From THE DEPARTMENT OF CLINICAL NEUROSCIENCE

Karolinska Institutet, Stockholm, Sweden

POSITRON EMISSION TOMOGRAPHY: DEVELOPMENT, EVALUATION AND APPLICATION OF QUANTIFICATION METHODS

Pontus Plavén-Sigray

Stockholm 2018

Positron Emission Tomography: Development, evaluation and application of quantification methods

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Pontus Plavén-Sigray

Principal Supervisor:
Dr. Simon Cervenka
Karolinska Institutet
Department of Clinical Neuroscience

Co-supervisors:
Prof. Lars Farde
Karolinska Institutet
Department of Clinical Neuroscience

Prof. Petter Gustavsson
Karolinska Institutet
Department of Clinical Neuroscience

Dr. Martin Schain
Copenhagen University Hospital
Neurobiology Research Unit

Opponent:
Prof. Mark Slifstein
Stony Brook Neurosciences Institute
Department of Psychiatry

Examination Board:
Prof. Mark Lubberink
Uppsala University
Department of Surgical Sciences

Dr. Daniel Quintana
University of Oslo
Institute of Clinical Medicine

Dr. Andreas Frick
Uppsala University
Department of Psychology

Abstract

Positron Emission Tomography (PET) is an imaging technique that allows for *in vivo* quantification of biochemical and physiological processes in the brain. Examples of targets in the brain that can be imaged using PET are dopamine receptors and the translocator protein 18kDa (TSPO). Following intravenous injection of a radio-labeled ligand and the ensuing PET examination of a subject, kinetic models are often used to estimate parameters of interest. Example of such parameters are binding potential or distribution volume, which are estimates of the availability of receptors in a specific region or volume-element of the brain. These parameters can then be inserted into statistical models to infer e.g. differences in target availability between patients and controls or relationships to behavioral traits. In order to detect effects of interest, it is important that the estimation of these parameters is precise, reliable and valid. The aim of this thesis was to evaluate different methods for estimating such parameters, and apply them on clinical data.

The thesis consists of two different themes. The focus of theme I was the quantification of dopamine receptor in striatum and the cortex, and their relationship to normal and dysfunctional social behavior.

Study I and **Study II** examined the relationship between dopamine D1 receptor availability and self-rated pro and anti-social behavior in healthy subjects. **Study I** found a positive correlation between striatal D1 receptor availability and Social Desirability, and a negative correlation to Trait Aggression. **Study II** did however fail to replicate these results.

In **Study III**, dopamine D2 receptor availability in limbic and cortical regions in patients with social anxiety disorder and healthy controls were compared. Exploratory analyses suggested that patients had higher D2 receptor availability in the lateral and orbitofrontal cortex, although the results warrant replication in a larger sample.

The focus of theme II was the quantification of TSPO in patients with psychosis and healthy subjects. The level of TSPO in the brain has been hypothesized to function as an index of microglial cell activity, which in turn is believed to be a proxy for immune activation in the central nervous system.

In **Study IV**, [^{11}C]PBR28 binding in the whole grey-matter in patients with first-episode psychosis and healthy controls were compared. Contrary to the hypothesis of elevated microglia activity, patients were found to have lower TSPO levels.

Study V evaluated the test-retest reliability and convergent validity of different methods to measure TSPO levels using [^{11}C]PBR28. Distribution volume ratios and standardized uptakes value ratios, derived using pseudo-reference regions, showed both poor reliability and convergent validity.

Study VI carried out a meta-analysis of TSPO in patients with schizophrenia and psychotic disorders compared to healthy controls. Again, contrary to the hypothesis of higher microglia activity, strong evidence was found in favor of patients having lower TSPO levels in both cortical and subcortical regions.

In **Study VII**, the test-retest reliability and convergent validity of different methods to estimate TSPO levels using (R)-[^{11}C]PK11195 were evaluated. Outcomes derived using pseudo-reference region approaches were unreliable and showed no convergent validity to outcomes derived using arterial input function.

Finally, **Study VIII** evaluated the reliability and accuracy of a new modeling method, applied to [^{11}C]PBR28 data, in order to estimate specific binding without requiring a reference region. Simulations, a pharmacological challenge and test-retest analysis showed that non-displaceable distribution volume, and ensuing specific distribution volume values, derived using this method were accurate, precise and reliable.

Taken together, the results of the studies illustrate the importance of evaluating quantification methods prior to applying them on clinical data. The thesis also shows how robust kinetic and statistical modeling, and the use of direct replications or multi-center collaborations, can yield more trustworthy and reliable findings in PET.

List of publications

I **Plavén-Sigray P**, Gustavsson P, Farde L, Borg J, Stenkrona P, Nyberg L, Bäckman L, & Cervenka S (2014). Dopamine D1 receptor availability is related to social behavior: A positron emission tomography study. *NeuroImage*, *102*, 590-595.

II **Plavén-Sigray P**, Matheson GJ, Gustavsson P, Stenkrona P, Halldin C, Farde L, & Cervenka S (2018). Is dopamine D1 receptor availability related to social behavior? A positron emission tomography replication study. *PLoS One*, *13*, e0193770.

III **Plavén-Sigray P**, Hedman E, Victorsson P, Matheson GJ, Forsberg A, Djurfeldt D, Rück C, Halldin C, Lindefors N, & Cervenka S (2017). Extrastriatal dopamine D2-receptor availability in social anxiety disorder. *European Neuropsychopharmacology*, *27*, 462–469.

IV Collste K, **Plavén-Sigray P**, Fatouros-Bergman H, Victorsson P, Schain M, Forsberg A, Amini N, Aeinehband S, Karolinska Schizophrenia Project (KaSP) consortium, Erhardt S, Halldin C, Flyckt L, Farde L, & Cervenka S (2017). Lower levels of the glial cell marker TSPO in drug-naive first-episode psychosis patients as measured using PET and [^{11}C]PBR28. *Molecular Psychiatry*, *22*, 850-856.

V Matheson GJ, **Plavén-Sigray P**, Forsberg A, Varrone A, Farde L, & Cervenka S (2017). Assessment of simplified ratio-based approaches for quantification of PET [^{11}C]PBR28 data. *EJNMMI Research*, *7*.

VI **Plavén-Sigray P**, Matheson GJ, Collste K, Ashok AH, Coughlin JM, Howes OD, Mizrahi R, Pomper MG, Rusjan P, Veronese M, Wang Y, & Cervenka S (2018). Positron Emission Tomography Studies of the Glial Cell Marker Translocator Protein in Patients With Psychosis: A Meta-analysis Using Individual Participant Data. *Biological Psychiatry*, *In press*.

VII **Plavén-Sigray P**, Matheson GJ, Cselényi Z, Jučaite A, Farde L, & Cervenka S (2018). Test-retest reliability and convergent validity of (R)-[^{11}C]PK11195 outcome measures without arterial input function. *Manuscript on bioRxiv, 298992*.

VIII **Plavén-Sigray P**, Schain M, Zanderigo F, Karolinska [^{11}C]PBR28 study group, Rabiner I, Gunn R, Ogden T, & Cervenka S (2018). Accuracy and reliability of [^{11}C]PBR28 specific binding estimated without the use of a reference region. *Manuscript on bioRxiv, 345645*.

Additional publications not included in the thesis

Plavén-Sigray P*, Matheson GJ*, Schiffler BC*, & Thompson WH (2017). The readability of scientific texts is decreasing over time. *eLife*, *6*, *e27725*. *Authors contributed equally.

Matheson GJ, Stenkrona P, Cselényi Z, **Plavén-Sigray P**, Halldin C, & Farde L. (2017) Reliability of volumetric and surface-based normalisation and smoothing techniques for PET analysis of the cortex: A test-retest analysis using [^{11}C]SCH-23390. *Neuroimage*, *155*, 344–353.

Farde L, **Plavén-Sigray P**, Borg J, & Cervenka S (2018). Brain neuroreceptor density and personality traits: towards dimensional biomarkers for psychiatric disorders. *Phil Trans R Soc B*, *373*.

Thompson WH, Richter CG, **Plavén-Sigray P**, & Fransson P (2018). Simulations to benchmark time-varying connectivity methods for fMRI. *PLoS computational biology*, *14*, *e1006196*.

Stenkrona P, Matheson GJ, Cervenka S, **Plavén-Sigray P**, Halldin H, & Farde L (2018). [^{11}C]SCH23390 binding to the D1-dopamine receptor in the human brain - a comparison of manual and automated methods for image analysis. *EJNMMI Research*, *8*.

Matheson GJ, **Plavén-Sigray P**, Louzolo A, Borg J, Farde L, Petrovic P, & Cervenka S (2018). Dopamine D1 receptor availability is not associated with delusional ideation measures of psychosis proneness. *Manuscript on bioRxiv, 321646*.

Contents

Abbreviations

^{11}C	carbon-11
^{18}F	flour-18
2TCM	two-tissue compartment model
ABSS	arterial blood sampling system
AbsVar	absolute variability (or test-retest variability)
AIF	arterial input function
ANCOVA	analysis of covariance
AUC	area under the curve
BF	Bayes factor
BP_{ND}	binding potential non-displaceable
CER	cerebellum
D1-R	dopamine D1 receptor
D2-R	dopamine D2 receptor
D3-R	dopamine D3 receptor
DLPFC	dorsolateral prefrontal cortex
DVR	distribution volume ratio
FEP	first episode psychosis
FWHM	full width at half maximum
GM	whole-greymatter
HAB	high affinity binders
HR	ECAT Exact High Resolution system
HRRT	ECAT Exact High Resolution Research Tomograph system
ICC	intra-class correlation coefficient
IPD	individual participant data
KI	Karolinska Institutet
MAB	mixed affinity binders
MD	minimum detectable difference
MRI	magnetic resonance imaging

OFC	orbitofrontal cortex
PET	positron emission tomography
ROI	region of interest
SAD	social anxiety disorder
SEM	standard error of measurement
SNR	signal-to-noise ratio
SPM	statistical parametric mapping
SRTM	simplified reference tissue model
SSP	Swedish universities Scales of Personality
SUV	standardized uptake values
SUVR	standardized uptake value ratio
SVCA	supervised cluster analysis
TAC	time-activity curve
TSPO	translocator protein 18kDa
Vb	fractional blood volume
V_{ND}	non-displaceable distribution volume
V_S	specific distribution volume
V_T	total distribution volume
WB	whole-brain

Chapter 1

Introduction

Essentially, all models are wrong, but some are useful.

– George E.P. Box

1.1 Dopamine receptors

The human brain contains around 86 billion neurons, approximately making up 50% of the total cell count (*1*). After depolarization, some neurons can release transmitters from the terminal part of the axon into the synaptic cleft. These neurotransmitters bind to receptors that are located on the post-synaptic neuron. When the receptors are stimulated they evoke excitatory or inhibitory responses, either through direct (ionotropic receptors) or indirect (G-protein coupled receptor) gating of ion channels (*2*).

The monoamine neurotransmitter dopamine mainly binds to two families of receptors: dopamine D1-Receptors (D1-R) and dopamine D2-Receptors (D2-R). D1-R are coupled to a stimulatory G-protein structure that activates adenylyl cyclase enzyme, leading to a cascade of second messenger functions within the post-synaptic neuron. D2-R are coupled to a G-protein structure that inhibits the activation of adenylyl cyclase formation. Activation of D1-R and D2-R lead to changes in the membrane potential and the biochemical state of the post-synaptic neuron (*3*), propagating the signal further. There is however still much to be discovered about the exact mechanisms and effects of dopamine transmission.

Dopamine is produced inside cell bodies in the substantia nigra and the ventral tegmental area. Three major pathways in the primate brain arises from the axon connections of these cells: 1. the nigrostriatal pathway, projecting from substantia nigra to the dorsal striatum, 2. the mesolimbic pathway, projecting from the ventral tegmental area to the ventral (limbic) striatum and other limbic structures, and 3. the mesocortical pathway, connecting from the ventral tegmental area to cortical regions (*4*). Dopamine and dopamine receptors are involved in a wide array of behavioral phenotypes, such as movement (*5*), attention (*6*), motivation and learning (*7*), addiction (*8*), and many different psychiatric and neurological disorders (*9–11*).

1.2 Glial cells

The human brain contains a large number of glial cells, of which microglia and astrocytes are common subtypes (*12*). One important role of microglial cells is to respond to infections (*13*). As resident macrophages they exterminate cellular debris and are involved in the presentation of antigens and signaling of cytokines (*14*). They play an integral role in attacking invading cells, removing damaged cells and fragments, and by promoting neural tissue repair (*15*). Microglia activity is thought to be upregulated when there is an infection or inflammation in the central nervous system (*16*). However, healthy subjects, without any acute immune-activation, can also express microglia throughout the entire brain (*17*). Astrocytes function as support cells in the brain and constitute around 20% of glial cell population of the human brain (*18*). Among other things, they provide nutrients to neurons and are involved in repair of damaged tissue (*19*). Importantly, activation of glial cells could have both beneficial and detrimental effects on the central nervous system pathology (*20,21*).

1.3 Overview of positron emission tomography

Positron Emission Tomography (PET) is a functional imaging technique that allows for *in vivo* examination of neurochemical targets and physiological processes, such as receptors and transporters, neurotransmitter release, blood flow and glucose metabolism. PET imaging relies on the decay of isotopes with short half-life, such as Carbon-11 (^{11}C) or Flour-18 (^{18}F). By replacing an atom on a medical compound with a suitable isotope, a radioligand (or radiotracer) is constructed and can be

used as an imaging marker. Examples of such radioligands are [11C]SCH23990 which has high affinity for D1-R, [11C]FLB457 which has high affinity for D2-R, and [11C]PBR28 which binds to the Translocator Protein 18kDA (TSPO), found in microglia, astrocytes and white blood cells (*22–24*).

1.3.1   The PET examination

During a PET examination, the radioligand is normally administered to the subject by injection, and delivered to the brain via the blood stream. When the isotope attached to the ligand decays it releases a positron that travels approximately 1mm, before interacting with an electron. The ensuing collision annihilates both particles and two resulting photons are emitted in a random direction separated by approximately $180 \pm 0.25°$ (*25*). The photons travel through tissue, cerebral spinal fluid, skull and air before being detected by the PET system and registered as a coincidence count. During a single PET measurement, multiple coincidences are registered and form the measured projections from the 3D object positioned in the system. This is the raw data from which the image can be reconstructed.

The goal of PET reconstruction is to recreate to 3D object that gave rise to the measured projections. Simplified, a frame of a PET examination can be thought of as a set of 2D images (slices) that when stacked on top of each other forms the full 3D image. There are different ways to reconstruct these slices. For the Karolinska Institutet (KI) PET examinations included in this thesis, two procedures have been used: 1) filtered back projection (used for the ECAT Exact High Resolution system (HR), Siemens Molecular Imaging, Knoxville, TN, USA) and 2) ordered subset expectation maximization (used for the ECAT Exact High Resolution Research Tomograph (HRRT) system, Siemens Molecular Imaging, Knoxville, TN, USA).

A full PET measurement often takes between 50 to 120 min, most often depending on the isotope that is being used. During this time the brain uptake and washout of radioligand can be measured by obtaining the radioactivity concentration in short time intervals called "time frames". The length of each frame is determined by the preferred temporal resolution (often set to be high in the initial phase of the examination and low in the end) and the half-life of the isotope.
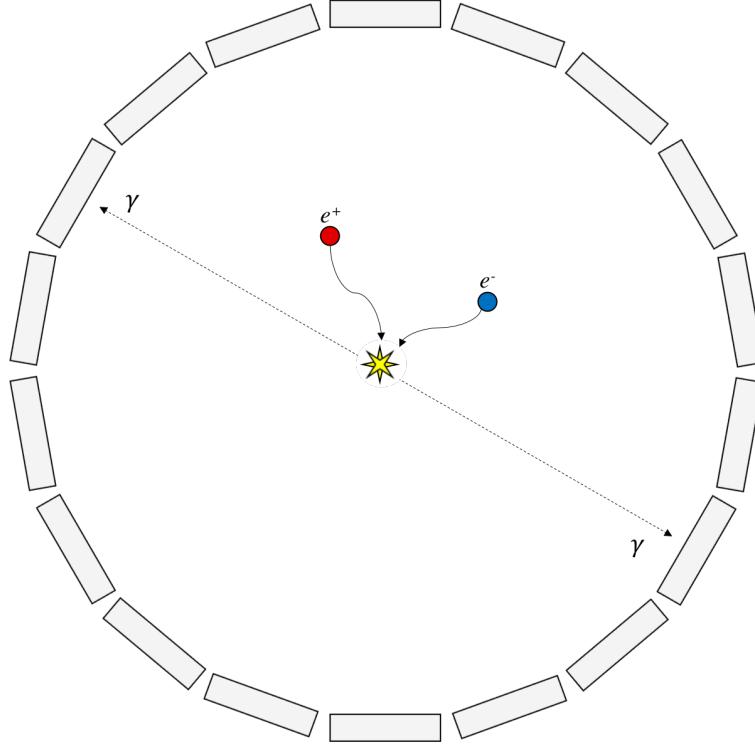
Figure 1.1: Schematic of a coincidence count in the PET system. When the isotope on the ligand complex decays, a positron (e^+) is emitted. The positron travels approximately 1mm before colliding with an electron (e^-) and both particles are annihilated, resulting in a photon pair. If both photons hit the detectors within a specified time-window, a coincidence is registered and a line-of-response can be established.

1.3.2 Compartmental models

In order to obtain the regional estimates of radioligand concentration, regions of interests (ROIs) are usually delineated on a magnetic resonance imaging (MRI) T1 weighted image. The T1 image is then co-registered to the PET image and regional concentration of radioactivity (measured in e.g. MBq/cc) of frames can be extracted. This results in ROI time-activity curves (TACs).

In order to quantify the metabolic or biochemical parameter of interest in the ROI, such as the density of a receptor, a kinetic model is used. The task of the kinetic model is to describe the TAC while differentiating the concentration of radioligand into various compartments. Depending on which ligand that is being used, different number of compartments and different descriptions of the transfer rates of concentration between compartments are necessary (*26*). A common and widely applicable kinetic model for describing the TAC and the concentration in
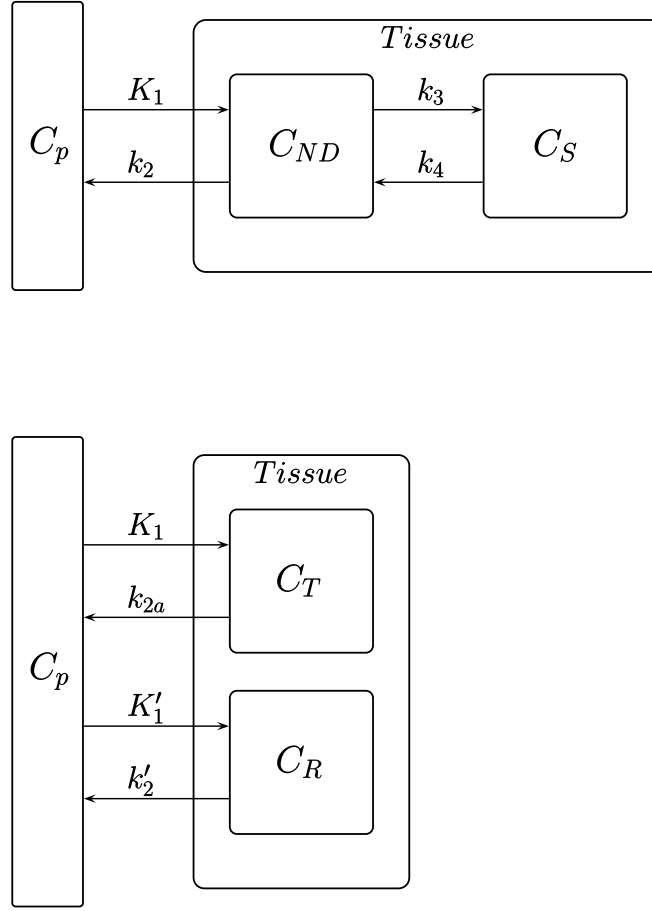
Figure 1.2: Upper panel: The two-tissue compartment model. Lower panel: The simplified reference tissue model. The boxes represent radioligand concentration in plasma (C_P), the non-displaceable compartment (C_{ND}), the specific compartment (C_S), the target region (C_T) and the reference region (C_R).

tissue is the two-tissue compartment model (2TCM; Figure 1A, (27)).

