the half-life of $^{11}\text{C}$ (i.e., $\lambda_{1/2} = 20.3$ minutes). We identified a positive association between number of frames and striatal $\text{BP}_{\text{ND}}$ and ACC $\text{BP}_{\text{ND}}$ but closer inspection revealed that this observation was driven mainly by the HRRT-scanner settings from Denmark using 36 frames and the high-resolution SHR12000-scanner from Japan using 38 frames. As high-resolution scanners increase the binding, this effect may be at least partially explained as a scanner x frames interaction.

The scan duration of these dynamic PET studies also varies substantially, ranging from 30 minutes to 120 minutes, i.e., duration $\in \{30, 60, 80, 90, 95, 100, 110, 120\}$ minutes. Although, Ogden et al. 2007$^{26}$ argued that 100 minutes of scanning time was sufficient, this recommendation has not been followed in all subsequent studies, with most studies choosing 90 minutes of acquisition time (Figure 2). The required duration of scanning time, however, depends on the neuroscientific question, as various brain structures will have different uptake dynamics$^{27}$. 
The injected dose also varies substantially between approximately 100 MBq to 740 MBq, and the dose varies substantially not only between studies, but also between subjects (Table S1).

All studies reported high molar radioactivity, however it has never been formally established in a test-retest study with substantially different doses, if a higher injected mass (or mass/kg body weight) leads to a reduction in cerebral $^{11}$C]DASB binding. In a large sample of 108 individuals from our own group, we found no evidence for an association between global $^{11}$C]DASB binding and injected mass/kg (McMahon et al. 2017, unpublished work). In our analysis, we found no evidence for an impact of injected dose on BP$_{ND}$ or between-subject variability.

Depending on the scanner sensitivity, the injected dose can impact signal-to-noise ratio (SNR). Information about, e.g., range of average counts per minute per study/subject could be interesting information to have access to for analysis, as we did not have access to the individual injected doses, but rather the within-study average injected doses.

The reconstruction of the PET images from the scanner has also been differently performed. Morimoto et al. 2006$^{28}$ compared $^{11}$C]DASB binding in 7 healthy subjects using images reconstructed with either a filtered backprojection (FBP) algorithm or the ordered subsets expectation maximization (OSEM) algorithm.

The study by Morimoto and colleagues was executed using the data acquisition workflow parameters: ECAT EXACT HR+, 1.5T MRI, 90-min dynamic PET acquisition in 2D-mode, 27 frames, injected dose of 170.2 ± 56.1 MBq. While there have been several reports suggesting a small bias using some versions of OSEM$^{29,30}$, Morimoto reported no statistically significant differences in any regions between images reconstructed with FBP and OSEM, suggesting that these two algorithms may be used interchangeably in the reconstruction of 4D PET data. Certain PET scanners (e.g. the HRRT) do not allow for a direct use of FBP due to the inherent geometry of the scanner, thereby restricting image reconstruction to the use of iterative reconstruction
techniques such as the OSEM algorithm. However, techniques have been developed that allow 3D-FBP on the HRRT, but due to poor noise performance they are not widely used. To summarize the data acquisition workflow section, the most widely published workflow consists of: 1.5T MRI (43%), ECAT EXACT HR+ (43%), 90 minutes acquisition (65%), 26 frames (17%), and FBP to reconstruct the 4D PET data (72%).

THE "PREPROCESSING PIPELINE" FOR $^{11}$C DASB PET QUANTIFICATION

Motion Correction

Motion correction (MC) algorithms for dynamic PET studies have been developed to remove inherent motion artefacts from the data. The most popular head MC technique is between-frame-correction where either all or a subset of the remaining images are registered to a chosen reference image. Of the 105 studies, forty-three studies (41%) leave out any type of motion correction, arguing that fixing the subject in the scanner using, e.g., a thermoplastic mask sufficiently limits motion. Twenty-nine studies use between-frame-correction to correct for motion without explicitly specifying the exact procedure (e.g. James et al. 2017). Twenty-one studies use between-frame-correction to correct for motion, by aligning all frames to a frame with high SNR (e.g. Frokjaer et al. 2015). Ten studies use either a mean or a summed PET image over all frames to correct for motion (e.g. Cannon et al. 2007), and two studies use either a partially summed image or a reference frame to perform between-frame motion correction, but only frames where the researcher observed motion are aligned, leaving the frames without motion untouched. The latter method not only introduces a user-dependent bias, it also raises the question: given that motion is present in the data, how much movement is needed in order to perform MC?