The rate constants (K_1, k_2, k_3, k_4 in Figure 1.2) describe the rate of transfer of radioligand to and from the different compartments. Using these rate constants, a set of differential equations can be set up to mathematically describe the exchange of radioligand between compartments:

$$\frac{dC_{ND}}{dt} = K_1 C_p(t) - k_2 C_{ND}(t) - k_3 C_{ND}(t) + k_4 C_S(t) \qquad (1.1)$$

$$\frac{dC_S}{dt} = k_3 C_{ND}(t) - k_4 C_S(t) \qquad (1.2)$$

where C_S is the radioligand concentration in the specific compartment (specifically bound to the target of interest) and C_{ND} is the radioligand concentration in the non-displaceable compartment (not specifically bound the the target of interest). In practice, however, the PET system cannot differentiate between C_S and C_{ND}. Instead the sum of the two becomes the total concentration of radioligand in a ROI:

$$C_T(t) = C_{ND}(t) + C_S(t) \tag{1.3}$$

The solution of equation 1.1 and equation 1.2 leads to a formula for describing the total concentration (C_T) of radioligand in a ROI:

$$
\begin{aligned}
C_T(t) = C_p(t) &\otimes (ae^{-ct} + be^{-dt}) \\
with \quad a &= r(k_3 + k_4 - c) \\
b &= r(d - k_3 - k_4) \\
c &= (s - e)/2 \\
d &= (s + e)/2 \\
e &= \sqrt{s^2 - q} \\
q &= 4k_2k_4 \\
r &= K_1/e \\
s &= k_2 + k_3 + k_4,
\end{aligned}
\tag{1.4}
$$

where $C_p(t)$ represents the concentration of unchanged radioligand in arterial plasma and is obtained by sampling arterial blood during the PET measurement. As such, $C_p(t)$ is also known as the arterial input function (AIF) and the exponential to the right of the convolution in equation 1.4 is known as the impulse response function.

The ROI will also contain a substantial amount of intravascular radioactivity, called whole-blood concentration (C_a) in the beginning of the PET measurement that is not accounted for by the model described in equation 1.4. The fractional blood volume (V_b) can therefore be added to the model:

$$C_{measured}(t) = (1 - V_b)C_T(t) + V_bC_a(t) \tag{1.5}$$

where $C_{measured}$ is the measured radioactivity obtained by the PET system during an examination. The rate constants are obtained by fitting equation 1.4 to a TAC, which has been corrected for V_b. Once obtained, the rate constant can be used to

derive parameters of interest, such as the total distribution volume (V_T). For the 2TCM, V_T is calculated as:

$$V_T = \frac{K_1}{k_2}\left(1 + \frac{k_3}{k_4}\right)$$ (1.6)

V_T is defined as the total concentration of radioligand in the ROI over the total concentration in plasma. A V_T of e.g. 2 means that 2ml of plasma is needed to account for the radioactivity in 1ml of tissue. As such V_T is an index of radioligand binding to target. This outcome can be used in inferential statistical designs, e.g. in order to quantify differences in target availability between patient and control groups.

1.3.3   Simplified compartmental models

Many different expansions and simplifications have been derived from the 2TCM. Consider for example a region of the brain with negligible specific binding of radioligand. If such a region exists the input function can be described using only rate constants and the concentration of radioligand in this region, a.k.a the non-displaceable compartment or "reference region" (Figure 1.2). This allows for quantification of non-displaceable binding potential (BP_{ND}) without the need for arterial blood sampling to establish an AIF (*27*). BP_{ND} is a measure of specific radioligand binding. It is defined as the concentration in the specific compartment over the concentration in the non-displaceable compartment. BP_{ND} is proportional to availability of the target:

$$BP_{ND} = \frac{B_{avail}}{K_d} f_{ND}$$ (1.7)

where B_{avail} is the density of the available binding sites, $1/K_d$ is the affinity of the radioligand and f_{ND} is the fraction of free radioligand in the non-displaceable compartment. Assuming that the affinity and free fraction of radioligand is similar between the subjects or the groups that are being compared, BP_{ND} can be thought of as an index of the density of the available target sites.

If a radioligand has very fast kinetics between the non-displaceable and the specific compartments, the model can be simplified by collapsing these two compartments into one. One such model is the Simplified Reference Tissue Model (SRTM) (*28*):

$$C_T(t) = R_1 C_R + \left\{ k_2 - \frac{R_1 k_2}{1 + BP_{ND}} \right\} C_R(t) \otimes e^{-\frac{k_2 t}{1 + BP_{ND}}} \qquad (1.8)$$

The number of parameters have now been reduced, leaving only R_1 (the ratio of K_1 over K_1'), k_2 and BP_{ND} to be fitted. The SRTM often has the advantage of showing less variability (higher reliability), although this can come with a cost of increased bias (lower accuracy) of the outcome measure BP_{ND} (*29*).

1.3.4 Linear models

Fitting the models described in equation 1.4 and 1.8 requires non-linear optimization. This can be a time-consuming process if there is a substantial amount of TACs that have to be modeled. An example of this is when a BP_{ND} value has to be derived for each separate volume element (voxel) in the brain. For this reason, full-tissue compartment models and simplified tissue compartmental models have been linearized, allowing for substantially faster optimization. There are also "compartmental free" linear models, that make less assumptions about the radioligand kinetics in different tissue types in the brain. Examples of linear models are Multilinear Analysis 1 and 2 (*30*), Multilinear Reference Tissue Model (*31*) and the Logan Plot (*32*).

1.3.5 All models are wrong

All kinetic models described above are only approximations of real biological processes. Such models might be useful in order to estimate a biological entity of interest but are not guaranteed to yield precise or even meaningful outcomes. The fact that one model has worked well to describe the availability of target for one radioligand under certain conditions, does not mean that the same model will be useful for other radioligands. All outcomes will also have more or less uncertainty attached to them, either due to biological variability, measurement error or because of faulty model assumptions. Hence, the e.g. point-estimated V_T or BP_{ND} value of a subject does not represent the absolute "true" binding, availability or density of target, but rather only a likely value among a set of many likely values. This inherent uncertainty around PET outcomes measures are rarely acknowledged in literature (or in this thesis), although there exist methods to account for it (*33*).

1.4   Imaging dopamine receptors in social behavior

1.4.1   Interpersonal personality traits

In personality literature, social behavior is commonly classified into two orthogonal dimensions, 1. agency (or dominance) and 2. affiliation (or agreeableness). This means that a person can display social behavior that are combinations of these two traits, such a high agency and low agreeableness (i.e. being aggressive and hostile) or high agency and high affiliation (i.e. being extroverted and gregarious) (*34*). The interpersonal traits described by this model are thought to be both descriptive and predictive of psychiatric disorders (*35,36*).

1.4.2   D2-R and Social Desirability

Previous PET studies have investigated the relationship between the pro-social personality traits Social Desirability and D2-R availability in striatal (*37–41*) and extrastriatal regions (*40*). In the two-dimensional theory mentioned above, measurements of Social Desirability are positioned as a combination of low dominance and high agreeableness (*40*). The trait reflects how a person acts in order to gain approval or acceptance by others (*42*). Most previous studies found a negative relationship between Social Desirability and D2-R binding in the striatum (*37–40*), with one study reporting a null-finding (*41*). Importantly, these previous studies found significant associations in different parts of the striatum. When using the whole-striatum as ROI for all studies, and including unpublished personality data from a samples of twins examined with [^{11}C]raclopride at the KI PET center, a random effects meta-analysis suggest that there is a modest negative correlation between D2-R availability and Social Desirability (Figure 1.3). One caveat to this analysis is that the funnel-plot (not shown) reveals potential publication bias. Hence, if the true effect is not null, it is likely to be smaller than presented in Figure 1.3.

1.4.3   D1-R, Social Desirability and aggression

The relationship between pro- and anti-social personality traits and the D1-R has not been as thoroughly investigated as for the D2-R. One study in mice suggest that D1-R expression is associated with anti-social behavior and may have an opposite role to D2-R in mediating aggression (*43*). In humans, measurements of Social

20

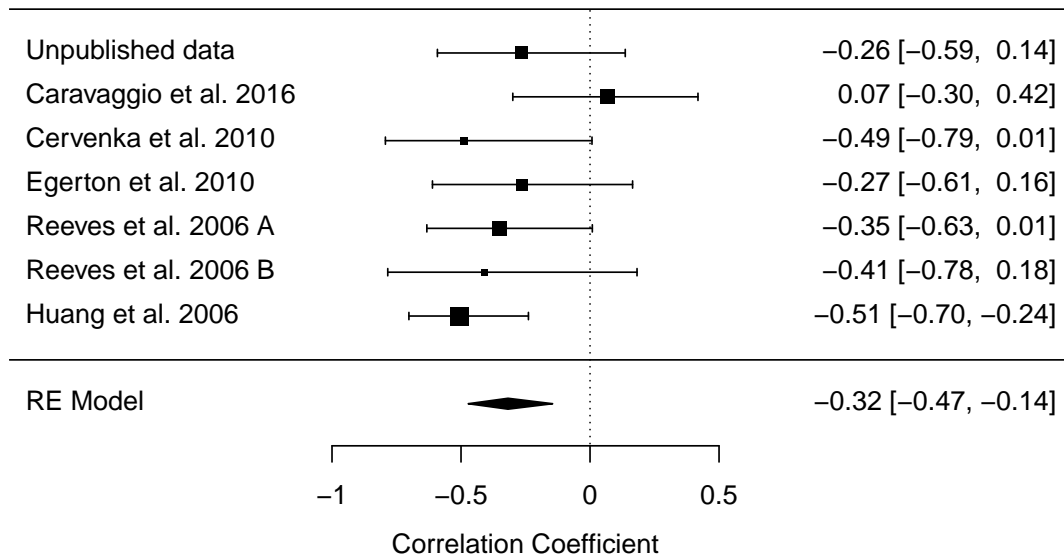|		
Unpublished data		−0.26 [−0.59, 0.14]
Caravaggio et al. 2016		0.07 [−0.30, 0.42]
Cervenka et al. 2010		−0.49 [−0.79, 0.01]
Egerton et al. 2010		−0.27 [−0.61, 0.16]
Reeves et al. 2006 A		−0.35 [−0.63, 0.01]
Reeves et al. 2006 B		−0.41 [−0.78, 0.18]
Huang et al. 2006		−0.51 [−0.70, −0.24]
RE Model		−0.32 [−0.47, −0.14]

Correlation Coefficient

Figure 1.3: Meta-analysis of correlations (with 95% CIs) between self-rated reports of Social Desirability and D2-R availability in the whole-striatum. Social Desirability was measured by self-reported personality questionnaires and striatal D2-R availability was measured using [^{11}C]raclopride or [^{123}I]iodo-benzoaminde.

Desirability have consistently been found to have high but negative loadings with aggression scales in exploratory component analyses (44). This suggests that the Social Desirability and aggression scales tap a common underlying pro-social (or anti-social) dimension, although in opposite direction of each other. In the mice study (43), increased levels of D1-R were related to higher levels of pro-social behavior, while the opposite pattern was found for aggression. If this relationship was also to be found in humans, it would suggest opposite regulatory mechanisms for the D1-R and D2-R systems in mediating pro- and antisocial behavior. This in turn could have implications for diagnosing and treating psychiatric conditions associated with dysfunctional social behavior, such as antisocial personality disorder or social anxiety.

1.4.4   D2-R and social anxiety disorder

Social anxiety disorder (SAD), also known as social phobia, is one of the most common anxiety disorders with a 12 months prevalence of around 18% (*45*). SAD is characterized by fear and avoidance of social interaction and scrutiny by other people (*46,47*). The dopamine system is involved in both fear processing (*48*) and avoidance learning (*49*). Together with the tentative role for D2-R in Social Desirability, these results suggest that an altered D2-R system may be involved in the pathophysiology of SAD.

Single-photon emission tomography and PET studies on the D2-R in SAD have however produced conflicting results. There have been reports of both lower levels (*50*) and no change (*51*) in D2-R availability in patients as compared to healthy controls. Importantly, these studies have focused on the striatum, while blood flow and activation studies have mainly shown that extrastriatal regions are involved in the pathophysiology of SAD. Specifically, many functional MRI studies have reported altered activation in voxels within the amygdala, anterior cingulate, insula, orbitofrontal (OFC) and medial frontal cortex in the disorder (*52,53*). Prior to this thesis, there have been no studies investigating extrastriatal D2-R availability in SAD patients compared to healthy controls.

1.5   Imaging TSPO in psychosis

1.5.1   Schizophrenia

Schizophrenia is a heterogeneous neuropsychiatric disorder characterized by delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior or negative symptoms, such as lack of motivation or poverty of speech. In order for people to be diagnosed with schizophrenia, they should display at least two of these symptoms for a minimum period of six month (or less if successfully treated) (*54*). The lifetime prevalence of schizophrenia is about 0.5% (*55*) and the disorder is associated with significant impairment of social and occupational life.

1.5.2 Immune hypothesis of schizophrenia

An altered or aberrant immune system has been hypothesized as one of underlying reasons for the genesis or progression of schizophrenia. This hypothesis gains support many different fields of research: 1) studies have found that genetic loci associated with schizophrenia were expressed in tissues critical for immune function (*56*), 2) epidemiological studies have shown that pre- and postnatal infection is associated with later onset of the disorder (*57,58*), 3) pharmacological studies have shown that Cox-inhibitors have positive add-on effects to anti-psychotics (*59*) and 4) studies have shown aberrant expression of immune markers in patients, such as up-regulated cytokine levels in plasma in chronic (*60*) and in drug-naive (*61*) patients. In the brain, cytokines are mainly released by microglia and astrocytes. Due to these increased levels of cytokines, it has been hypothesized that microglia and astrocytes will also show up-regulated numbers or activity in patients with schizophrenia (*62,63*).

Initial post-mortem studies have reported increases in microglial cells, and in some cases also astrocytes, in brains of schizophrenic patients (*64–67*). However, later large-scale meta-analyses on post-mortem brains have revealed mixed results regarding the expression of these cells in psychotic disorders, showing substantial heterogeneity in the direction of patient-control differences (*68,69*). Changes in cell expression at the time of death, methodological differences in labeling and counting cells, as well as the use of different immunohistochemical markers could in part explain some of the variation in results. Common limitations in these *in vitro* studies were that control subjects and patients often were old at the time of death, and the cause of death was often suicide (*70*).

1.5.3 Translocator protein 18kDa

TSPO is located on the outer mitochondrial membrane of many different cell types in the mammalian body, such as microglia and astrocyte brain cells. It has been hypothesized that when there is an on-going inflammation in the central nervous system, microglia become activated and the levels of brain TSPO is increased (*71,72*). If true, this would make the TSPO a suitable *in-vivo* biomarker for neuroinflammation or brain immune-activation. It is however important to note that upregulated levels of TSPO might reflect pathological processes other than immune activation, such as discrete neurotoxic damages or abnormalities in cell metabolism, energy production and oxidative stress (*73*). In fact, even healthy individuals, without an activated

immune system, might also show high levels of TSPO in e.g. dormant microglial cells (*17*).

1.5.4 First generation TSPO radioligands

The (R)-[^{11}C]PK11195 PET radioligand has been used since the mid 90s with the aim of imaging TSPO expression in the human body and brain. (R)-[^{11}C]PK11195 has been applied in a range of studies aiming to investigate neuroinflammation in a wide set of disorders, such as Alzheimer's disease (*74*), multiple sclerosis (*75*), major depression (*76*) and schizophrenia (*77–81*).

1.5.5 Second generation TSPO radioligands

Concerns regarding the low signal to noise ratio (*82–84*), and poor reliability (*85*) of (R)-[^{11}C]PK11195, sparked the development of a set of second-generation TSPO radioligands. These included tracers such as [^{11}C]PBR28 and [^{11}C]DPA713, which show greatly improved specific binding (*86*) as well as higher reliability (*87*), compared to (R)-[^{11}C]PK11195. However, the binding affinity of all second-generation radioligand are affected by a genetic polymorphism, the rs6971-allele, resulting in three binding-types: high affinity binders (HABs), mixed affinity binders (MABs) and low affinity binders (LABs) (*88,89*). E.g. HABs has double the affinity for [^{11}C]PBR28 radioligand binding to TSPO, which result in higher binding estimates compared to MABs, even if the true density of TSPO is the same. The distribution of HABs, MABs and LABs in the Caucasian population is approximately 49%, 42% and 9% respectively, while the presence of the low affinity allele is generally lower in the Afro-American, Han Chinese or Japanese populations (*90*). These differences in affinity makes it important to screen and control for the TSPO genotype when performing PET studies using the second-generation TSPO radioligands.

1.5.6 Arterial input function or pseudo-reference region approaches

TSPO is expressed throughout the entire brain, and for this reason no brain region can serve as reference in a kinetic model. Instead, arterial samples of metabolite-corrected plasma have to be obtained and used as AIF in a kinetic model. V_T, which is one of

the the most widely applied outcome in PET, is often derived from a kinetic model using such an input function. Obtaining an AIF is difficult as it can be uncomfortable for the research subjects, require specialized research equipment and skills, laborious for the research personnel, and prone to measurement error. For these reasons, there have been many suggestions of alternative methods to quantify the binding of first and second generation TSPO radioligands in the brain, that do not make use of an AIF. One of the most common and simple approaches is to use standardized uptake values (SUVs). A SUV is calculated by dividing the radioactivity concentration in the TAC (c_{TAC}) with the injected radioactivity (c_{inj}) and the body weight (BW) of the subject for all time points:

$$SUV(t) = \frac{c_{TAC}(t)}{c_{inj}/BW} \tag{1.9}$$

The average, or the area under the curve (AUC), of the SUV over a time interval is then often taken as the final outcome measure.

Another approach to quantify radioligand binding in the abscence of an AIF or reference region is to use a "pseudo-reference region" (*91,92*). A pseudo-reference region is simply a region that contains specific binding, but the binding is assumed to be either low, or show negligible differences between the subjects or the groups that are being compared.

One way to use such a pseudo-reference region is to divide the SUV in a target ROI with the SUV in the pseudo-reference region. For [11C]PBR28, this approach has been suggested to produce an outcome, standardized uptake value ratios (SUVRs), with higher sensitivity to detect clinical effects (*91,92*). Another way is to use the pesudo-reference region in a kinetic model, such as the SRTM, and derive "pseudo-BP_{ND}" values. This has been suggested to be suitable method for estimating e.g. (R)-[11C]PK11195 binding (*79,81*).

Even if V_T values have been obtained, the use of this approach (e.g. dividing a V_T value in a target ROI with that in a pseudo-reference region) has been suggested to be a suitable variance-reducing strategy allowing for more precise comparison of subjects or groups (*93*). However, such suggestions have not been without criticism (*94*). One important question is how much relevant biological signal remains in the target ROI after the division with a pseudo-reference region.

1.5.7 TSPO imaging in psychosis

The first PET studies to examine TSPO expression in schizophrenia were published over a decade ago and used the first-generation TSPO radioligand (R)-[^{11}C]PK11195. In a sample of 10 patients with recent onset schizophrenia and 10 control subjects (*77*), the authors reported higher levels of whole grey-matter TSPO binding in recent-onset schizophrenia (p=0.048), using specific distribution volume (V_S) as outcome measure. This was followed by a study with 7 patients with psychotic disorder and 8 healthy controls where patient (R)-[^{11}C]PK11195 BP_{ND} (k_3/k_4) was shown to be higher in the hippocampus after normalizing to whole grey-matter BP_{ND} (p$_{\text{corrected}}$=0.04) (*78*). Later studies, which have not made used of AIFs, have reported mixed findings, with patients showing both higher and no-differences in pseudo-BP_{ND} values (*79–81*). Since no reference region exists for TSPO, these pseudo-BP_{ND} values were derived using SRTM with TACs from the supervised-cluster analysis (SVCA) (*95,96*), or the cerebellum, as reference input.

Table 1.1 shows all published (R)-[^{11}C]PK11195 studies in schizophrenia or psychosis using the whole of grey-matter as ROI. In these studies, different outcome measures have been used, and it's not clear whether or not the effect sizes can be entered into the same meta-analysis model: First, the reliability and precision of the outcomes measures have not yet been established. Second, the outcome measures likely show different sensitivity and precision for patient-control differences.

Table 1.1: (R)-[^{11}C]PK11195 patient-control comparisons of schizophrenia or psychotic disorder using the whole grey-matter as ROI. Pseudo-BP_{ND} values were derived using either a cerebellum (CER) TAC or a TAC from the supervised cluster analysis (SVCA) as reference input.