Overall, this results in five different ways in which motion correction has been applied/not applied in the $^{11}$C DASB literature. In our analysis, we grouped the analysis into
motion vs. no motion and observed a trend for significance (β = -0.032, P = 0.064) of the use of motion correction and striatal between-subject variability, suggesting that motion correction lowers between-subject variability in the striatum with 0.035 compared to without motion correction (Figure S2). This translates into 26% fewer subjects needed in a group analysis to obtain similarly powered statistical tests (See calculation in supplementary).

MC in the absence of motion will lead to some degree of smoothing, which may to some extent account for the observed reduced between-subject variability. In addition, potential effects of motion within a frame are often neglected, even though several solutions have been suggested such as MOLAR75 (Motion-Compensation OSEM List-mode Algorithm for Resolution-Recovery Reconstruction) or Tracoline76 (List-mode PET motion correction using markerless head tracking), given that list-mode data is available.

MC is often carried out using different software packages (AIR, FSL and SPM), which all have different implementation and precision of similar methods but based on different cost functions. To our knowledge, the effect of various software packages on MC performance has not yet been investigated in dynamic 4D PET imaging. In addition, frame-by-frame motion correction without re-doing the image reconstruction may result in errors in attenuation correction, which is often neglected (Van den Heuvel et al. 201373).

Co-registration

Accurate co-registration of PET and MR images is an important step, not the least when PET partial volume correction (PVC) and parcellation of regions is carried out, when integrating multimodal neuroimaging data36. Ninety-eight percent of all studies use a normalized mutual information (NMI) registration algorithm to perform the co-registration but the explicit procedure differs across studies that use various software packages, including FSL and SPM. Each co-registration technique is
based on a cost-function, aiming to minimize the registration error (e.g. sum of least-squares or mutual information) of the two datasets being aligned (MRI and PET). This cost-function is often based on shared information between the two datasets being aligned (e.g. cortical boundaries), making them somewhat dependent on the intensity distribution and resolution of the acquired data. The remaining two percent\textsuperscript{37,38} used a boundary-based registration (BBR) algorithm to co-register the T1-weighted MRI with the PET image. BBR also contains a mutual information component, but puts an additional cost on the cortical boundaries being aligned. The co-registration preprocessing step potentially depends on the spatial- and temporal distribution of the PET signal, and will therefore be sensitive to the chosen cost-function. For example, the serotonin transporter is only modestly expressed in the neocortex, and the boundary-based algorithm may therefore not be the optimal registration algorithm to capture cortical folding patterns, particularly not if the PET scanner resolution is limited. In addition, brain areas located in close vicinity to ventricles and cerebrospinal fluid (CSF) will suffer more from PVEs, depending on the resolution of the PET scanner and the radiotracer being used, especially when data with non-isotropic spatial resolution was acquired.

**Delineation of Volumes of Interest**

Many neuroimaging experiments are based on hypotheses relating to specific anatomical brain regions, often referred to as volumes-of-interest (VOIs). As mentioned previously, for PET this generally requires co-registration with a structural MRI scan with anatomical labels. However, there is currently no consensus in the \textsuperscript{[11]C}DASB PET community about which atlas generates the best set of VOIs. Whereas a single study used the probabilistic Harvard-Oxford atlas to delineate VOIs\textsuperscript{31}, 14 published papers used PVElab, which is a data-driven anatomical probability-based labeling approach based on MRI templates from 10 healthy volunteers (e.g. Frokjaer et al. 2009\textsuperscript{39}).
Nine studies used the Desikan/Killiany atlas (e.g. from FreeSurfer) which involves a data-driven technique, providing the researcher with a subject-specific anatomical labeling, given that they have acquired a subject-specific T1-weighted MRI (e.g. Ganz et al. 2017\textsuperscript{37}). Seven studies used the anatomical automatic labeling (AAL) atlas offered by the SPM software (e.g. Savli et al. 2012\textsuperscript{40}). The AAL atlas does not provide unique subject-specific anatomical labeling, but can be used for group analyses, where all subject-specific PET scans have been normalized to AAL standard space. Seven studies used the Hammers atlas, which is a probabilistic brain atlas based on 83 manually delineated regions drawn on MR images of 30 healthy subjects in native space, subsequently spatially normalized to a standard brain from the Montreal Neurological Institute (MNI) (e.g. Hinz et al. 2008\textsuperscript{41}). Fourteen studies used an atlas-based procedure, without explicitly stating the exact labeling approach, mostly being based on local procedures and study-specific atlases (e.g. Takano et al. 2006\textsuperscript{42}). These atlases are often based on data obtained from a set of young and healthy subjects, in which manually delineated regions have been drawn in native subject-space prior to spatial normalization to MNI-space. Ten published studies used an "automatic method" to obtain VOIs, stating that the anatomical labeling was unbiased with respect to any user interactions (e.g. Tyrer et al. 2016\textsuperscript{14}).