Study and subsample[a]	N controls/patients[b]	Used AIF	Outcome	Patient Mean (SD)	Control Mean (SD)	reported p	Hedges' g
van Berckel et al., 2008							
	10/10	yes	V_S (BP_P)	1.89 (0.32)	1.62 (0.3)	<0.05	0.83
Doorduin et al., 2009							
	8/7	yes	BP_{ND}	1.99 (0.64)	1.54 (0.41)	0.122	0.80
Holmes et al., 2016							
Medicated	16/8	no	pseudo–BP_{ND} CER	0.16 (0.11)	0.09 (0.05)	0.032	0.95
Drug-free	16/8	no	pseudo–BP_{ND} CER	0.09 (0.07)	0.09 (0.05)	0.981	0.08
van der Doef et al., 2016							
	17/19	no	pseudo–BP_{ND} SVCA	0.17 (0.09)	0.14 (0.09)	0.3	0.33
di Biase et al., 2017							
Chronic	15/18	no	pseudo–BP_{ND} CER	0.9 (0.04)	0.9 (0.04)	NA	0.12
Rescent onset	12/15	no	pseudo–BP_{ND} CER	0.87 (0.05)	0.89 (0.04)	NA	-0.29

[a] di Biase et al., 2017 (81) did not examine whole grey-matter. Instead reported ROI pseudo-BP_{ND} values have been averaged.
[b] The same healthy control sample was used for both patient samples in Holmes et al., 2016 (79)

To date, five studies have been published on differences between psychosis patients and healthy controls using second-generation TSPO radioligands, excluding Study IV from this thesis (*93,97–100*). Four of these studies did not find support for the hypothesis of higher levels of TSPO (estimated using V_T) in the brain of patients compared to control subjects (*93,97–99*). One study (*100*) did however find a significant interaction effect of age and patient-control groups on V_T (with older patients showing higher V_T values in grey matter). This finding emerged after entering all subjects as random effects into a linear mixed effects model and covarying for TSPO genotype, age, cohort by age, cohort by genotype and age by genotype.

1.6 Statistical inference of PET outcome measures

Two schools of inference have been applied in this thesis: frequentist and Bayesian statistics. In frequentist inference the aim is to calculate the probability of obtaining the observed, or more extreme data, given that the null-hypothesis is true. Broadly speaking, frequentist statistics can be divided into two groups, 1) Fisherian inference (*101,102*) and Neyman-Pearson Null-Hypothesis-Testing (*103*). In Fisherian statistics, the p-value is meant to capture surprise. I.e. assuming that the null-hypothesis is true, the researchers obtain an estimate on how surprised they should be by the observed data. If the data is deemed to be very surprising (i.e. the p-value is very low), this is taken as indirect evidence against the null-hypothesis. In a Neyman-Pearson framework, the p-value allows researchers to perform error-control, conditioned on the null-hypothesis being true. I.e. if the p-value falls below a pre-defined threshold (such as 0.05), then the researcher knows that if they act like the null is false they will not be wrong more than 5% of the time.

In Bayesian inference, statistical parameters, such as regression coefficients, are represented by probability distributions. A probability distribution reflects both the researcher's belief in, and the uncertainty around, different values for that parameter. The observed data is then used to update these distributions, to arrive at posterior distributions. A posterior distribution represents the researcher's updated belief about the probability, and uncertainty, of different values of the parameter (*104,105*).

1.6.1 All models are wrong, again

Contrary to semi-popular belief, there exist no statistical model or inference that is inherently "objective". I.e. there are no statistical models were data only speaks for itself and results are independent of the choices and beliefs of the researchers. All models, frequentist and Bayesian, involves making a wide array of subjective choices and assumptions, such as the use of likelihood function, regularizing coefficient, hierarchical or non-hierarchical treatment of data, alpha thresholds, null-intervals of interest, division of training and testing data sets, fit-metrics and procedures for model-averaging, and prior distributions. Using a pre-packaged statistical model, such as an ANOVA, might give the illusion of producing objective results, but applying it "off the shelf" merely entails that somebody else made subjective choices, often hidden in the model.

1.6.2 Pre-registration, reproducibility and replication

There is a growing awareness in neuroscience that pre-registration, reproducibility and direct replication of studies are important for improving the trustworthiness and robustness of findings (*106–108*). Pre-registration is when researchers publicly pre-register the study design, data and/or statistical analysis procedures. I.e. they declare what they are going to do before they do it. Ideally, pre-registration should be done prior to collecting any data. This helps in combating three major issues in the scientific literature: 1) publication bias (also known as the "file drawer problem"), 2) creating hypotheses and presenting results as confirmatory only after the results are known (also known as "Hypothesizing After the Results are Known" (*109*)), and 3) analytical flexibility, such as switching outcomes and covariates or dropping outliers in order to obtain positive findings (also known as "p-hacking").

If a study or analysis is reproducible, it means that it possible for the results to be reproduced and recreated by an independent researcher. This includes practices of open data sharing and public sharing of analysis code and study procedures (*108*), which in turn makes post-publication peer review by the scientific community possible.

A direct replication is when an identical (or very similar) study design, data and/or statistical analysis procedures are applied on a new and independent data set, with the aim of replicating the results of a previous published study (*110*). This helps in assessing the robustness of published findings and combating p-hacking as well as

other questionable research practices (*106,111,112*). In PET literature there is to date little practice in pre-registering the study or statistical analysis protocol, make analyses reproducible or initialize and publish direct replication studies.

Aims

The thesis consists of two themes:

 I. Examine the relationship between dopamine receptors and normal as well as dysfunctional social behavior (Study I-III).

 II. Apply and evaluate methods for examining brain levels of TSPO in psychosis (Study IV-VIII).

The first aim of theme I was to examine the relationship between D1-R and pro and anti-social behavior, measured using scales from self-rated personality questionnaires. The second aim was to examine differences in extrastriatal D2-R levels between patients with SAD and healthy control subjects.

The aim of theme II was to examine brain TSPO levels, hypothesized to be an index of glial cell activity, in patients with psychosis or schizophrenia as compared to healthy control subjects. Since there exist many different approaches to quantify TSPO binding, a further aim was to examine the reliability, accuracy and validity of such approaches. This was done in order to both aid the interpretation of already published clinical findings, and to evaluate which quantification method would be suitable to apply in future *in vivo* studies of TSPO in humans.

Chapter 2

Methods

2.1 Research participants

This thesis includes subjects that participated in different PET projects at KI (Study I-VIII), as well as subjects examined at PET centers in the UK, USA and Canada (sub-samples included in Study VI and VIII). All subjects included in the thesis gave written informed consent according to the Helsinki declaration prior to participating. All studies performed at KI were approved by the Stockholm regional ethics board and the Karolinska Hospital radiation safety committee. Table 2.1 outlines the demographic data for all studies and participants in this thesis.

Study I and Study II The subjects included in Study I (n = 23) and II (n = 26) were healthy controls examined with [^{11}C]SCH23390. In addition to the PET examinations, all participants were given personality questionnaires to fill out. A subset of subjects included in Study II underwent two PET examinations as part of a test-retest study (n = 16) that has been published previously (*113*), but only data from the first examination was included in this thesis.

Study III Twenty-eight subjects (16 healthy controls and 12 patients with SAD) were examined using [^{11}C]FLB457. The patients were medication-free at the time of PET. Eleven patients were examined prior to participating in a study about internet based cognitive behavioral therapy for SAD (*114*).

Study IV Sixteen drug-naive first episode psychosis patients and 16 control subjects were examined with [^{11}C]PBR28. Patients were recruited from hospital wards and outpatient clinics around the Stockholm region. Mean duration of illness was 7.9 ±

9.6SD months. None of the patients had been exposed to anti-psychotic medication and in order to ease participation, anxiolytics and sedatives were allowed during the course of the study. Subsequently, 5 patients received occasional benzodiazepines. Fourteen of the controls also participated in a [11C]PBR28 test-retest study (partly overlapping with the sample used for Study V) and two controls were obtained from different ongoing [11C]PBR28 study where they also participated as control subjects.

Study V Data from an already published [11C]PBR28 test-retest study (*87*) was reanalyzed. Fourteen subjects were recruited to the original study, but for two subjects data from the first PET examinations had to be discarded due to technical reasons. The remaining 12 (6 HABs and 6 MABs) participated in two PET examinations taking place on either the same day (n = 6) or 2-5 days apart (n = 6).

Study VI This study was an individual participant data (IPD) meta-analysis of all published studies up to 2017 that examined TSPO binding in schizophrenia or psychosis using second-generation radioligands. All participants from Study IV were included, as well as an additional 28 subjects (14 patients, and 14 controls) (*93*), 30 subjects (14 patients and 16 controls) (*115*), and 68 unique subjects (35 patients and 33 controls) (*97,99*).

Study VII Data from an already published (R)-[11C]PK11195 test-retest study (*85*) was reanalyzed. Six healthy controls were examined with the first generation TSPO radioligand (R)-[11C]PK11195 and underwent two PET examinations with approximately 6 weeks in between.

Study VIII Different samples of healthy subjects examined with [11C]PBR28 were included: A) 11 out of 12 subjects from the test-retest used in Study V (excluding one subjects who's second PET was only 60 minute long), B) 54 subjects from the KI [11C]PBR28 database and C) 5 subjects that participated in a previously published pharmacological competition challenge with XBD173 at IMANOVA Ltd London, UK (*116*).

2.2   Personality assessment

In Study I and II, participants were given the Swedish universities Scale of Personality (SSP) (*117*) to fill out. The SSP is a self-rating questionnaire that measures 13 different personality traits using scales consisting of 8 items each. Respondents rate how much they agree with statements about their thoughts, behavior and emotions

Table 2.1: Demographic data for all samples included in the thesis. f = females; m = males

Study and subsample	Radioligand	Target	N total	N controls f/m	N patients f/m	Age controls Mean (SD)	Age patients Mean (SD)
Study I							
-	[¹¹C]SCH22390	D1R	23	10/13	-/-	39.2 (21.9)	-
Study II							
-	[¹¹C]SCH22390	D1R	26	0/26	-/-	26.2 (3.2)	-
Study III							
-	[¹¹C]FLB457	D2/D3R	28	7/9	7/5	37.8 (15.2)	33.8 (11.6)
Study IV							
-	[¹¹C]PBR28	TSPO	32	9/7	5/11	26.4 (8.4)	28.5 (8.4)
Study V							
-	[¹¹C]PBR28	TSPO	12	6/6	-/-	23.9 (2.9)	-
Study VI							
All	-	TSPO	152	35/42	24/51	35.4 (15.1)	33.9 (12.6)
Bloomfied et al., 2016	[¹¹C]PBR28	TSPO	26	3/11	3/9	46.2 (13.6)	47 (9.3)
Collste et al., 2017	[¹¹C]PBR28	TSPO	32	7/9	8/8	26.4 (8.4)	28.5 (8.4)
Coughlin et al., 2016	[11C]DPA713	TSPO	26	5/9	3/9	25.4 (4.9)	24.3 (3.3)
Hafizi et al., 2017	[¹⁸F]FEPPA	TSPO	37	10/8	7/12	27.3 (9.1)	27.5 (6.8)
Kenk et al., 2015	[¹⁸F]FEPPA	TSPO	31	8/7	6/10	54.3 (9.5)	42.5 (14)
Study VII							
-	(R)-[¹¹C]PK11195	TSPO	6	0/6	-/-	25.8 (3.9)	-
Study VIII							
KI [11C]PBR28 database	[¹¹C]PBR28	TSPO	54	22/32	-/-	45.2 (17)	-
Test-Retest	[¹¹C]PBR28	TSPO	11	5/6	-/-	24.5 (3)	-
XBD173 competition	[¹¹C]PBR28	TSPO	5	0/5	-/-	25.2 (7.3)	-

over time and different contexts. Subjects respond by choosing one alternative on a 4-point Likert-scale ranging from "Strongly agree" to "Strongly disagree". The measured personality traits are assumed to be semi-stable over time and different contexts. They are thought to predict how vulnerable a person is for developing a psychiatric condition (*118,119*). The SSP is an updated version of the Karolinska Scale of Personality (*120*). It shows improved psychometric properties compared to its predecessor, such as higher internal reliability and a more coherent factor structure (*117*).

In Study I, three different SSP scales were considered: Social Desirability, Physical Trait Aggression and Verbal Trait Aggression. The Social Desirability scale assesses how a person represents herself in a social situation in order to gain approval and acceptance by others. The Physical Trait Aggression and Verbal Trait Aggression scales measures the tendencies of a person to react with physical violence or intimidation when provoked or unfairly treated. Study II aimed to replicate the findings from Study I, and only the Social Desirability and Physical Trait Aggression scales were analyzed.

2.3 Radioligands

In Study I and II, [11C]SCH23390 (*22*) was used. [11C]SCH23390 has high affinity for the D1-R (*121*).

In Study III, [11C]FLB457 was used. [11C]FLB457 is an analogue of epidepride and has very high affinity for D2 and D3 receptors (D2/D3-R), making it a suitable radioligand for examining binding outside the striatum where the density of D2-R is lower (*122*).

In Study IV, V and VIII, the second-generation TSPO radioligand [11C]PBR28 (*123*) was used and in Study VII the first-generation TSPO radioligand (R)-[11C]PK11195 (*124*) was used. Both bind to the TSPO receptor (also known as the peripheral benzodiazepine receptor), although [11C]PBR28 has tenfold the affinity for TSPO compared to (R)-[11C]PK11195 (*123,125*).

Study VI included participants examined with three different second-generation TSPO radioligands: [11C]PBR28, [11C]DPA713 (*126*) and [18F]FEPPA (*84*).

2.4 PET image acquisition

All participants included in studies performed at KI were examined with either the ECAT Exact HR (Study I-II and VII) or the ECAT Exact HRRT (Study III-VI and VIII) PET system at the KI PET center, located at the Karolinska University Hospital in Solna, Stockholm. The HR system provides a transaxial field of view of ~4 mm full width at half maximum (FWHM) (*127*). HR data was reconstructed using filtered back projection with a Hann filter of 2mm cut off frequency. The HRRT system provides transaxial field of view of ~2 mm FWHM (*128*). HRRT data was reconstructed using 3D ordinary Poisson ordered subset expectation maximization (10 subsets, 16 iterations), including modeling of the system's point spread function (*129*).

For all examinations, the radioligand was delivered as a rapid bolus, with an injection time that ranged between 5-10 seconds. Transmission scans were performed prior to all emission scans to correct for attenuation. To minimize movement during the KI PET examinations, a plaster helmet was made for each participant and used to fixate the head during the scan (*130*). In Study I and II the total emission scan was 51 minutes ([11C]SCH23390), in Study III 87 minutes ([11C]FLB457), in Study IV

and V 90 minutes ([^{11}C]PBR28), in Study VII 60 minutes ((R)-[^{11}C]PK11195), while in study VIII 75 minutes ([^{11}C]PBR28) of the examinations were used. In all studies, brain radioactivity was measured in a series of consecutive frames. To further correct for head movement, all frames were realigned to the first minute of acquisition (*131*).

2.5   Arterial blood sampling

In Study IV, V, VII and VIII, blood samples were obtained using an arterial catheter. The radioactivity levels in blood were then measured using an automated blood sampling system (ABSS, Alogg Technology, Mariefred, Sweden). Only ABSS data collected for the first 5 minutes of the examination was used. From the start to the end of the scan, manual blood samples were also drawn at subsequent time points. Samples of plasma were obtained by centrifugation of a subset of manual blood samples. Radioactivity levels in these blood and plasma samples were measured in a well counter. Plasma curves were derived by multiplying the blood curves with plasma-to-blood radioactivity ratio values. Curves of the remaining parent fraction of radioligand in plasma were obtained using high-performance liquid chromatography. The plasma curves were then corrected for metabolites by multiplying them with the parent fraction curves. These metabolite-corrected plasma TACs were then used as AIFs in the kinetic modeling.

2.6   Definition of regions of interest

All participants included in this thesis underwent MRI scans and in order to derive structural T1 weighted MR images. For studies carried out at KI, three different MRI systems were used. All subjects in Study I were examined using the 1.5T GE Signa system (Milwaukee, WI). Subjects in Study II and VII were examined using the 1.5T Siemens Magnetom Avanto system (Erlangen, Germany). Subjects in Study III were examined using either the 1.5T GE Signa (all patients and six healthy controls) while the remaining controls (n=10) were examined using a 3T GE Signa (GE MR750, GE Medical Systems, Milwaukee, WI, USA). All subjects in Study VI, V and VIII were examined using the 3T GE Signa.

ROIs were then deliniated on the T1 weighted MR images. In Study I, striatal ROIs were manually drawn using previously published guidelines (*132*), while the

extrastriatal ROIs were delineated using the FreeSurfer software (version 5.0.0, http://surfer.nmr.mgh.harvard.edu/). In all remaining studies carried out at KI, all ROIs were automatically defined using either the FMRIB FSL software (*133*) (Study II) or FreeSurfer (Study III-V, VII and VIII).

A manually delineated cerebellum was used as reference region in Study I and III. Remaining studies that included a reference region used either a FreeSurfer (Study V and VII) or FSL (Study II) defined grey-matter cerebellum, or the FreeSurfer whole-brain ROI (Study V).

Following ROI delineation, all T1 weighted MR images were co-registered to the PET images. In Study I-II, IV-V and VII-VIII TACs of radioactivity concentration were then extracted for all ROIs.

2.7 Kinetic modeling

In this thesis, three different kinetic models have mainly been used to estimate radioligand binding to target: the unconstrained 2TCM, SRTM (*28*), and the non-invasive Logan plot fitted with a multilinear regression (*134,135*).

For all studies involving the [11C]PBR28 or (R)-[11C]PK1195 radioligands (Study IV-VIII), the unconstrained 2TCM with AIF was used. For each subject and ROI, a V_T value was derived. Although V_T, in theory, should have lower signal-to-noise ratio (SNR) than e.g. V_S, it is still widely applied as outcome measure. This is because V_T often show higher reliability and less variability compared specific binding outcomes obtained from the 2TCM (*136,137*). Hence, in Study IV and VI, the V_T values were used to compare patients with psychosis or schizophrenia to healthy control subjects.

In Study I and II, the SRTM was used with cerebellum as reference region, yielding ROI [11C]SCH23390 BP_{ND} values. These BP_{ND} values were then correlated with SSP scale scores, to examine the relationship between pro and anti-social personality traits and D1-R availability in striatum and extrastriatal regions.

2.7.1 Pseudo-reference region and supervised cluster-analysis

When non-negligible levels of target are expressed throughout the brain, it is not possible to establish a true reference region. For example, in the case of [11C]FLB457,

the radioligand can be displaced in the cerebellum, which suggests the presence of specific binding (*138,139*). In the case of [¹¹C]PBR28 and (R)-[¹¹C]PK11195, TSPO is expressed throughout the brain and there is no region that show even small levels of uptake. When this is the case, it might still be possible to use a region as a "pseudo-reference region", but only under the assumption that the availability of target in this region does not differ between the groups that are being compared, or that the difference is minuscule. In Study III, which compares SAD patients to healthy control subjects, [¹¹C]FLB457 BP_{ND} values were derived using the cerebellum as pseudo-reference region.

For (R)-[¹¹C]PK11195, the cerebellum has also been suggested to function as a suitable pseudo-reference region. The SRTM, with cerebellum as reference input, has hence been used in many clinical studies, such as comparisons between controls and patients with schizophrenia or psychosis (*79–81*). It is however unclear if this is a sensible approach, since 1) the radioligand uptake is cerebellum is very high, 2) there is little to no evidence that the cerebellum be spared of pathology in any psychiatric disorder, and 3) it has not yet been established if the resulting BP_{ND} values are reliable.

Another alternative approach to quantify (R)-[¹¹C]PK11195 expression in the brain, without the need for arterial sampling, is to use SVCA4 method (*96*). The SVCA4 method is performed on brain-masked dynamic PET images and aims to segment the voxels into different classes characterized by distinct kinetic profiles: 1. grey-matter with high specific binding 2. grey-matter with low specific binding, 3. white-matter, 4. soft tissue. In order to do so, the method compares voxel TACs with a set of pre-defined kinetic classes, derived from healthy subjects or a sample of subjects thought to have inflammated brain tissue (*95*). The goal is to identify all voxel TACs most similar to the kinetic class assumed to contain negliable levels of specific binding. These voxels are then combined into a ROI from where a TAC is extracted that can function as reference input to a kinetic model.

In Study VII, (R)-[¹¹C]PK11195 BP_{ND} values were derived using SRTM with either cerebellum or SVCA4 grey-matter TAC as reference input. The test-retest reliability and convergent validity of these BP_{ND} values were then evaluated.

2.7.2 Parametric imaging

In Study III, a BP_{ND} value was calculated for each voxel in the brain, using cerebellum as pseudo-reference region. The HRRT yields PET files that consist of about 2 million voxels, and using a non-linear model to fit each voxel TAC (such as the SRTM) would take considerable amounts of time. Instead, the non-invasive multilinear Logan analysis (*134,135*) was employed:

$$\int_0^t C_T(\tau)d(\tau) = DVR\left(\int_0^t C_R(\tau)d(\tau) + \frac{C_R(t)}{k_2'}\right) + bC_T(t) \qquad (2.1)$$

where C_T is the concentration of radioligand in the target ROI, and C_R is the concentration in the reference region. The original non-invasive Logan graphical analysis yields a plot of data which become linear after pseudo-equilibrium has been reached (t*). In Study III, t* was set to 27 minutes. From t* to the end of the examination, a linear regression model can be fitted, and the regression-coefficient of the first predictor in equation 2.1 is the estimated distribution volume ratio (DVR = V_T/V_{ND}). Subtracting one from this parameter yields the BP_{ND} estimate for the voxel of interest. Applying the Logan analysis to all voxels in the brain in Study III yielded a parametric map of [^{11}C]FLB457 BP_{ND} values for each participant.

Since there is substantial amount of noise in each individual voxel TACs, the voxel-wise estimation of BP_{ND} becomes prone to error. For this reason, many different methods for noise-reduction have been developed for parametric imaging in PET (*140*). In Study III, the 3D stationary wavelet transform approach was used, as it has shown to effectively reduce noise for [^{11}C]FLB457 imaging (*134*).

In Study III, the final parametric maps of BP_{ND} were used to both apply statistical parametric mapping (SPM) analysis on voxel-level, and extract mean voxel BP_{ND} values using larger ROIs, in order to compare SAD patients with control subjects.

2.7.3 Ratio-based outcome measures

If arterial samples of radioactivity have not been collected, and no reference region exist, the commonly applied outcome measure for radioligand uptake is ROI SUV. However, the SUV has some critical limitations: 1) it often show large variability and 2) it contain signal that is not related to the true density of the target, such as radioactivity contribution from vasculature. Using SUV as outcome also relies

on the assumption that there are no differences in radioligand delivery to the brain between the individuals or groups that are being compared.

In order to reduce the high variability and low sensitivity of outcomes from [11C]PBR28 studies without arterial blood samples, it has been suggested that ROI SUVs can be divided or normalized by a SUV stemming from a pseudo-reference region. The idea is that this would yield more reliable outcomes measure with higher sensitivity to detect effects ($91,92$). [11C]PBR28 V_T values, obtained by using AIFs, also show large variability even after accounting for differences in the TSPO affinity genotype. V_T values have therefore also been suggested to benefit from being normalized by V_T from a pseudo-reference region, such as the cerebellum, occipital cortex or the whole-brain ($92,141$).

In Study V, SUVRs and "distribution-volume-ratios" (DVRs) were derived by dividing target ROI SUV or V_T values by the SUV or V_T from the whole-brain. The test-retest reliability and convergent validity of the ensuing ratios were then evaluated.

2.7.4   Simultaneous estimation

The simultaneous estimation (SIME) technique aims to model TACs from several brain regions simultaneously to estimate a brain-wide V_{ND} value for a subject (142). This is done by replacing the rate constant K_1 in equation 1.4 with $V_{ND} \cdot k_2$. A unique k_2 is then fitted for each ROI, but V_{ND} is constrained to be the same across all ROIs for a subject. A set of different V_{ND} values are then evaluated for each subject, and the value that results in the best fit across all ROIs is taken as the final brain-wide estimate of V_{ND}. This value ($V_{ND:SIME}$) can then be used to derive outcome measures of specific binding:

$$BP_{ND} = \frac{V_T - V_{ND:SIME}}{V_{ND:SIME}} \tag{2.2}$$

$$V_S = V_T - V_{ND:SIME} \tag{2.3}$$

where V_T is calculated using the unconstrained 2TCM.

In Study VIII, the accuracy and reliability of the SIME method applied on [11C]PBR28, were evaluated. First, simulations were used to examine if SIME could estimate an known "true" value of V_{ND}. Second, data from an already published

pharmacological challenge study (116) was re-analysed to compare estimates of V_{ND} using SIME, the revised Lassen plot (143) and the Likelihood Estimation of Occupancy (144) techniques. Third, the test-retest data used in Study V was re-analysed to examine the reliability and precision of SIME derived BP_{ND} and V_S values.

2.8 Statistical analyses

2.8.1 Correlations and replications

In Study I and II, Pearson's correlation coefficients were used to examine the relationship between SSP scale scores and [11C]SCH23390 BP_{ND} in the ROIs. In Study I, partial correlations were used, with age and gender as covariates.

Study II was a direct replication attempt of the relationships found in Study I. First, Pearson's r were employed to assess if these relationships were significant in a new and independent data set. However, since a p-value in itself does not convey much information on how successful a replication is, Study II also employed a procedure known as replication Bayes factor (BF). The replication BF examines how much support there is in data for the original finding (H1), relative to a correlation of zero (H0) (145). BF is the extent to which data updates the prior odds of two competing models, to posterior model odds (104). As such, it can be used as an estimate of how much support (or evidence) there is in data for one hypothesis over another. For the correlation replication BF, H1 was defined as the posterior distribution of the correlation coefficient from the original study (calculated using a uniform prior), and H0 was defined as a point null hypothesis of no effect (146). A BF above 3 is commonly interpreted as providing moderate support for one hypothesis over the other, and a BF above 10 as strong support. Hence, if BF10 = 10, this means that the observed data is 10 times more likely under the original hypothesis compared to the null, signifying strong evidence in favor of a successful replication.

2.8.2 Patient-control differences

In Study III, multiple univariate two-tailed one-way independent samples ANCOVAs were used to assess patient-control differences in extrastriatal [11C]FLB457 BP_{ND}, while covarying for age.

In Study IV, a two-tailed one-way independent samples ANCOVA was used to assess patient-control differences in whole grey-matter [^{11}C]PBR28 binding, while covarying for TSPO genotype and gender. In the article (*147*), four additional ROIs were examined: frontal cortex, temporal cortex, hippocampus and whole white-matter. Due to the high interregional correlations in V_T only the results from grey-matter are presented in this thesis.

The ANCOVA models used in Study III and IV are similar to an independent samples t-test (i.e. two group means are being compared), with the addition of allowing covariates to be entered into the model.

In Study VI, Bayesian linear mixed-effects models were used to assess patient-control differences in TSPO levels, while taking into account the hierarchical and nested structure of data. Four different models of increasing complexity were evaluated, where PET center and TSPO genotype were entered as random effects, allowing intercepts and slopes to vary. The model with the best out-of-sample deviance was selected and used to assess patient-control differences. Bayesian hypothesis testing, using BFs (*148,149*), was employed to assess the evidence in favor of patients with psychosis having higher or lower TSPO binding, compared to control subjects. BFs of standardized patient-control differences in V_T were computed using the Savage-Dickey Ratio (*150*).

The study design and statistical modeling of Study VI were pre-registered prior to performing any literature search or analysis of data. The pre-registration protocol can be found on https://github.com/pontusps/TSPO_psychosis.

2.8.3 Test-retest metrics

In Study V, VII and VIII, test-retest properties of radioligands and quantification methods were evaluated using a set of different metrics. The absolute variability (AbsVar) is an estimate of repeatability and was calculated by the following formula:

$$AbsVar = \left(\frac{|PET1 - PET2|}{(PET1 + PET2)/2} \right) \times 100 \qquad (2.4)$$

where PET1 is the ROI outcome measure (such as V_T or BP_{ND}) in the first PET examination, and PET2 is the ROI outcome measure in the follow-up examination. As such, AbsVar is a measure of absolute percentage change. A value of 0 indicates

that there has been no change between test and retest, and a value of 100 indicates that the change is as large as the mean of both PET examinations.

The intra-class correlation coefficient (ICC) is an estimate of the reliability or differentiability of subjects. It aims to calculate the proportion of variance that is due to "true" signal of the total signal (true + error):

$$ICC = \frac{MS_B - MS_W}{MS_B + MS_W} \approx \frac{\sigma_{true}}{\sigma_{true} + \sigma_{error}} \tag{2.5}$$

where MS_B is the between subjects mean sum of squared variance and MS_W is the within subjects mean sum of squared variance. An ICC of e.g. 0.5 indicates that only half of the variance of the outcome measure is due to "true" signal, while the remaining is due to error. According to previously suggested guidelines (*151*), an ICC between 0.5 and 0.75 should be interpreted as indicating poor to moderate reliability, 0.75 to 0.9 as good, and above 0.9 as excellent and suitable for usage in clinical applications.

The standard error of measurement (SEM) (*152*) is related to the ICC but reflects the precision of individual scores in the same unit of the outcome measure:

$$SEM = SD\sqrt{1 - ICC} \tag{2.6}$$

where SD is the standard deviation of the sample. The SEM can be thought of as estimate of the standard deviation around each point-estimated outcome value. E.g. if BP_{ND} for a subject is 1, and the SEM is 0.5, we expect 68% of all follow-up measures to, on average, fall within 1 ± 0.5 assuming that the true density of target is kept constant. The SEM can be recalculated to the minimal detectable differences (MD) metric:

$$MD = SEM \times 1.96 \times \sqrt{2} \tag{2.7}$$

This translates the SEM into the difference between two measurements that is considered sufficiently large in order for the points to be significantly different from each other, according to 95% confidence interval. A small MD indicates that the outcome measure is precise, and that it will be easier to detect a difference between two subjects, or two time-points, compared to an outcome showing a large MD. The MD is often scaled so that it is expressed as percentage of the sample mean of the

outcome measure. This is done to aid comparison between different outcomes.

2.9 Statistical disclosure statements

When studies lack pre-registration protocols it can be difficult for the reader to assess how much evidential value the reported results contain, due to potential use of intended and non-intended questionable research practices (*153*). In order for the reader of this thesis to be able to better judge the robustness of the inference reported in the included studies, this section presents statistical disclosure statements. The following is a transparent listing of all analytical choices carried out in the studies, but not reported in the published articles.

Study I Additional statistical analyses were performed as part of this study. First, additional personality scales were analyzed (all extroversion and agreeableness sub-scales from the NEO-PI-R inventory). Second, 6 additional subjects were initially included but later excluded from the study due to the time interval between PET and personality assessment being too long, which is reported in the article. However, additional statistical analyses were also run with these subjects included, and this was not reported. Third, all reported correlations had been run both with and without controlling for gender and age. Forth, a Statistical Parametric Mapping (SPM) analysis was performed but not reported in the final version of the article.

Study III Additional statistical analyses were performed as part of this study. First, age as a covariate in the main ROI based and voxel-based analysis was not included in the initial analysis but added at a later stage. Second, two additional subjects were initially included in the healthy control group but later found to be taking diabetes and SSRI medication respectively, and therefore excluded from the study. The family-wise error rate correction method was additionally tried out to correct for multiple comparisons on voxel-level but results from this were not reported.

Study II and IV-VII All statistical analyses that were performed have been reported in the articles. All data from all subjects that was planned to be part of the studies has been included and reported, unless otherwise stated in the articles. However, in Study VI, two additional analyses that were not in the pre-registration protocol were performed following requests from reviewers. These deviations have been time-stamped and logged together with the pre-registration protocol (see https://github.com/pontusps/TSPO_psychosis).

2.10 Data and code availability

Study I The data has not been made publicly available due to current institutional regulatory restrictions.

Study II All derived data (e.g. individual BP_{ND} values and personality scores) and code used for Study II can be found on https://osf.io/te5q7/.

Study III The data has not been made publicly available due to current institutional regulatory restrictions. However, data extracted from Figure 3.3 can be found on https://osf.io/2jm9h/. The original code for reproducing the statistical analysis was not saved at the time of the study.

Study IV The data has not been made publicly available due to current institutional regulatory restrictions. However, data extracted from Figure 3.6 can be found on https://osf.io/2jm9h/. The original code for reproducing the statistical analysis was not saved at the time of the study.

Study V All derived data (e.g. TACs and genotype information) and code is available on https://github.com/mathesong/PBR28_RatioMethods.

Study VI The majority of data used in this study was not collected by the authors, and the first author does not have permission to share it publicly. However, data extracted from Figure 3.9, and code for performing the main analysis of the study can be found on https://osf.io/2jm9h/.

Study VII All derived data (e.g. TACs) and code is available on https://osf.io/gcn4w/.

Study VIII The majority of data used in this study was not collected by the authors, and the first author does not have permission to share it publicly. The code for executing the main model evaluated in the article can be found on https://github.com/martinschain/SIME.

Chapter 3

Results and Discussion

3.1 Overview

In this thesis, several different radioligands and quantification methods have been implemented, evaluated and applied on clinical data.

Study I and II examined the relationship between D1-R availability and self-rated pro and anti-social behavior in healthy subjects. Study I found a positive correlation between striatal D1-R availability and Social Desirability, and a negative correlation to Trait Aggression. Study II did however fail to replicate these results.

In Study III, D2-R availability in limbic and cortical regions in patients with SAD and healthy controls were compared using both ROIs and voxel-wise analyses. SAD patients were found to have higher D2-R availability in the DLPFC and OFC, although the results warrant replication in a larger sample.

In Study IV, TSPO expression in the whole of grey-matter in patients with first-episode psychosis and healthy controls were compared, and patients were found to have lower TSPO binding.

In Study V the test-retest reliability and convergent validity of alternative ways to measure [^{11}C]PBR28 binding were evaluated. DVRs and SUVRs, derived using pseudo-reference regions, showed both poor reliability and convergent validity.

Study VI carried out a meta-analysis of TSPO expression in patients with schizophrenia and psychotic disorders compared to healthy controls. Strong evidence was found in favor of patients having lower TSPO binding in both cortical and

subcortical regions.

In Study VII the test-retest reliability and convergent validity of alternative ways to estimate (R)-[11C]PK11195 binding were evaluated. BP_{ND} values, derived using pseudo-reference TACs with the SRTM, were unreliable and showed no convergent validity to outcomes derived using AIFs.

Finally, Study VIII aimed to evaluate the reliability and accuracy of a new modeling method to estimate specific [11C]PBR28 binding without requiring a reference region. Simulations, a pharmacological challenge and test-retest data showed that V_{ND} and ensuing V_S values derived using this method, were accurate, precise and reliable.

3.2   Study I

Study I explored relationships between [11C]SCH23390 BP_{ND} in a set of ROIs and SSP Social Desirability, as well as Physical and Verbal Trait Aggression scale scores (Table 3.1). Social Desirability was positively correlated to D1-R availability in all striatal regions, when controlling for age and gender (using an uncorrected alpha threshold of 0.05). Physical Trait Aggression showed a negative relationship to D1-R availability in the limbic striatum, while Verbal Trait Aggression was not related to BP_{ND} in any ROI. In extrastriatal regions, Social Desirability showed significant positive correlations with D1-R availability in amygdala and medial frontal cortex, after using the "p-plot" method (*154*) to correct for multiple comparisons (Table 3.1).

The results from Study I should however be considered exploratory and not confirmatory. The reasons for this are that additional statistical analyses were performed but not presented in the study (see section on Statistical Disclosure statements in Methods), and because the alpha threshold was not adjusted for multiple comparisons for the striatal ROI analyses. The results hence carry limited evidential value. Instead, they can be thought of as hypothesis-generating: [11C]SCH23390 BP_{ND} in the limbic striatum will be positively correlated to the pro-social personality trait Social Desirability, and negatively correlated to anti-social traits, such as aggression.
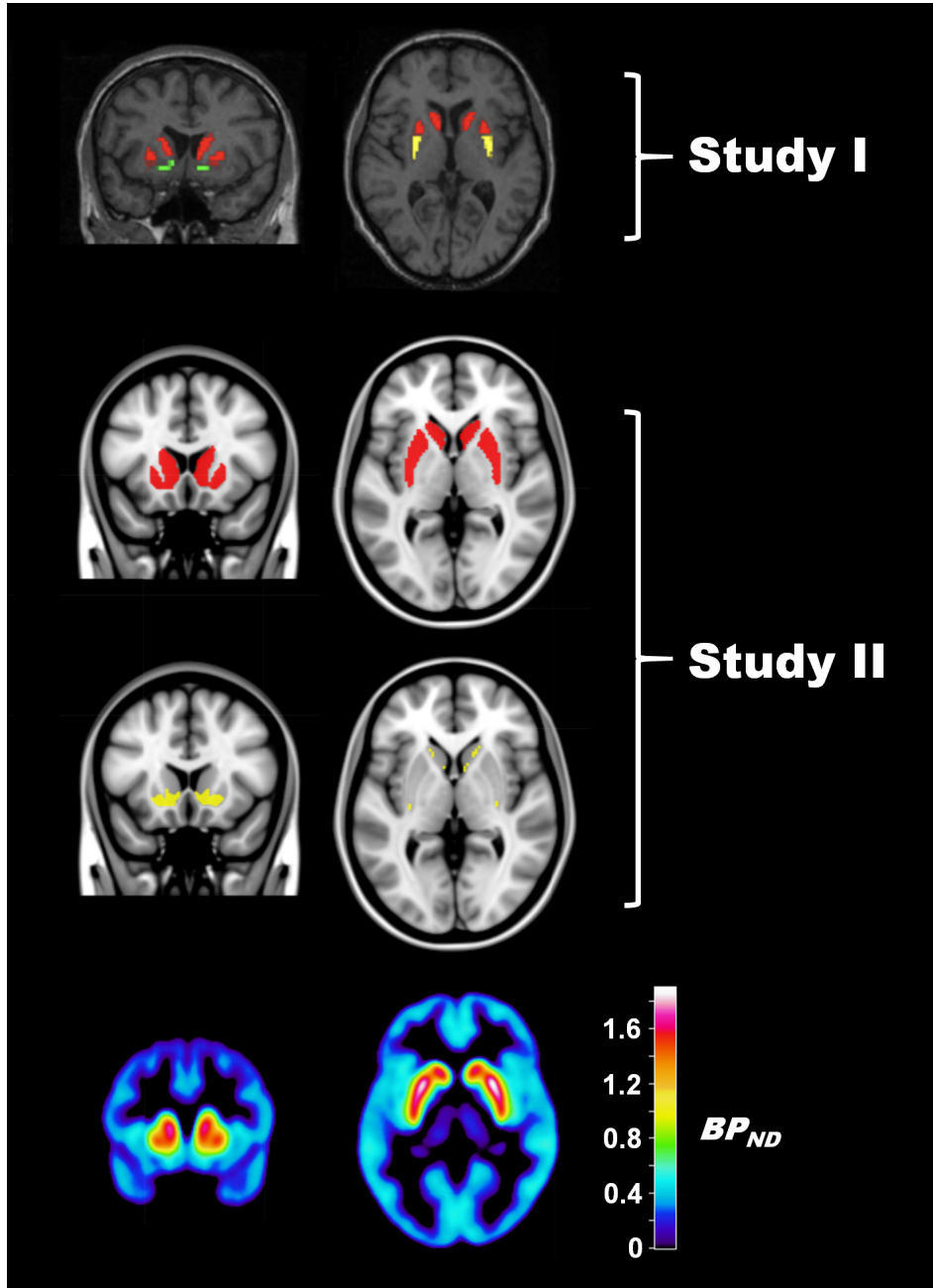
Figure 3.1: Example of ROI delineation in Study I (manually drawn on an individual T1-weighted MR image) and II (defined using the FSL Oxford-GSK-Imanova connectivity atlas of striatum). In Study I the limbic striatum is depicted in green, and in Study II in yellow. The bottom panel displays an average [^{11}C]SCH23390 BP_{ND} brain map from all subjects in Study II.

Table 3.1: Pearson's partial correlations (controlling for age and sex) between D1-R availability and personality traits reflecting pro and anti-social behavior. Alpha was set to 0.05, uncorrected for striatal regions and family-wise-error rate corrected for extrastriatal regions. LST = limbic striatum; AST = associative striatum; SMST = sensorimotor striatum; ACC = anterior cingulate cortex; HIP = hippocampus; INS = insula; MFC = medial frontal cortex; LFC = lateral frontal cortex; OFC = orbitofrontal cortex

Region	Social Desirability			Physical Trait Aggression			Verbal Trait Aggression		
	r	p	df	r	p	df	r	p	df
Striatal ROIs									
Striatum	0.54	0.012	19	0.36	0.106	19	0.13	0.578	19
LST	0.52	0.015	19	0.51	0.019	19	0.37	0.101	19
AST	0.55	0.009	19	0.31	0.165	19	0.09	0.700	19
SMST	0.67	0.001	19	0.35	0.120	19	0.02	0.926	19
Extrastriatal ROIs									
ACC	0.51	0.018	19	0.55	0.010	19	0.29	0.199	19
AMG	0.60	0.006	18	0.41	0.072	18	0.24	0.316	18
HIP	0.49	0.033	17	0.46	0.048	17	0.50	0.028	17
INS	0.47	0.031	19	0.50	0.021	19	0.46	0.038	19
MFC	0.60	0.004	19	0.39	0.077	19	0.17	0.461	19
LFC	0.55	0.010	19	0.35	0.117	19	0.01	0.955	19
OFC	0.53	0.013	19	0.50	0.020	19	0.19	0.398	19

3.3   Study II

Study II aimed to replicate and confirm the exploratory findings from Study I. The relationships between Social Desirability and Physical Trait Aggression and [^{11}C]SCH23390 BP_{ND} in whole-striatum and, in particular, the limbic striatum were deemed to be most interesting and carry the highest chance of replicating. Hence, these relationships were analyzed using Pearson's r and replication BFs in a new and independent sample of subjects (Figure 3.2).