Somewhat surprisingly, 38% of all published [\textsuperscript{11}C]DASB studies included in this review, manually define their own VOIs, also in some recent studies\textsuperscript{43,44}.

In our analysis, we found a striatal BP\textsubscript{ND} x VOI interaction, suggesting that some definitions of volumes produce either higher or lower BP\textsubscript{ND} compared to others (Figure S3). Since this step may interact with all previous steps, we are cautious to make any firm conclusions based on this. Hammers atlas and manual delineations contributed to the most variation, but should ideally also be split into additional sub-categories depending on the operational criteria, and whether the delineation was performed in PET or MRI space (Table S1). In addition, it is expected that the
variability will increase as the size of the VOI decreases, but with limited reports on size of VOIs in
the published studies, this reduces our ability to assess the impact of atlas choice. Nevertheless,
what we can conclude is that the choice of atlas can produce widely different outcomes, as
highlighted in Table S1.

Even though manual anatomical labeling seems to be the most popular, it may impose an interrater
variability/bias in the subsequent data analysis and interpretation, unless well-defined operational
criteria and blindness to subject diagnosis are applied.

Another potential issue with both manual delineation and atlases is that even though the tracer
distribution within an anatomical VOI is assumed to be homogeneous, this is often not the case and
accordingly, structural homogeneously defined VOIs may therefore misrepresent the radioligand
concentration within that region (e.g. the thalamus). Correct anatomical labeling is critically
important in many dynamic PET studies, because the PET data suffers significantly from PVEs.
We recommend that researchers provide explicit specifications about VOI definitions in the
supplementary material, and if possible, attach the 3D anatomical labelings in appropriate formats.
This is an approach also supported by researchers in the fMRI field45.

Partial Volume Correction

In PET studies, it can be difficult to assess the extent to which an observed difference in PET signal
is caused by a change in the imaging target distribution, if it is due to less GM, or if it is due to
limited PET scanner resolution causing the PET signal to spill in or out of relatively homogeneous
tissue regions. Partial Volume Correction (PVC) is not commonly used in [11C]DASB PET imaging
(Figure 3). Only four published [11C]DASB studies have used Muller-Gartner PVC, to correct for
partial volume effects46,47,48,74.
If there is little evidence for differences in brain volumes, the application of PVC techniques may lead to noise amplification, and extreme care should therefore be taken when interpreting the results. In addition, PVC is MR scanner and sequence dependent due to variability of segmentation results from the MRI. For an in-depth discussion of PVC techniques in PET imaging we refer the reader to the paper by Erlandsson et al. 2012.

Quantification of $[^{11}\text{C}]\text{DASB}$ PET data

The final step in processing of $[^{11}\text{C}]\text{DASB}$ PET data is kinetic modeling which is applied to the preprocessed 4D PET data. All the kinetic modeling approaches used for quantification of $[^{11}\text{C}]\text{DASB}$ PET data are displayed in Figure 3, including the frequency of their use. The quantification of tracer kinetics of the serotonin transporter in vivo has been applied extensively and in various formats, providing information about binding in specific VOIs. The gold standard is to obtain arterial blood samples in parallel with the dynamic PET scan, providing an arterial input function (AIF) for subsequent kinetic modeling.

However, the use of arterial sampling requires invasive techniques, which often imposes additional discomfort to the subject being scanned. Furthermore, blood sample analysis (on-line vs. manual sampling, including frequency of sampling), metabolite estimation (HPLC or fraction-collector) and interpolation (fitting a power function) can add additional variation to the data analysis. Two-tissue compartment modeling (2TCM) with an AIF is considered state-of-the-art in the PET literature, but once validated, tissue reference methods may be used instead. The kinetic models used in $[^{11}\text{C}]\text{DASB}$ neuroimaging include both reference tissue methods and methods with an AIF (Figure 3). Reference tissue methods obviate invasive arterial sampling, but they rely on the assumption and identification of a reference region with non-specific binding characteristics.