The results showed that none of the correlations between personality scale scores and D1-R availability were significant (Figure 3.2). Instead, replication BF revealed that there was moderate to strong evidence in favor of failed replications. The data supported to null-hypothesis up to 12 times more than it supported the original findings from Study I (Table 3.2).

In conclusion, the results of Study II did not confirm the exploratory findings from Study I. Instead, the results indicated evidence in favor of no correlation between D1-R availability and pro and anti-social personality traits, as measured using the SSP. A potential reason for this failed replication is that a more diverse sample was used in Study I, and the subsequent loss of variance in Study II meant that

Table 3.2: Correlations between Social Desirability and Physical Trait Aggression scores and ROI [^{11}C]SCH23390 BP_{ND} from Study I and Study II. The replication BFs denotes how much support there is for a successful replication (BF10) compared to that of no correlation (BF01). Note that the correlation between Physical Trait Aggression and STR was not significant in the original study but has still been included here for completeness. STR = whole-striatum; LST = limbic striatum

	Original study[a]			Present study [b]			Replication BF	
	r	df	p-value	r	df	p-value	BF01	BF10
Social Desirability								
STR	0.54	19	0.012	-0.12	24	0.73	12.4	0.08
LST	0.52	19	0.015	-0.03	24	0.57	7.2	0.14
Physical Trait Aggression								
STR	-0.36	19	0.106	-0.08	24	0.35	2.0	0.51
LST	-0.51	19	0.019	-0.09	24	0.32	3.3	0.31

[a] two-sided test
[b] one-sided test in direction of original study

associations were difficult to detect. However, another likely explanation is that the results in Study I were false positives and that there is no direct relationship between striatal D1-R availability and Social Desirability or Trait Aggression in humans.

3.4 Study III

Study III examined the difference in extrastriatal [^{11}C]FLB457 BP_{ND} between patients with social anxiety disorder and healthy controls. Univariate ANCOVAs, controlling for age, showed no significant patient-control differences in BP_{ND} in any of the the ROIs (Figure 3.3).

Using the wavelet-aided parametric imaging approach (*134*), BP_{ND} values in each voxel were calculated for each participant. These individual brain-wide voxel-maps were normalized into the Montreal Neurological Institute standard stereo-tactical space. All maps were then smoothed with a cubic Gaussian kernel (FWHM 8x8x8 mm). The ROIs presented in Figure 3.3 were concatenated into a brain mask that was used in the SPM analysis of voxel-wise group differences (Figure 3.4).

An ANCOVA, controlling for age, revealed significant group differences (using false discovery rate correction, FDR with alpha = 0.05) on voxel-level, with patients showing higher BP_{ND} in multiple voxels in cortical regions compared to controls
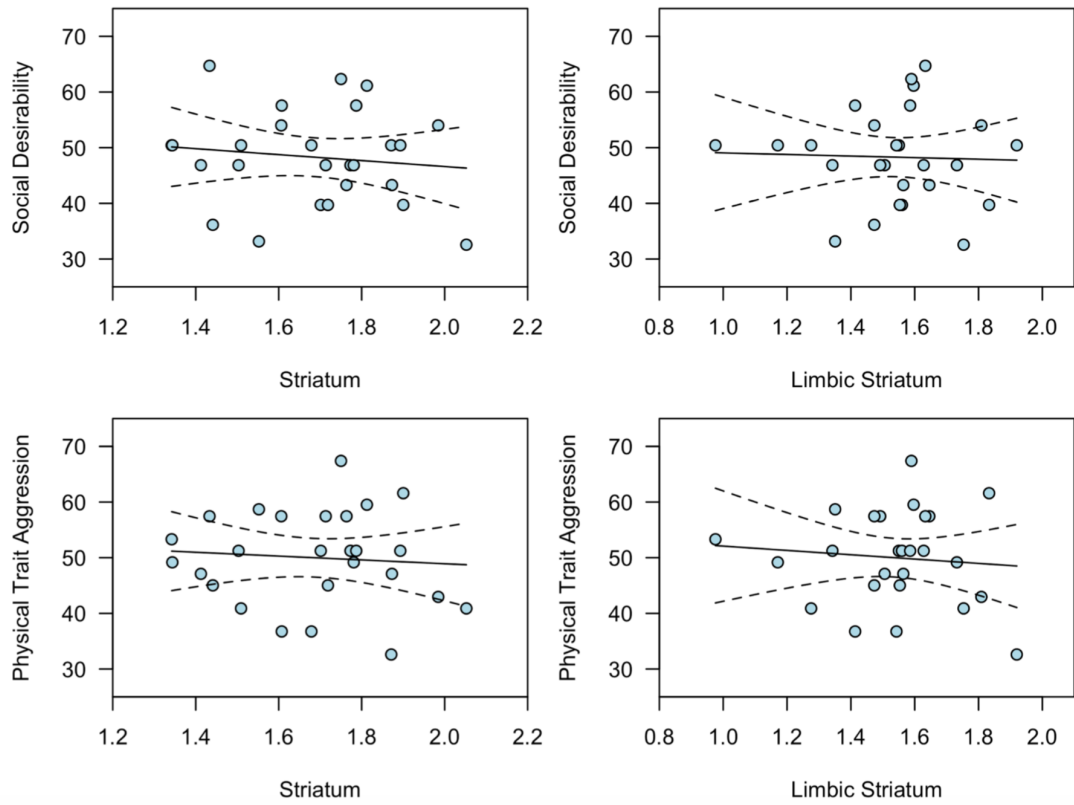
Figure 3.2: Relationships between [¹¹C]SCH23390 BP_{ND} in striatum and Social Desirability and Physical Trait Aggression. The dotted lines indicate the 95% confidence intervals. Raw scale scores have been transformed to T-scores for illustrative purposes in this figure.

(Figure 3.4). The FDR correction method does however come with an important caveat: Contrary to FWEr correction, it does not control error rates in such manner that the risk of obtaining any false positive voxel is, on average, set to 5%. Instead, FDR correction sets the error rate so that, on average, not more than 5% of significant voxels will be false positives. This entails that a small and confined cluster of significant voxels could, in theory, all be false positives (*155*). For this reason, FWEr correction on cluster-level was also employed in this study, revealing significant differences in the OFC and right dorsolateral prefrontal cortex (DLPFC).

All patients (n = 12) and a subset of control subjects (n = 10) filled out the self-rating version of the Liebowitz Social Anxiety Scale (LSAS-SR) (*156*) prior to participating in the PET examination. There was no significant correlation between LSAS-SR scores and voxel [¹¹C]FLB457 BP_{ND} values within the pre-defined brain mask (results not shown), after correction for multiple comparisons.

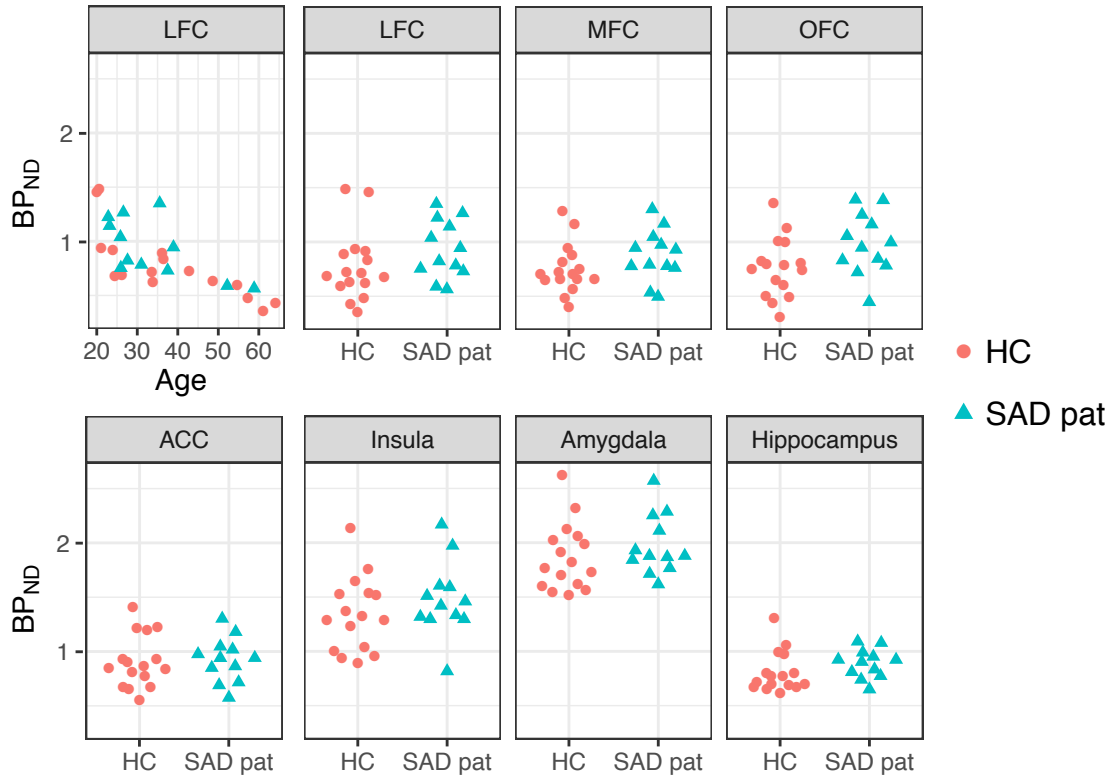One important caveat with the study is that healthy control subjects showed

Figure 3.3: Comparisons of [^{11}C]FLB457 BP_{ND} values between 12 SAD patients (SAD pat) and 16 healthy controls (HC) in cortical and subcortical ROIs. None of the ROIs showed significant differences between the groups. There was a negative correlation between age and BP_{ND} in all regions (here only LFC is shown), and age was therefore included as covariate in the statistical analysis.

numerically higher SUVs in the reference region (cerebellum). The group difference in SUVs was not significant (t=1.64, df=26, p=0.11, Cohen's d=0.65), but a visual inspection of the data (plot not displayed) showed that three control subjects had substantially higher values compared to the remaining participants. SUVs are not suitable as direct estimates of specific binding, but this result still suggest that healthy control subjects might have higher specific uptake in the reference region. This could in turn fully or partly cause the observed finding of lower BP_{ND} in control subjects. However, when excluding the three healthy control showing higher SUVs in cerebellum, the patient-control differences in cortical BP_{ND} still remained significant in the SPM analysis, although with smaller effect sizes.

The results from Study III suggest that patients with SAD might have higher D2-R availability in frontal cortex, compared to healthy controls. In SAD, a large part of the experienced distress and avoidance behavior expressed by patients is likely caused by anticipatory anxiety (*46*). OFC is essential for signaling expectations of
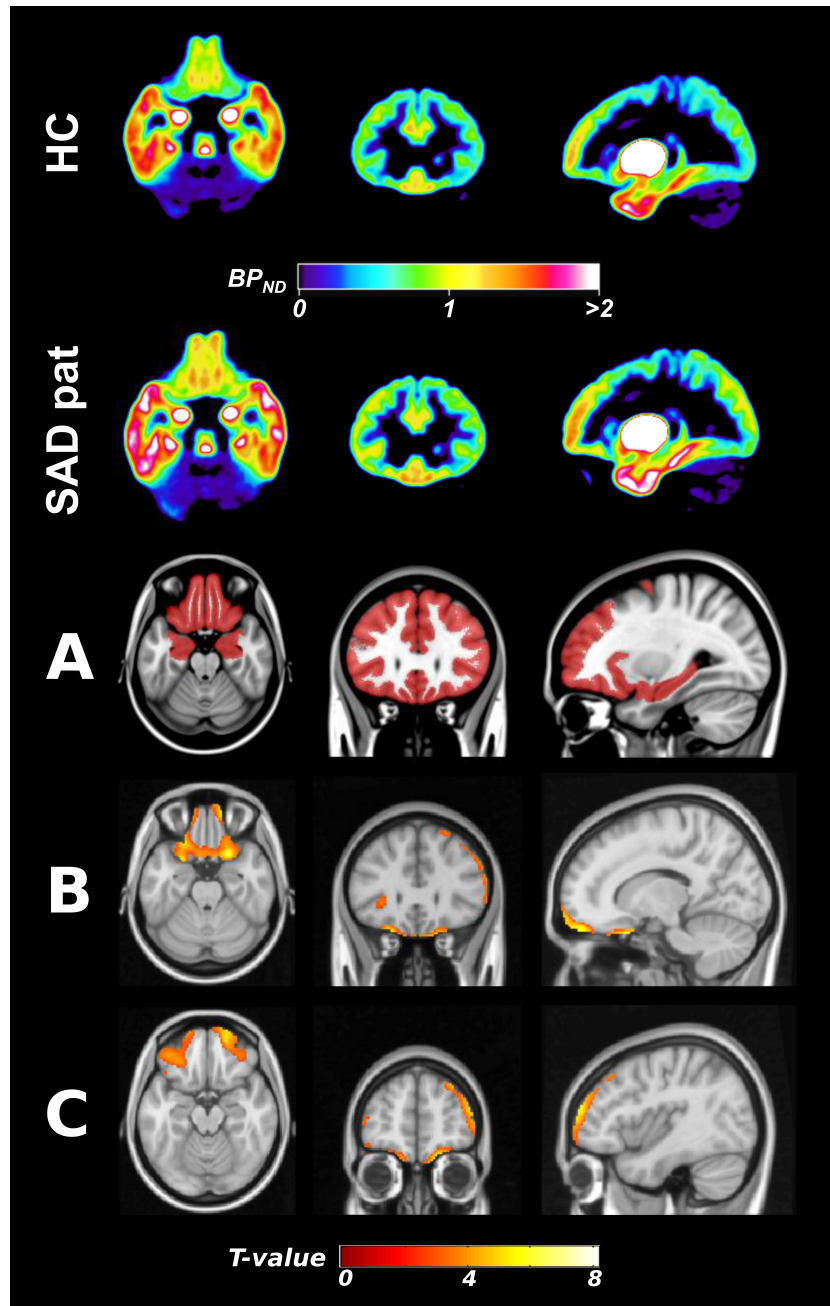
Figure 3.4: In the two upper panels, average [^{11}C]FLB457 BP_{ND} brain maps for 16 healthy control subjects and 12 social anxiety disorder patients (SAD pat) are displayed. The ROIs were concatenated into a mask (A), which was then used to restrict the voxel-wise group comparison. In the SPM analysis, SAD patients expressed elevated levels of D2-R availability within bilateral orbitofrontal cortex (B) and right dorsolateral prefrontal cortex (C).

future outcomes, such as predictive representations of sensory input and emotional valence (*157*). The region is also part of processing anticipatory anxiety by inflating predictions of threats (*158*). These predictions are relayed to the DLFC, a region shown to be responsible for signaling the mismatch between expectations and actual sensory input, i.e. prediction errors (*159*). Prediction errors of aversive stimuli have in turn been shown to be mediated by the midbrain dopamine system (*160,161*). Hence, it can be speculated that an aberrant extrastriatal dopamine D2-R signaling system in cortical regions is linked to critical disease mechanisms in SAD, such as processing of fearful social stimuli or avoidance responses.

In addition to the group difference in cerebella SUVs, two caveats in this study were that a small sample size was used, and that analysis plan was not pre-registered. Another important limitation is that the smoothing of parametric PET images can induce substantial amount of variability, especially on the edge of the fronto-cortical surface. Hence, 3D smoothing can make cortical voxel BP_{ND} values unreliable and prone to produce outliers (*113*). The findings from Study III therefore need to be replicated in order to assess the robustness of the results. Preferable, such a replication should be carried out in a much larger sample and using only ROIs.

3.5   Study IV

In Study IV, the TSPO levels in drug-naive patients with first-episode psychosis (FEP) (n = 16, 11 male, 5 female, mean age = 28.5 ± 9.4SD) was compared to healthy control subjects (n = 16, 7 male, 9 female, mean age = 26.4 ± 8.4SD) using the whole grey-matter as ROI. TSPO expression is thought to function as an index of microglial activation, which in turn is considered to indicate immune activation in the brain. Hence, in line with the immune hypothesis of schizophrenia, the hypothesis was that patients would show higher [^{11}C]PBR28 binding compared to controls. However, previous to this study, another article had just been published reporting no differences in [^{11}C]PBR28 V_T between patients with schizophrenia and controls. Instead, the effect size of difference suggested that patients might express lower levels of TSPO in the whole grey-matter (*93*). For this reason, we employed a two-tailed statistical test, examining group differences in V_T both directions.

All patients were recruited from psychiatric wards and out-patient clinics in the Stockholm region. At the time of the study, all patients were naive to anti-psychotic medication. According to DSM-IV, they met the diagnostic criteria for schizophrenia
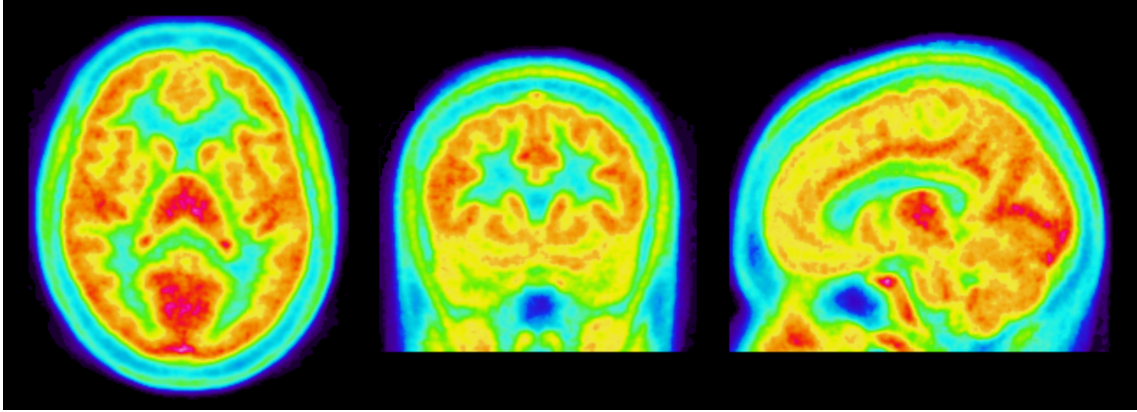
Figure 3.5: Average summation PET image of 32 subjects examined with [^{11}C]PBR28.

(n=4), schizophreniform psychosis (n=7), psychosis NOS (n=4) or brief psychosis (n=1). During the course of the study, antidepressants as well as occasional medication with sedatives and anxiolytics were allowed.

Two different outcome measure of TSPO expression in the brain were used: [^{11}C]PBR28 V_T and DVR. DVRs were calculated by dividing the whole grey-matter [^{11}C]PBR28 V_T with whole-brain [^{11}C]PBR28 V_T. The reason for including DVR was because a previous study suggested that this outcome measure yielded higher sensitivity to detect patient-control differences compared to V_T (*93*). In that study, the difference in DVR between patients with schizophrenia (n=14) or subjects with high-risk of developing a psychotic disorder (n = 14) and healthy controls were found to be of a very large effect size. Hence, Study IV aimed to replicate this finding, using a slightly larger sample of patients.

Patient-control differences in whole grey-matter [^{11}C]PBR28 V_T and DVRs were examined using an ANCOVA, controlling for gender and the TSPO genotype (patients = 8 HABs and 8 MABs; controls = 9 HABs and 7 MABs). The alpha level was set to 0.05. The patient group showed significantly lower [^{11}C]PBR28 V_T in the whole-grey-matter ROI, compared to control subjects (F = 6.19, df = 1.28, P = 0.019). The difference between HAB patient-controls ($\eta^2 = 0.38$) was larger than that of MAB patient-controls ($\eta^2 = 0.02$) (Figure 3.6). There was however no group difference in [^{11}C]PBR28 DVR (F = 0.07 df = 1.28, p = 0.80), using grey-matter V_T in the numerator and the whole-brain V_T in the denominator (Figure 3.6). As a result, this study did not replicate the previous finding of higher DVR in patients (*93*).

Contrary to the hypothesis of higher TSPO levels in patients with psychosis, this
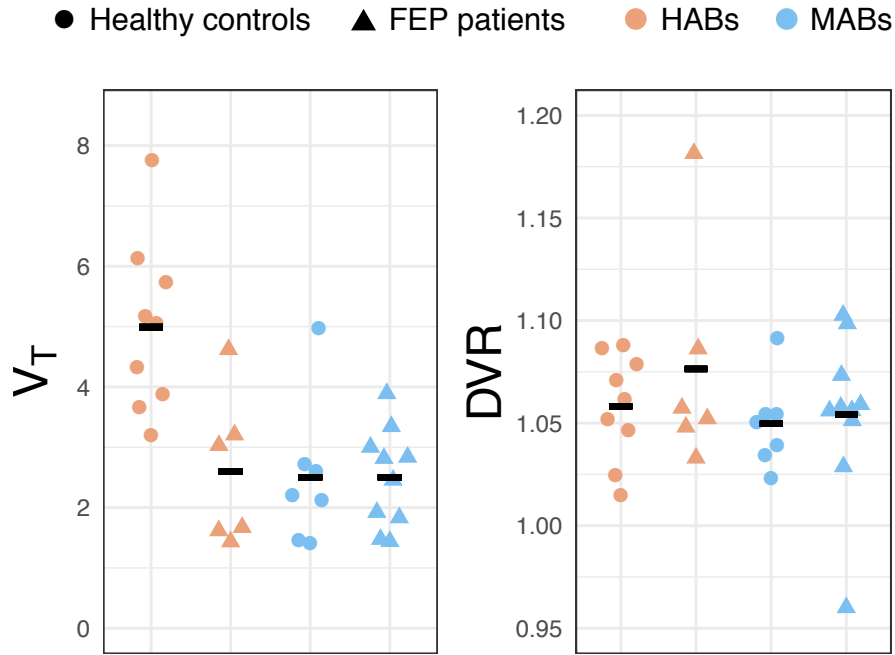
Figure 3.6: Differences between FEP patients and healthy controls in $[^{11}\text{C}]$PBR28 grey-matter V_T and DVR values. DVRs were calculated by dividing V_T in grey-matter with V_T in the whole-brain, to replicate the previous findings from Bloomfield et al. (*93*).

study found the opposite pattern. FEP patients showed significantly lower levels of $[^{11}\text{C}]$PBR28 V_T in grey-matter compared to controls. Despite a previously reported patient-control difference in grey-matter DVR of large magnitude (*93*), this study found no difference in DVR between the diagnostic groups. Although the patients in the previous study were medicated with anti-psychotics, it is unlikely that the lack of an observed effect in the current study was due to medication-status since drug-naive high-risk individuals also differed from controls in grey-matter DVR (*93*). However, looking at Figure 3.6 it becomes clear that all DVR values are very close to one, and that the variance between subjects has almost entirely been eliminated (DVR mean of whole sample $= 1.05 \pm 0.036$SD v.s. V_T mean of whole sample $= 3.2 \pm 1.59$SD). When examining the effect sizes reported in the revised supplementary information of the previously published study (*93*), the difference in V_T between patients and controls seemed to be mostly confined to the denominator region and not the target region.

In light of these concerns, and in order to aid interpretation of the contradictory DVR findings from Study IV and Bloomfield et al. (*93*), we set out to evaluate the

reliability, precision and validity of [11C]PBR28 DVR values.

3.6 Study V

The objective of Study V was to evaluate the test-retest reliability and convergent validity of [11C]PBR28 ratio-outcomes. Specifically, the DVR and SUVR outcome measures were investigated by re-analyzing an already published test-retest [11C]PBR28 dataset (*87*). Twelve healthy subjects (mean age = 23.9 $\pm$ 2.99SD, 6 males, 6 females, 6 HABs, 6 MABs) had been examined twice with [11C]PBR28. The examinations took place either on the same day (n = 6) or with 2-5 days apart (n = 6). Due to technical reasons, one individual had a shorter follow-up examination. This subject was excluded from the test-retest analysis but the first examination was included in all correlational analyses.

The cerebellum and whole-brain were used in the denominator of the ratio-outcomes, as both have been suggested to be suitable pseudo-reference regions for [11C]PBR28 (*91,93*). The frontal cortex was selected as the primary target ROI in this study and used in the numerator of the ratios. Frontal cortex was chosen instead of the whole grey-matter in order to have at least one ratio-outcome (frontal-cortex/cerebellum) that was not based on nested regions.

Figure 3.7 shows the interregional correlations for V_T and SUVs. Frontal cortex, cerebellum and whole-brain V_T were highly correlated (all Pearson's r > 0.98) for both HAB and MAB subjects. A similar pattern was observed for SUVs, calculated from 40-60 minutes (all Pearson's r > 0.95).

Table 3.3 displays the test-retest metrics for all outcome measures evaluated in Study V. Both frontal cortex V_T and SUVs showed high reliability as estimated using ICC. DVR showed poor reliability, with ~46-48% of the variance estimated as being attributable to error. SUVR showed medium to high reliability. Both V_T and SUV showed higher variability, as compared to the ratio-based outcome measures. There were no correlations between Frontal Cortex V_T and SUVR or DVRs (Figure 3.8), except for SUVR$_{\text{WB}}$. However, examining at HABs and MABs separately revealed that the relationship between SUVR$_{\text{WB}}$ and V_T was confined only to MAB subjects (HAB R^2 = 0.01; MAB R^2 = = 0.33). V_T and SUV were however correlated to each other (HAB R^2 = 0.64; MAB R^2 = 0.86, Figure 3.8).

[11C]PBR28 DVRs, calculated using either cerebellum or whole-brain V_T in the
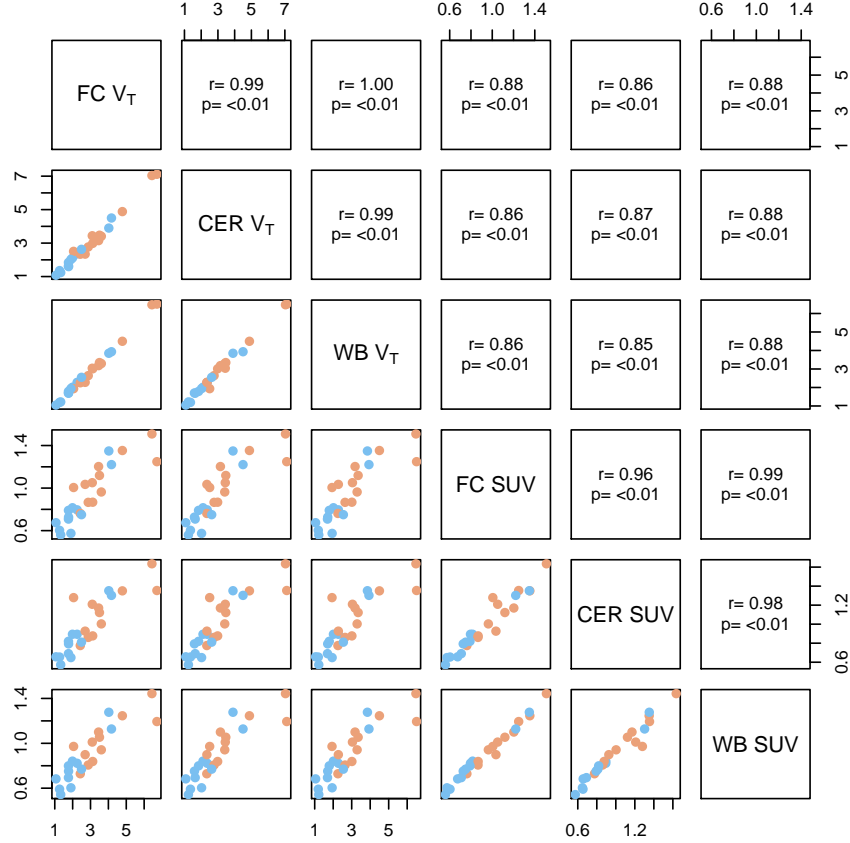
Figure 3.7: Interregional correlations of [^{11}C]PBR28 V_T and SUV in frontal cortex (FC), cerebellum (CER) and whole-brain (WB). HABs are depicted in orange and MABs in blue.

denominator, showed poor reliability and no convergent validity to V_T or SUVs. SUVRs showed an apparent moderate to high reliability, but also lacked association to SUV or V_T. The most concerning finding for [^{11}C]PBR28 ratio-based outcomes is the high interregional correlations for V_T and SUVs. Creating ratios out of outcomes that correlate r>0.95 likely mean that little to no biological signal remains after the division. Instead, the ensuing outcomes might reflect little else but measurement error. In addition, due to these high interregional associations for V_T and SUVs, the inter-individual variance is almost entirely eliminated after division, and all subject's values are set to be close to one. This partly explains the low variability observed for the ratio-outcomes. However, this also means that small changes specific to the numerator or the denominator that are induced by methodological artifacts (such as partial volume effects or head movement) can lead to very large differences in the total outcome between individuals or groups.

Table 3.3: Mean values, variability and test-retest metrics for V_T, SUV, SUVR and DVR using the frontal cortex as the target region and the cerebellum (CER) or whole-brain (WB) as the denominator region for the ratio-based outcomes.

Genotype	Denominator	Mean	SD	ICC	AbsVar	SEM
V_T from 2TCM						
HAB	-	3.90	1.64	0.89	21.0	14.0
MAB	-	2.20	1.01	0.93	17.0	12.0
All	CER	0.98	0.07	0.54	4.7	4.9
All	WB	1.00	0.04	0.52	3.0	2.5
$SUV\ 40-60min$						
HAB	-	1.10	0.24	0.76	13.0	11.0
MAB	-	0.80	0.25	0.91	13.0	9.1
All	CER	0.95	0.06	0.63	4.4	3.8
All	WB	1.00	0.04	0.89	1.5	1.3

The findings of this study suggest that strong caution is warranted when interpreting results from clinical [^{11}C]PBR28 studies making use of ratio-based outcome measures. This caution should especially be applied to studies including healthy control subjects, or a patient group where there is little evidence of the pseudo-region being spared of pathology that affect TSPO levels.

3.7 Study VI

Study V found that [^{11}C]PBR28 DVR is unreliable and likely to reflect little to no biological signal, suggesting that the previous finding of higher DVR in schizophrenia (*93*) likely contained limited evidential value. The results from Study IV instead suggested that patients with psychosis have lower levels of TSPO in the brain compared to healthy controls, contrary to the hypothesis of an elevated microglia activation in schizophrenia. During the course of Study IV, a set of studies was published that also examined microglia activation in psychosis or schizophrenia using second generation TSPO radioligands (*97,99,115*). In line with the results from Study IV and Bloomfield et al. (*93*), none of these studies found the expected elevation of brain TSPO levels in patients. However, all published studies, including Study IV, used small sample sizes (all patient consisted of n < 20). The power to detect a medium-sized difference (e.g. Cohen's D $\approx$ 0.5 with an alpha = 0.05) in V_T between patients and controls ranged from 23% to 34% in previous designs (*93,97,99,115,147*). This level of power severely limits any conclusion that can be
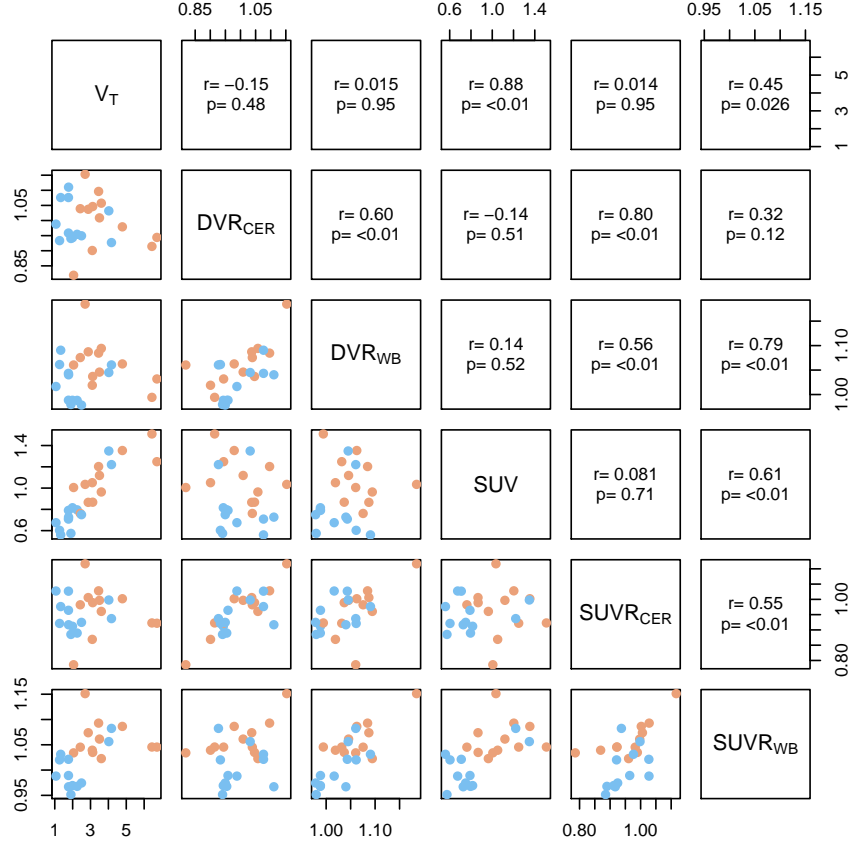
Figure 3.8: Associations between frontal cortex [11C]PBR28 V_T, SUV and ratio-based outcomes, using the whole-brain and cerebellum as denominator regions. HABs are depicted in orange and MABs in blue.

drawn these previous studies.

Hence, Study VI set out to perform an IPD meta-analysis (*162*) on all second generation TSPO studies on psychosis or schizophrenia. The goal was to synthesize all previously published data, and examine if patients showed no difference (H0), higher (H1) or lower (H2) TSPO levels in the brain as compared to healthy controls.

A literature search on PubMed was performed to ascertain that all published second generation TSPO studies on psychosis or schizophrenia were included. Inclusion criteria for studies were that they 1) used a second-generation TSPO radioligand, 2) reported brain V_T values in the in subjects with psychosis or schizophrenia as compared to healthy controls, and 3) reported TSPO affinity type of all participants. At the time of the search (June 2017) five published articles fulfilled these criteria. The corresponding authors of all articles were contacted and all agreed to share

the V_T values, age, gender and TSPO genotype information, clinical ratings and medication status for all individuals included in the original articles. The frontal cortex, temporal cortex and hippocampus were selected as ROIs, as the majority of articles reported V_T values from these three regions. For Bloomfield et al. (*93*), that did not report results from these regions, V_T values derived from the 2TCM were obtained and used in this meta-analysis.

The final sample amounted to 75 patients with schizophrenia or a psychotic disorder (mean age = 33.88 ± 12.57SD; 52 HABs; 23 MABs) and 77 healthy controls (mean age = 35.42 ± 15.12SD; 56 HABs; 21 MABs). To account for range and magnitude differences in V_T among the different radioligands used across studies, all ROI V_T values were standardized (z-scored) within each genotype group of each study (Figure 3.9).

Four different Bayesian linear mixed-effects models examining patient-control differences in V_T (ΔV_T) were evaluated to account for the hierarchical structure of data (Figure 3.10). In model 1 (M1) standardized ROI V_T was specified as the dependent variable, diagnostic group as the fixed effect, and genotype and study as random effects allowing the intercepts to vary. M2 was identical to M1 but added varying slopes of the random effect of genotype (i.e., allowing for differences in ΔV_T between HAB and MAB subjects). M3 was also identical to M1 but added varying slopes of the random effect of PET center (i.e., allowing for differences in ΔV_T between included study samples). M4 allowed for varying slopes of both genotype and PET center.

M1 showed the best fit to data, as determined by the leave-one-out cross-validation fit procedure. Converting the fits of all models to Akaike Weights revealed that M1 showed only slightly better out-of-sample deviance compared to M2 and M3. M1 was therefore used to calculate BFs to evaluate the support for H0, H1 and H2 in data. H1 was specified as a folded-normal distribution centered at zero with an SD of 0.5, expecting higher V_T in patients. H2 was similarly specified but expecting lower V_T in patients. H0 was specified as the point null. M3, allowing the random slopes of PET center to vary, was used to evaluate the between-sample heterogeneity and produce forest-plots of ΔV_T (Figure 3.11). In M3, the prior over ΔV_T was specified as a weakly regularizing normal distribution (mean = 0, SD = 10).

In all regions, BFs showed strong evidence in favor of H2 (patients having lower TSPO levels) relative to H1 (patients having higher TSPO levels). BFs also indicated strong evidence in favor of H2 compared to H0 (no group difference in TSPO levels).
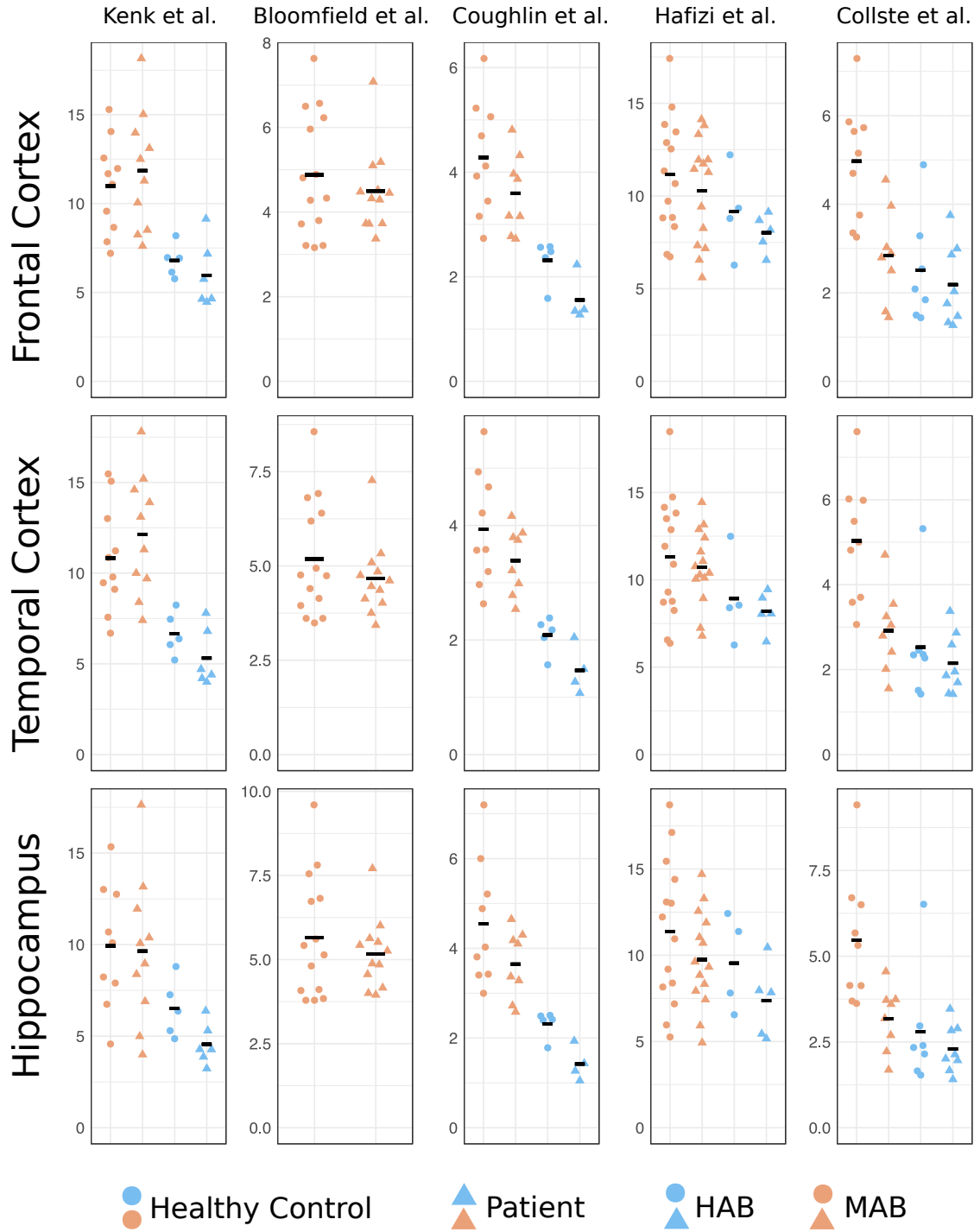
Figure 3.9: All IPD V_T data included in the meta-analysis. All V_T values have been z-scored within study-sample and genotype to produce the pooled plots to the right (the means of HAB and MAB subjects have hence been set to zero).
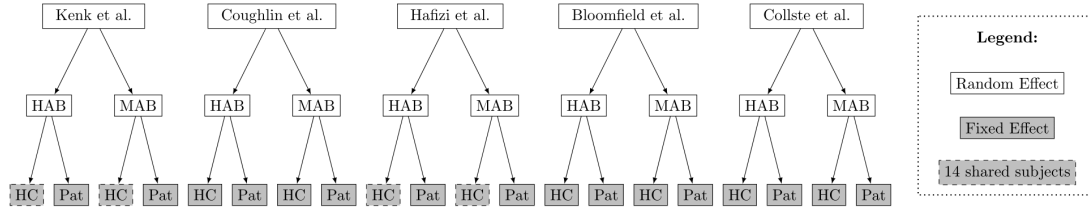
Figure 3.10: Schematic show the included studies and hierarchical structure of data. Bloomfield et al. (*93*) only included two MAB patients but no MAB control subjects. These two subjects were excluded from the analysis since Z-scoring within genotype group was not meaningful. The studies by Kenk et al. (*97*) and Hafizi et al. (*99*) shared 14 healthy control subjects. These subjects were allocated to either Kenk et al. (*97*) or Hafizi et al. (*99*) in this IPD meta-analysis.