In the $[^{11}\text{C}]\text{DASB}$ literature, cerebellum (possibly excluding vermis) serves as a reference region,
because it is considered to be devoid of serotonin transporters. However, there is currently no consensus among researchers about the validity of cerebellum as a reference region. Some researchers argue for and others against, as DASB binding in the cerebellum has been shown to be displaced by SSRIs. Even among the researchers using cerebellum as a reference region, there is no consensus about how exactly the reference region must be defined. A recent investigation of cerebellar heterogeneity and its impact on PET data quantification of 5-HT receptor radioligands, based on a large sample of 100 [11C]DASB HRRT scans, concluded that there are differences in radioactivity uptake between cerebellar subregions.

New kinetic models are continually being developed and refined, and to date nine different approaches have been applied. Published studies that include blood sampling and use of an AIF over the last couple of years have become less common, with currently four different approaches used to perform the kinetic modeling of [11C]DASB. Only three published studies have used a one-tissue compartment model (1TCM) with an AIF to capture the features of the serotonin transporter. Eight studies used a 2TCM with an AIF (e.g., van de Giessen et al. 2014), and eight studies have used the Logan method with an AIF (e.g., Murthy et al. 2010). Finally, the Likelihood Estimation Graphical Analysis (LEGA) method (maximum likelihood estimation of the Logan) using an AIF has been used in eight published studies, including one of the three test-retest studies that evaluate reproducibility of [11C]DASB, as discussed in more detail below.

Forty-four published studies (38% in total) have used the multilinear reference tissue model 2 (MRTM2), developed by Ichise in 2003, to quantify tracer kinetics of [11C]DASB (e.g., Fisher et al. 2017). Twelve studies used the simplified reference tissue model (SRTM) developed by Lammertsma and Hume in 1996, and eight studies a constrained version of the same model, SRTM2. The non-invasive Logan method is used in 22 published studies. Finally, four studies have used
the ratio of standardized uptake values (SUVR) defined by the SUV of a given VOI to the SUV of a reference VOI (i.e., the cerebellum has been used as a reference for DASB binding). When using SUVR as a direct measure for binding, the arterial input concentration is assumed to have a consistent shape between studies/subjects, and the area under the arterial input curve is assumed to be proportional to the injected dose / kg body weight. This assumption applies to all reference techniques, but may be violated as a function of age and/or disease. In terms of equilibrium, one should be careful when selecting the time frame of interest, as this should coincide with the transient equilibrium of the tracer in all subjects.

The SUVR also depends on the rate of peripheral clearance of the tracer; unlike parameters derived from most kinetic models of brain uptake and binding, SUVR is not purely a function of brain parameters, though the extent to which differences in clearance between subjects affects study results has not been carefully examined for \[^{11}\text{C}]\text{DASB.}

Studies that have used SUVR include Lee et al. 2015, Hesse et al. 2012, Ginovart et al. 2001 and Houle et al. 2000. To sum up, nine different methods have been applied to quantify \[^{11}\text{C}]\text{DASB PET. In our analysis, we find that the choice of kinetic model was associated with between-subject variability of ACC BP}_{ND}. SRTM and non-invasive Logan (with Muller-Gartner PVC) produced the highest between-subject variabilities (Figure S4). When adding BP_{ND} as a covariate in the analysis, we also found a trend for a positive association \((P = 0.11)\) between variation and BP_{ND}, highlighting a potential bias-variance trade-off in ACC BP_{ND} (Figure S5). The identified bias-variance trade-off as a function of neuroimaging workflow warrants further investigation.

TEST-RETEST STUDIES FOR \[^{11}\text{C}]\text{DASB PET}
To date, three test-retest studies for $[^{11}C]$DASB PET imaging have been published$^{26,56,67}$. These studies involve two different scanners (ECAT HR+ and GE Advance), one fixed time duration (120 minutes), two different dynamic framings (21 and 33 frames) and a range of 185 MBq to 740 MBq in injected dose. The studies included between 8 to 11 healthy subjects (aged 18-50) with a nearly 50/50 gender distribution. All test-retest scans were performed on the same day.