Hence, there was >422 times more support in data for the hypothesis of lower levels of TSPO in patients as compared to higher levels. A robustness check of BFs revealed that strong support in favor of H2 was maintained when varying the widths (SD = 0.2 and 0.8) of the prior distributions over H1 and H2.

The forest plot (Figure 3.11) revealed similar results. The overall patient-control differences in standardized V_T values were centered around -0.48 for frontal cortex, -0.47 for temporal cortex and -0.64 for hippocampus, roughly corresponding to medium-sized differences (*163*). The credible intervals did however show considerable uncertainty around these point-estimates (Figure 3.11). Despite this, the posteriors indicated that it is still unlikely that patients have higher levels of TSPO compared to controls in any of the brain regions. The probability of e.g. temporal cortex ΔV_T being above zero (i.e. patients > controls) is only 0.013, conditioned on the data and the model. M3 also showed low between-study heterogeneity (posterior mean = 0.23, 95% credible interval [0, 0.60]).

The results of Study VI suggest that there is strong evidence in favor of patients with psychosis or schizophrenia having lower TSPO levels throughout the brain. Assuming that TSPO binding is a suitable index of glial cell activity, this suggest that patients have lower density, or altered function, of microglia in cortical and subcortical regions.
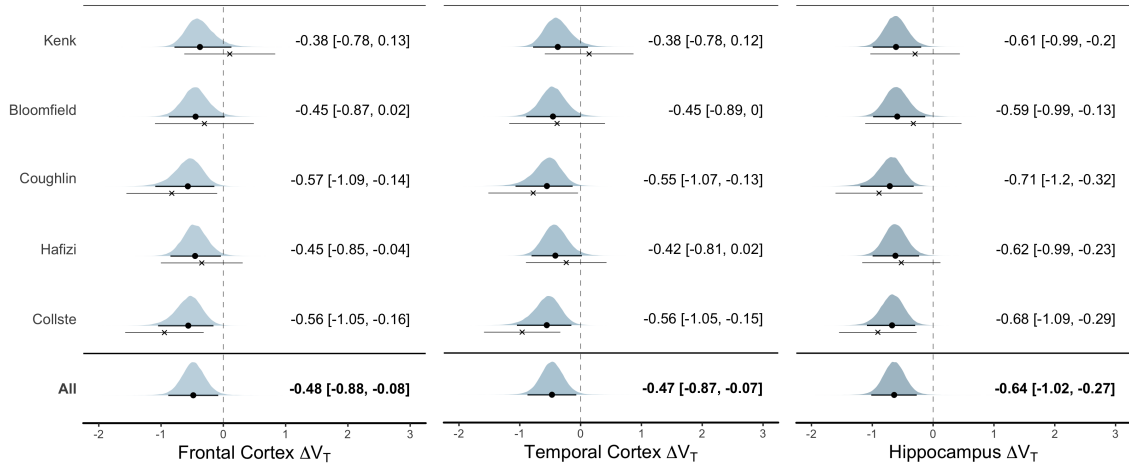
	Frontal Cortex ΔV_T	Temporal Cortex ΔV_T	Hippocampus ΔV_T
Kenk	-0.38 [-0.78, 0.13]	-0.38 [-0.78, 0.12]	-0.61 [-0.99, -0.2]
Bloomfield	-0.45 [-0.87, 0.02]	-0.45 [-0.89, 0]	-0.59 [-0.99, -0.13]
Coughlin	-0.57 [-1.09, -0.14]	-0.55 [-1.07, -0.13]	-0.71 [-1.2, -0.32]
Hafizi	-0.45 [-0.85, -0.04]	-0.42 [-0.81, 0.02]	-0.62 [-0.99, -0.23]
Collste	-0.56 [-1.05, -0.16]	-0.56 [-1.05, -0.15]	-0.68 [-1.09, -0.29]
All	**-0.48 [-0.88, -0.08]**	**-0.47 [-0.87, -0.07]**	**-0.64 [-1.02, -0.27]**

Figure 3.11: Standardized difference in TSPO levels (estimated using V_T) between patients with psychosis or schizophrenia and healthy control subjects. The posterior distribution for each study-specific difference in V_T estimate (random slopes) from the Bayesian linear mixed model (M3) is presented. The black circle denotes the posterior mean, and the thick line denotes the 95% credible interval; these are also presented in text next to the plots. The cross denotes the patient–control difference in raw data (together with its 95% credible interval) without performing linear mixed-effects modeling. Hence, the difference between the dot and the cross show the model shrinkage toward the mean. The overall effect size suggest that patients show lower TSPO levels in all three brain regions.

3.8   Study VII

Study VI only included studies using second-generation TSPO radioligands. A set of studies have however examined patients with psychotic disorders using the first generation TSPO radioligand (R)-[^{11}C]PK11195 (*77–81*). In these studies, a wide range of different outcome measures were used, such as V_S or BP_{ND} derived from 2TCM, BP_{ND} derived using SVCA or BP_{ND} derived using cerebellum as pseudo-reference region. The reason for excluding these articles from Study VI was as follows: in a meta-analysis model, all sample-specific effect sizes have to come from the same underlying distribution of effects. Since very different quantification methods and outcome measures had been used, it was unlikely that the effect from these previous (R)-[^{11}C]PK11195 studies could be synthesized into the same model. However, peer reviewers of Study VI pointed out that BP_{ND} estimates of (R)-[^{11}C]PK11195 data were both precise and showed high sensitivity to detect patient-control differences, and that the meta-analysis hence omitted an important part of the literature. It was difficult to assess the credibility of such claims, as there is to date no thorough evaluation of reliability, sensitivity and validity of
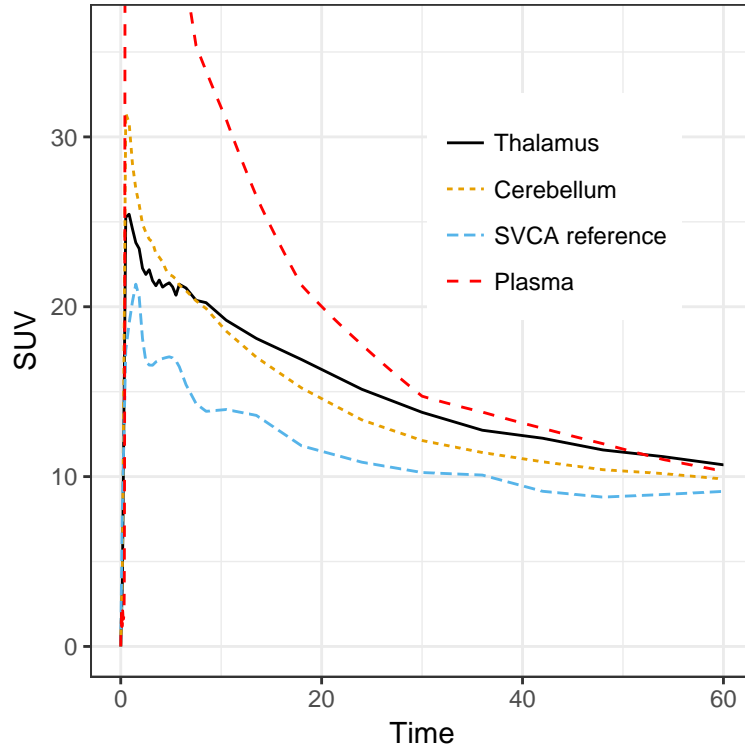
Figure 3.12: Average (R)-[^{11}C]PK11195 thalamus, cerebellum and SVCA4-reference TACs. The metabolite corrected plasma curve (truncated at SUV = 37) is also shown. All TACs are expressed in SUV units to allow for averaging.

(R)-[^{11}C]PK11195 outcome measures used in previous psychosis studies.

Study VII hence set out to investigate the test-retest properties and convergent validity of different (R)-[^{11}C]PK11195 outcome measures. Specifically, the reliability of BP_{ND} derived from the 1) 2TCM, 2) from using a SVCA reference TAC, and 3) from a pseudo-reference region (cerebellum) was evaluated (Figure 3.12). Study VII also aimed to correlate these BP_{ND} estimates with V_T and V_S derived from the 2TCM using an AIF.

Data from six healthy subjects (mean age = 25.8 ± 3.9SD, all males) that participated in an already published test-retest study (*85*) was re-analyzed. All subjects performed two PET examinations, approximately 6 weeks apart. Frontal cortex, whole grey-matter, hippocampus, striatum and thalamus were selected as ROIs. V_T, V_S and BP_{ND} were calculated from rate constants extracted from the 2TCM. BP_{ND} values were also derived using the SRTM with cerebellum as reference region, or by using a SVCA4-derived reference TAC.

The results showed that BP_{ND} values derived using cerebellum or a SVCA4 reference
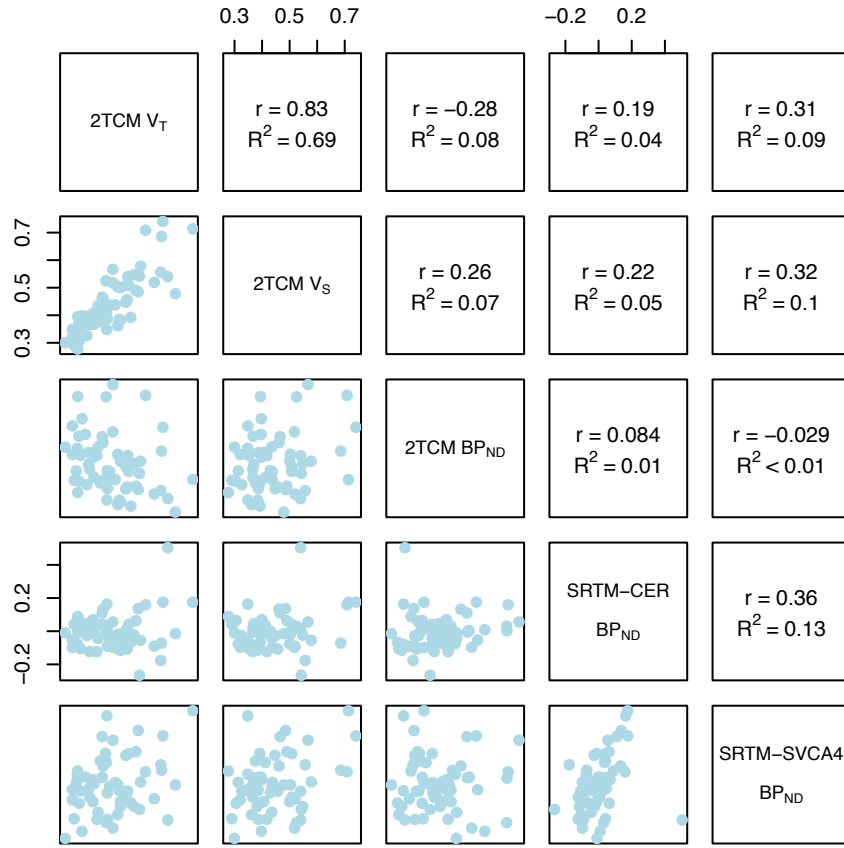
Figure 3.13: Correlations between the different (R)-[^{11}C]PK11195 outcome measures. Outcomes derived from using AIF were not correlated to outcomes derived without AIF. BP_{ND} from the 2TCM was not correlated to any other outcome. All BP_{ND} hence showed low convergent validity towards V_T or V_S from the 2TCM with AIF.

TAC were close to zero, or even negative (Table 3.4). The overall regional reliability, as estimated using ICC and AbsVar, was poor for all measures of BP_{ND} derived with or without an AIF. The precision, as estimated using SEM and MD, was also low. V_T and V_S from the 2TCM did however show moderate reliability (Table 3.4).

Figure 3.13 shows the correlations between the evaluated outcome measures. None of the BP_{ND} outcomes were correlated to V_T or V_S from the 2TCM (all $R^2 < 0.11$). V_T and V_S were correlated to each other, as were BP_{ND} using cerebellum and BP_{ND} using a SVCA4 derived TAC as reference input.

The results from Study VII suggest that (R)-[^{11}C]PK11195 BP_{ND}, estimated with or without AIF, showed low reliability and poor precision. BP_{ND} was not correlated to V_T or V_S, suggesting no to low convergent validity. Hence, if V_T or V_S, at least in part, reflects true binding to target, it is unlikely that BP_{ND} estimates does so as well. Rather, the results suggest that BP_{ND} estimates are so unstable and imprecise that they might represent little more than noise. Previous (R)-[^{11}C]PK11195 BP_{ND}

Table 3.4: Mean values all subjects (R)-[^{11}C]PK11195 PET examinations. The rest-retest reliability, repeatability and precision were estimated using the Intra-Class Correlation Coefficient (ICC), average absolute variability in percentage (AbsVar) and standard error of measurement (SEM), of different outcome measures derived with or without AIF. The minimum detectable difference (MD) denotes the difference (expressed as a percentage of the mean) needed between two measurements for them to be significantly different from each other.

Region	Mean	SD	ICC	AbsVar	SEM	MD
V_T (2TCM)						
FC	0.72	0.16	0.73	15	0.08	32
GM	0.70	0.17	0.78	15	0.08	31
HIP	0.72	0.19	0.66	21	0.11	44
STR	0.76	0.17	0.44	18	0.13	46
THAL	0.77	0.22	0.69	21	0.12	43
V_S (2TCM)						
FC	0.42	0.09	0.68	14	0.05	32
GM	0.42	0.09	0.67	15	0.05	34
HIP	0.45	0.10	0.35	21	0.08	51
STR	0.44	0.10	0.23	23	0.09	58
THAL	0.48	0.14	0.91	13	0.04	24
BP_{ND} (2TCM)						
FC	1.49	0.33	0.65	18	0.20	37
GM	1.62	0.40	0.31	29	0.33	56
HIP	2.02	0.77	-0.19	50	0.84	115
STR	1.41	0.39	0.32	22	0.32	63
THAL	1.79	0.67	-0.11	39	0.71	110
BP_{ND} (SRTM-SVCA4)						
FC	0.17	0.04	0.21	29	0.04	63
GM	0.21	0.06	0.34	27	0.05	59
HIP	0.17	0.09	-0.39	83	0.10	160
STR	0.21	0.09	-0.12	59	0.09	120
THAL	0.35	0.09	0.32	22	0.07	55
BP_{ND} (SRTM-CER)						
FC	-0.07	0.09	0.50	160	0.06	258
GM	-0.03	0.06	0.51	277	0.04	444
HIP	0.01	0.08	0.19	181	0.07	1920
STR	-0.02	0.17	-0.14	196	0.18	2963
THAL	0.09	0.06	0.67	494	0.04	112

studies making use of BP_{ND} values derived from 2TCM, SVCA4 or the cerebellum as pseudo-reference region should be interpreted with caution, at least when they include healthy control subjects as a comparison group.

3.9 Study VIII

Study VII showed that direct estimates of specific (R)-[^{11}C]PK11195 radioligand binding (such as BP_{ND}) were unreliable and unlikely to be useful in clinical studies. With regards to [^{11}C]PBR28, BP_{ND} from the 2TCM has previously been shown to be unreliable (87). For these reasons, V_T was used to estimate TSPO levels in the brain of patients with psychosis and control subjects in Study VI. V_T does however contain both specific (V_S) and non-displaceable signal (V_{ND}) and is therefore not a direct estimate of specific binding. The conclusions of Study VI relies on a critical assumption: that the patient and control groups did not differ in V_{ND}. This could not be directly tested, since V_T does not differentiate between V_S and V_{ND}. If reliable and valid estimates of [^{11}C]PBR28 specific binding could be obtained, it would allow for bettered powered and more precise detection of clinical effects.

The SIME method aims to derive estimates of specific binding (such as V_S or BP_{ND}) without requiring a reference region. The objective of Study VIII was to evaluate the accuracy and reliability of SIME applied to [^{11}C]PBR28 data. Specifically, Study VIII set out to 1) simulate data and examine how accurately SIME could estimate a known V_{ND} value, 2) compare SIME to "gold standard" estimates of V_{ND} using pharmacological challenge data 4) compare SIME derived V_{ND}, V_S and BP_{ND} values between HAB and MAB subjects and 5) examine the test-retest reliability of SIME derived outcomes.

First, realistic looking [^{11}C]PBR28 TACs were simulated with a known V_{ND} value. This was done as follows: ROI TACs from KI [^{11}C]PBR28 database (mean age = 45.2 ± 17.0SD; 30 HABs and 24 MABs; 32 males and 22 females) were fitted using the 2TCM. The residuals of the fits were saved into a library. Residuals from this library were then sampled and added to a set of noise-free model curves with a known V_{ND}. The sampling was done in such way that the variance-covariance structure between and within TACs were maintained (see article by Schain et al. (164) for more details). The goal of this procedure is to obtain a more "realistic" looking set of simulated [^{11}C]PBR28 ROI TACs, compared to only adding Gaussian noise to model curves. Using noise-free model curves from a HAB and a MAB subjects,

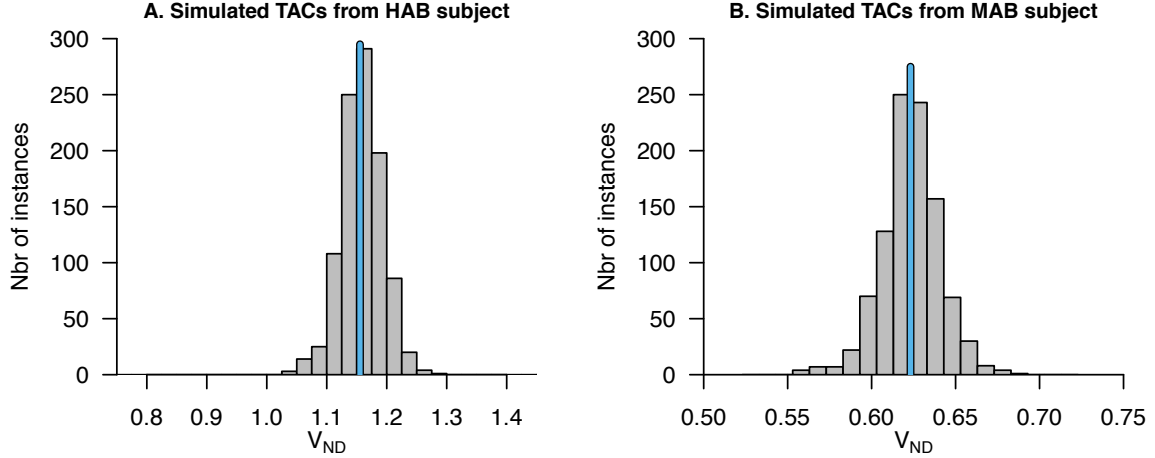Figure 3.14: In each plot, 1000 [^{11}C]PBR28 noise instances have been created and added onto a set of noise-free model curves obtained from a HAB subject (A) or a MAB subject (B). SIME has then been applied to each instance to estimate V_{ND} (histogram). The "true" underlying V_{ND} is shown be the blue vertical line.

the procedure was repeated 1000 times (instances). The results of the simulations can be thought of as a separate HAB and MAB subject, examined 1000 times with [^{11}C]PBR28, with a set of ROI TACs for each examination that have a known and common V_{ND}.

SIME was then applied to each simulated instance. The results show that SIME could estimate the "true" underlying V_{ND} value with high accuracy and precision (Figure 3.14; HAB $V_{ND:True}$ = 1.15, mean $V_{ND:SIME}$ = 1.17 ± 0.035SD; MAB $V_{ND:True}$ = 0.62, mean $V_{ND:SIME}$ = 0.63 ± 0.018SD).