Two out the three studies used an AIF for the kinetic modeling, whereas one study used MRTM2 with cerebellum as reference region. Ogden et al. 2007$^{26}$ reported that 100 minutes of scanning time was sufficient to obtain stable parameter estimates, and that the LEGA kinetic modeling approach produced the best results. However, the LEGA method produced a median percent difference in test-retest binding of approximately 20% ($n = 11$, range: 11 - 39.6%), when taken across all subjects and all VOIs. The median Intraclass Correlation Coefficient (ICC), was approximately ICC = 0.8 (range: 0.455 - 0.926), taken across all subjects and all VOIs, with the highest ICC’s in the dorsal caudate, thalamus and midbrain. Frankle et al. 2006$^{56}$ obtained slightly higher ICCs compared to Ogden et al. 2007$^{26}$ with a median ICC of 0.93 ($n = 9$, range: 0.79 - 0.97). Kim et al. 2006$^{57}$, investigating the reproducibility of $[^{11}C]$DASB binding modeled with MRTM2 ($n = 8$), also used ICC as performance metric for test-retest reliability, including the additional performance metrics test-retest bias and test-retest variability. The results showed a significant negative bias in binding across test-retest, and high test-retest reliability for regions such as striatum, thalamus, temporal cortex and occipital cortex (ICC = 0.84). In contrast, poor test-retest reliability measures were obtained in the raphe and frontal cortex (ICC = 0.445). The reported negative bias across test-retest was barely discussed by Kim et al. 2006 and neither Frankle et al. 2006 nor Ogden et al. 2007 observed a negative test-retest bias with lower binding at retest.

The overall conclusion by Kim et al. 2006 was that the MRTM2 was reproducible and reliable for $[^{11}C]$DASB studies.
Notably, these test-retest studies were all performed on a relatively small sample and they demonstrate that some methods (i.e. Ogden et al. 2007 (LEGA) and Kim et al. 2006 (MRTM2)) are better or equally performing compared to other methods. However, the chosen performance metrics are not consistent across test-retest studies, and no attempt is explicitly made to address possible interactions with other preprocessing steps and/or other steps of the workflow (i.e. subject selection and data acquisition), as data acquisition and preprocessing are not consistent across the three test-retest studies. For example, Kim et al. 2006 used a summed image to perform frame-based MC, whereas Frankle et al. 2006 used a reference frame. However, while Frankle and colleagues used VOIs manually determined on MRIs according to well defined operational criteria in conjunction with automated gray/white/CSF segmentation in cortex, Kim et al. 2006 instead used manual delineations without specifying the operational critera to obtain the VOIs. In addition, even though Frankle et al. 2006 and Ogden et al. 2007 used the same PET scanner to acquire the data, images were reconctructed into 1.7 x 1.7 x 2.4 mm (non-isotropic) and 2.5 x 2.5 mm, respectively, with no specification on the z-direction in the latter study. All these modifications from study to study, make it difficult for the reader to infer whether reported methodological improvements are causally related to the new proposed method, or if it is due to a difference in data acquisition and/or preprocessing, limiting the generalizability to other neuroimaging workflows and studies.

CONCLUSIONS
In this review, we highlight the remarkable variety of ways in which researchers are currently performing complex neuroimaging studies, while implicitly expecting generalizable results across research groups. We systematically reviewed 105 published $[^{11}C]$DASB studies from 21 different PET centres, outlining differences in subject selection, data acquisition and preprocessing. Data sharing initiatives may significantly contribute to the understanding of the generalizable impact on
such complex workflows, as the combined effects resulting from subject selection, data acquisition and preprocessing are unclear. We still need to understand the importance of bias-variance tradeoffs in neuroimaging experiments, and how neuroimaging workflows can be optimized for particular neuroscientific questions. The purpose of this study was not to identify a definitive PET preprocessing pipeline, but rather to establish workflow-dependent effects on binding and variation.

It is to be expected that the application of a new preprocessing pipeline will lead to different absolute binding measures, but the important question is whether the outcome of a study (i.e. difference between patients and controls) will remain.

In order to evaluate the extent to which any of the methodological factors described in this review matters, one needs to consider the given study aims. For example, if an investigator wishes to compare age effects on the serotonin transporter in the striatum between two studies, it might be tempting to use both data sets, given that the methodology is internally consistent. However, while the researcher may not be able to combine the two sets of data, he/she may be able to use the two data sets separately, assuming that the derived parameters while different, are scalable. For future data sharing initiatives, it would be beneficial in a large and complete data set across a large number of subjects to assess which differences the various methodological variations can lead to, e.g. how much of a difference does variation in scanner resolution impact on, e.g., the striatum.

Our review focused on the radioligand [11C]DASB, but the same considerations underlying the [11C]DASB workflows could be made for any given PET or SPECT radiotracer. The aim of our paper is to highlight the need for transparency, reproducibility and to support future data sharing opportunities in the PET neuroimaging community. It is our hope that this work can also be used as a tool for future studies to evaluate the extent to which a given study deviates significantly from the current literature. From the current literature, it can be difficult to infer whether an observed change
is physiological, or if it is driven by changes in subject selection and/or data acquisition and/or preprocessing. Data acquisition and preprocessing pipelines and their experimental interactions seem to be an overlooked aspect in modern PET neuroscence, and we believe that such testing is necessary in order to reliably provide new insights into human brain function.

DISCLOSURE/CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPLEMENTARY MATERIAL

Supplementary material for this paper can be found at

http://jcbfm.sagepub.com/content/by/supplemental/data

REFERENCES


Titles and legends to figures

Figure 1: Timeline of number of patient and healthy controls in the 105 published $^{[1]}C$DASB studies. The colors indicate either healthy controls, or a specific disorder as a function of time and sample size. Abbreviations: ADHD (Attention-Deficit/Hyperactive Disorder), MDD (Major Depressive Disorder), MDMA (ecstasy), HIV (Human Immunodeficiency Virus), OCD (Obsessive Compulsive Disorder), SAD (Seasonal Affective Disorder), PTSD (Post-Traumatic Stress Syndrome), PD (Parkinson’s Disease).

Figure 2: Schematic overview of the different data acquisition workflows used to acquire dynamic $^{[1]}C$DASB data. The workflow consists of scanners providing anatomical information, i.e., MRI scanners at various field strengths (Tesla), various PET scanners, duration of the dynamic PET acquisition, frame sequence used to temporally acquire 4D $^{[1]}C$DASB data, injected dose (ranging from approximately 185-740 MBq), and finally the reconstruction methods used to reconstruct the 4D PET sequence. The colors indicate the frequency per step that has been applied in a $^{[1]}C$DASB PET study out of the total 105 studies. Injected dose is filled as white, because it spans a continuous range and is highly subject-specific. The 4D imaging data is the output of the data acquisition workflow and input to the preprocessing workflow.

Figure 3: Schematic overview of the various preprocessing steps used in analyzing dynamic $^{[1]}C$DASB data. This ranges from different motion correction techniques, co-registration, volume-
of-interest definitions, partial volume correction, and kinetic modeling. The colors indicate the percentage, in which a given step has been applied in the 105 $^{11}$C-DASB PET studies.
Timeline of number of patient and healthy controls in the 105 published $^{[1]}$C]DASB studies. The colors indicate either healthy controls, or a specific disorder as a function of time and sample size. Abbreviations; ADHD (Attention-Deficit/Hyperactive Disorder), MDD (Major Depressive Disorder), MDMA (ecstasy), HIV (Human Immunodeficiency Virus), OCD (Obsessive Compulsive Disorder), SAD (Seasonal Affective Disorder), PTSD (Post-Traumatic Stress Syndrome), PD (Parkinson’s Disease).
Schematic overview of the different data acquisition workflows used to acquire dynamic [11C]DASB data. The workflow consists of scanners providing anatomical information, i.e., MRI scanners at various field strengths (Tesla), various PET scanners, duration of the dynamic PET acquisition, frame sequence used to temporally acquire 4D [11C]DASB data, injected dose (ranging from approximately 185-740 MBq), and finally the reconstruction methods used to reconstruct the 4D PET sequence. The colors indicate the frequency per step that has been applied in a [11C]DASB PET study out of the total 105 studies. Injected dose is filled as white, because it spans a continuous range and is highly subject-specific. The 4D imaging data is the output of the data acquisition workflow and input to the preprocessing workflow.
Schematic overview of the various preprocessing steps used in analyzing dynamic [11C]DASB data. This ranges from different motion correction techniques, co-registration, volume-of-interest definitions, partial volume correction, and kinetic modeling. The colors indicate the percentage, in which a given step has been applied in the 105 [11C]DASB PET studies.