Next, five healthy HAB subjects (mean age = 25.2 ± 7.3SD; all male) underwent a pharmacological challenge. All subjects performed a baseline [^{11}C]PBR28 followed by intake of the TSPO receptor agonist XBD173 (10 to 90mg) and a repeat PET examination. Using the baseline and blocking scans, the Lassen plot (*143*) and Likelihood Estimation of Occupancy (*144*) methods were used to estimate V_{ND}. SIME was then applied to only the baseline scans. Figure 3.15 show the results from all three methods. SIME underestimated V_{ND} by 19% compared to the Lassen plot, but showed similar values to the Likelihood Estimation of Occupancy method (-3%).

The differences in SIME derived V_{ND}, grey-matter V_S and BP_{ND} were compared between HAB and MAB subjects from the KI [^{11}C]PBR28 database. Unexpectedly, HABs showed higher average V_{ND} compared to MABs (Figure 3.16A, Hedges' g = 0.82). In order to investigate potential reasons for this difference in V_{ND}, the AUC, peaks and shapes of the AIF of HABs and MABs were compared. There were clear
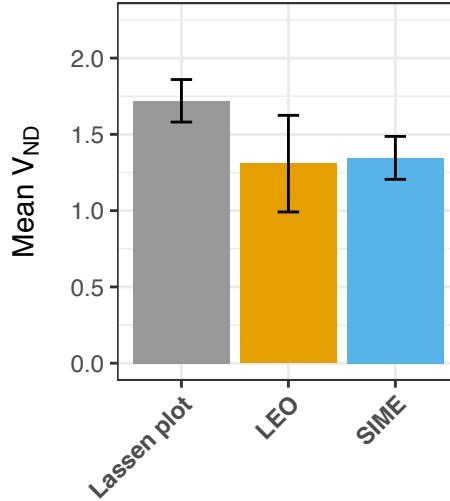
Figure 3.15: V_{ND} estimates from 5 HAB subjects undergoing a XBD173 blocking challenge, using the Lassen plot, Likelihood Estimation of Occupancy (LEO) and SIME. SIME was only performed on the baseline scans.

differences in AUC and peak-height between average HAB and MAB AIFs, with MABs showing higher values (data not shown). This is to be expected since HABs have much higher binding in the periphery. There was however also a difference in the shape of the AIF between genotype groups, with HABs showing a steeper post-peak decay (Figure 3.16B). When using an average AIF for all subjects, the difference in SIME derived V_{ND} observed in Figure 3.16A disappeared (Figure 3.16C, Hedges' g = -0.029). Assuming that SIME is valid and unbiased, these results suggest two possible explanations for the observed genotype difference in V_{ND}: 1) There is an actual biological difference in the non-specific binding or free-fraction of radioligand between HABs and MABs (due to e.g. difference in active transportation across the blood-brain barrier); 2) there is a systematic error in the measurement of the AIF for HABs and/or MABs that affects the estimated V_{ND}. To date, there exists no published [^{11}C]PBR28 blocking data examining V_{ND} in MAB subjects. Hence, the observed difference in V_{ND} between genotypes cannot be fully verified, and this phenomenon warrants further investigation.

V_T from the unconstrained 2TCM and SIME derived V_S showed large separation in the whole grey-matter between genotype groups, while to SIME derived BP_{ND} values showed lower separation (Figure 3.16D, E and F). Importantly, the average SIME V_S for HABs (mean = 2.69) was almost exactly double that of MABs (mean = 1.36). If the SIME method is valid, this is to be expected since the low-affinity allele has negligible binding of [^{11}C]PBR28 to TSPO, so that HAB subjects will effectively show twice as many binding sites compared to MAB subjects (*89*).

Table 3.5: Means, variability and test-retest metrics for grey-matter V_T from the unconstrained 2TCM and SIME derived outcome measures (V_{ND}, V_S and BP_{ND}).

Measure	PET1 Mean (SD)	PET2 Mean (SD)	ICC	AbsVar	SEM
2TCM V_T	3.41 (1.86)	3.65 (1.84)	0.94	17.54	0.43
SIME V_{ND}	1.29 (0.47)	1.35 (0.47)	0.86	18.07	0.18
SIME V_S	2.12 (1.55)	2.29 (1.47)	0.93	24.29	0.39
SIME BP_{ND}	1.61 (0.76)	1.63 (0.6)	0.65	24.00	0.39

The test-retest reliability and precision of SIME derived outcomes can be seen in Table 3.5. V_S showed excellent reliability, as estimated using the ICC, while BP_{ND} showed lower reliability.

The results from Study VIII suggest that SIME could accurately estimate V_{ND} for [^{11}C]PBR28 examination. V_S derived using SIME showed both large separation between genotype groups, and high reliability. SIME derived BP_{ND} values were however not as reliable and precise. One potential explanation for this is that small amounts of measurement error in the SIME derived numerator ($V_{S:SIME}$) and/or denominator ($V_{ND:SIME}$) used to calculate BP_{ND} leads to an amplified and larger error in the quotient, while this is not the case for subtraction carried out to calculate only V_S ($V_{T:2TCM}$ - $V_{S:SIME}$). In theory, V_S should show higher sensitivity and power for detecting clinical effects compared to V_T. Future clinical [^{11}C]PBR28 studies should therefore use SIME derived V_S values in preference, or in addition, to V_T from the unconstrained 2TCM.
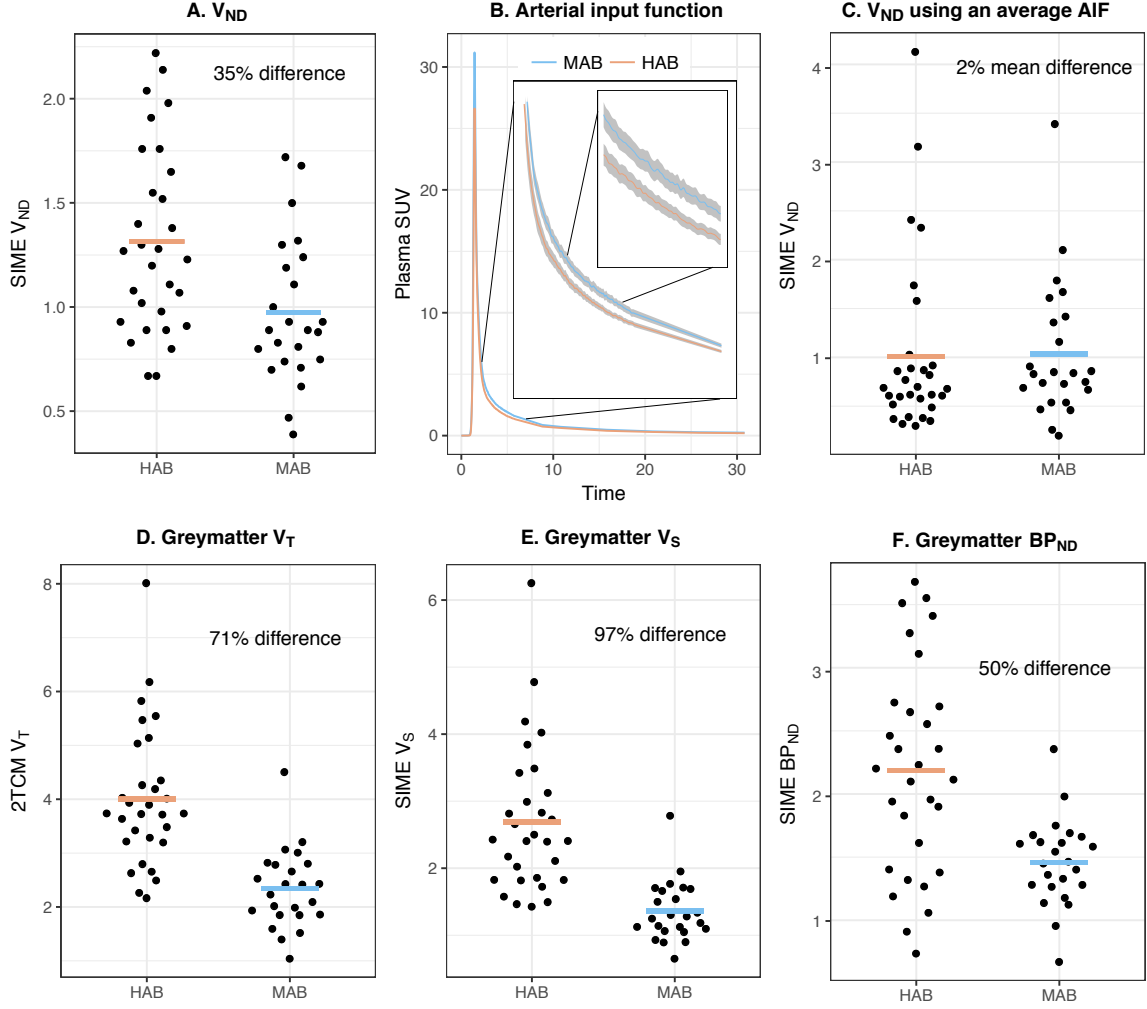
Figure 3.16: Difference between genotype groups in A) SIME derived V_{ND}, B) shape of AIFs, C) SIME derived V_{ND} when using an average AIF, D) grey-matter V_T from the unconstrained 2TCM grey-matter, and SIME derived E) V_S and F) BP_{ND} values. When an average AIF was used for all subjects, the genotype difference in V_{ND} disappeared (C). Shaded regions in B denotes ± 1SE around the average AIFs.

Chapter 4

Future perspectives

- The negative relationship between Social Desirability and D2-R availability in striatum is one of the most replicated findings in PET-personality literature. However, few to no PET-personality studies have been pre-registered, sample sizes are often very small and it is common that many different regions (even within the striatum) are examined with insufficient correction for multiple comparisons. It would therefore be of importance to perform a meta-analysis of the relationship between Social Desirability scale scores and striatal D2-R availability. This should preferable be done using both published and unpublished data. The latter could be obtained by e.g. contacting PET centers around the world and ask them to share potential unpublished data. In doing so, a clearer assessment of the robustness of the effect could be obtained and published.

- The SIME method showed promise in estimating [^{11}C]PBR28 V_S. A future study could hence apply SIME on patient-control data, to yield specific binding estimates of [^{11}C]PBR28. The hypothesis is that, regardless of the direction of the results, SIME V_S should show higher sensitivity compared to V_T. This would then translate into SIME outperforming the e.g. unconstrained 2TCM when it comes to quantifying support for, or against, an effect.

- Binding outcome measures are often treated as point-estimated values in PET literature. I.e. there are few methods or studies that takes the inherent uncertainty around e.g. V_T or BP_{ND} into account when performing statistical inference. Future research could focus on developing and evaluate methods where the uncertainty in the kinetic modeling is estimated, and accounted for

in the ensuing statistical inference. A question of interest would be how such an approach would change the statistical power of analyses and the assessment of certainty around effects, compared to the way kinetic and statistical modeling of PET data is carried out today.

- It will likely become more important to advocate for, and perform 1) pre-registered studies or "registered reports" (*165*), 2) open-data and code sharing as well as 3) practices leading to reproducible workflows and results, when doing research in PET. Such objectives would allow for more robust and trustworthy findings and likely limit questionable research practices, such as "p-hacking". In order to ease data sharing and reproducibility, one goal would be to create publicly available open-source software that can process data all the way from a standardized "raw" format (e.g. reconstructed dynamic PET images organized by BIDS structure, (*166*)) to outcome measures of interest (e.g. BP_{ND} values) in the same pipeline.

Acknowledgments

I would like to thank Simon Cervenka for having been an excellent supervisor and mentor. The freedom and responsibility you have entrusted me with during these years are the most important reasons to why I've been able to develop skills for doing science, while also finding it highly rewarding and meaningful. Your loyalty, patience and leadership have allowed me to always and only focus on science and nothing else. This is probably the greatest gift a supervisor can give a PhD student.

Granville Matheson – thank you for being my closest friend and companion in almost all my, more or less successful, scientific endeavors. As the arbiter of my scientific reasoning and works, my PhD (and everyday life) would have been significantly less fruitful and fun without you by my side.

Thanks to my co-supervisor Petter Gustavsson for fueling my interest in science, statistics and music, and to Lars Farde for your wisdom, encouragement and always saying "yes, go for it" when I asked for data to try out new ideas.

Thank you, Martin Schain, for the scientific and non-scientific guidance you have provided me with as a co-supervisor and as a friend. Few people have the questionable honor of being a role (playing?) model to me, but you are one of them.

I would like to thank everybody in room 1300.5: Emma Veldman, Patricimo Fazio, Max Andersson, Vera Kerstens, Ämma Tangen, and Jonas Svensson for being awesome friends and co-workers and for stoically withstanding mine and Granville's loud fights over many, many trivial scientific matters. You are the reason why I always enjoyed going to work in the morning, even during the tougher times of my PhD. Thanks to Miklós Toth, Lenke Tari and Anton Forsberg for all the fun discussions, for the few but intensive online play sessions, and for enriching my everyday work experience. Without you all, lunch would not have been near as fun.

Lieke, Nina, William and Björn – thank you for your friendship and for pushing me to become a better scientist. Spending time with you in the journal club, seminar-series,

geek-club or the pub have been among the best parts of my PhD. Remember that you are always allowed a couple of mulligans (approximately 3-4) in the beginning of your PhD.

I would like to thank Karin Collste and Pauliina Victorsson for always taking the time to answer basic and complex questions about psychiatry, sharing your clinical expertise while also putting my statistics teaching skills to the test.

Zsolt Cselényi, Katarina Varnäs and Andrea Varrone – thank you for the guidance, experience and technical expertise you have provided me. A big thanks to Urban Hansson and Göran Rosenqvist for patiently explaining bash code and the physics of the PET systems (respectively), while always providing technical and logistical support on very short notice. A big thanks to Karin Zahir and Nina Knave – without your help I would have stumbled into the pit of administrational issues, and likely never gotten up. Thanks to Eva Holmgaard for standing watch at the bottom of the pit, feeding my fear of falling in.

A very big thanks to current and former members of the PET group: Christer Halldin, Balazs Gulyas, Per Stenkrona, Magdalena Nord, Aurelija Jučaite, Kai-Chun Yang, Mikael Tiger, Patrik Mattson, Johan Lundberg, Ryosuke Arakawa, Jacqueline Borg, Sjoerd Finnema, Akihiro Takano, Mahabuba Jahan, Carsten Steiger, Patricia Miranda-Azpiazu, Pavitra Kannan, Camilla Gustafsson, Rafael Maior, Magnus Schou, Karin Olsson, Peter Johnström, Vladimir Stepanov, Sangram Nag, Arsalan Amir, Julio Gabriel, Sara Lundqvist, Zsolt Sarnyai, Matteo Ferrante, Nadja Hellsing, Marcello Venzi, Malena Kjellén, Jonas Ahlgren, Opokua Britton-Cavaco, Åsa Södergren, Siv Eriksson, Marie Svedberg and Jenny Häggkvist. A very big thanks to everybody in the PET group not listed here - without you all my work would not have been possible.

Thanks to Todd Ogden and Francesca Zanderigo for excellent collaboration, inspiration and supervision. Thanks to Rita Almeida, Erik Hedman, Ann Rudman, Predrag Petrovic, Karin Jensen and Bo Melin for scientific support or inspiration outside the immediate scope of my thesis.

I would like to thank a set of people, most of whom I never met, whose online courses, online presence, summer schools or participation in podcasts have had an especially strong impact on the way I think about and do science: Roger Peng, Hillary Parker, Andrew Conway, Michael Fitzpatrick, Richard McElreath and E.J. Wagenmakers.

References

1. Azevedo FAC, Carvalho LRB, Grinberg LT, et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled up primate brain. *Journal of Comparative Neurology.* 2009;513:532-541.

2. Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA, Hudspeth AJ. Principles of neural science. Vol 4. McGraw-hill New York; 2000.

3. Vallone D, Picetti R, Borrelli E. Structure and function of dopamine receptors. *Neuroscience & Biobehavioral Reviews.* 2000;24:125-132.

4. Malenka RC, Nestler EJ, Hyman SE. Chapter 6: Widely Projecting Systems: Monoamines, Acetylcholine, and Orexin. *Sydor A, Brown RY Molecular Neuropharmacology: A Foundation for Clinical Neuroscience (2nd ed) New York: McGraw-Hill Medical.* 2009:147-148.

5. Calabresi P, Picconi B, Tozzi A, Ghiglieri V, Di Filippo M. Direct and indirect pathways of basal ganglia: a critical reappraisal. *Nature neuroscience.* 2014;17:1022-1030.

6. Nieoullon A. Dopamine and the regulation of cognition and attention. *Progress in neurobiology.* 2002;67:53-83.

7. Wise RA. Dopamine, learning and motivation. *Nature reviews neuroscience.* 2004;5:483-494.

8. Volkow ND, Fowler JS, Wang G-J, Swanson JM. Dopamine in drug abuse and addiction: results from imaging studies and treatment implications. *Molecular psychiatry.* 2004;9:557-569.

9. Zhuang X, Mazzoni P, Kang UJ. The role of neuroplasticity in dopaminergic therapy for Parkinson disease. *Nature Reviews Neurology.* 2013;9:248-256.

10. Nutt DJ, Lingford-Hughes A, Erritzoe D, Stokes PRA. The dopamine theory of

addiction: 40 years of highs and lows. *Nature Reviews Neuroscience.* 2015;16:305-312.

11. Grace AA. Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nature Reviews Neuroscience.* 2016.

12. Herculano-Houzel S. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience.* 2009;3:31.

13. Gehrmann J, Matsumoto Y, Kreutzberg GW. Microglia: intrinsic immuneffector cell of the brain. *Brain Research Reviews.* 1995;20:269-287.

14. Raivich G. Like cops on the beat: the active role of resting microglia. *Trends in neurosciences.* 2005;28:571-573.

15. Hanisch U-K, Kettenmann H. Microglia: active sensor and versatile effector cells in the normal and pathologic brain. *Nature neuroscience.* 2007;10:1387.

16. Garden GA, Möller T. Microglia biology in health and disease. *Journal of Neuroimmune Pharmacology.* 2006;1:127-137.

17. Farde L. TSPO binding may also represent 'resting'microglia. *Clinical and Translational Imaging.* 2015;3:491-492.

18. Pelvig DP, Pakkenberg H, Stark AK, Pakkenberg B. Neocortical glial cell numbers in human brains. *Neurobiology of aging.* 2008;29:1754-1762.

19. Ransom B, Behar T, Nedergaard M. New roles for astrocytes (stars at last). *Trends in neurosciences.* 2003;26:520-522.

20. Graeber MB, Streit WJ. Microglia: biology and pathology. *Acta neuropathologica.* 2010;119:89-105.

21. Hamby ME, Sofroniew MV. Reactive astrocytes as therapeutic targets for CNS disorders. *Neurotherapeutics.* 2010;7:494-506.

22. Halldin C, Stone-Elander S, Farde L, et al. Preparation of 11C-labelled SCH 23390 for the in vivo study of dopamine D-1 receptors using positron emission tomography. *Int J Rad Appl Instrum [A].* 1986;37:1039-1043.

23. Olsson H, Halldin C, Swahn C-GG, Farde L. Quantification of [11C]FLB 457 Binding to Extrastriatal Dopamine Receptors in the Human Brain. *Journal of Cerebral Blood Flow and Metabolism.* 1999;19:1164-1173.

24. Chen M-K, Guilarte TR. Translocator protein 18 kDa (TSPO): molecular sensor

of brain injury and repair. *Pharmacology & therapeutics.* 2008;118:1-17.

25. Moses WW. Fundamental limits of spatial resolution in PET. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment.* 2011;648:S236-S240.

26. Gunn RN, Gunn SR, Cunningham VJ. Positron emission tomography compartmental models. *Journal of Cerebral Blood Flow & Metabolism.* 2001;21:635-652.

27. Lammertsma AA, Bench CJ, Hume SP, et al. Comparison of methods for analysis of clinical [11C] raclopride studies. *Journal of Cerebral Blood Flow & Metabolism.* 1996;16:42-52.

28. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. *Neuroimage.* 1996;4:153-158.

29. Salinas CA, Searle GE, Gunn RN. The simplified reference tissue model: model assumption violations and their impact on binding potential. *Journal of Cerebral Blood Flow & Metabolism.* 2015;35:304-311.

30. Ichise M, Toyama H, Innis RB, Carson RE. Strategies to improve neuroreceptor parameter estimation by linear regression analysis. *Journal of Cerebral Blood Flow & Metabolism.* 2002;22:1271-1281.

31. Ichise M, Liow J-S, Lu J-Q, et al. Linearized Reference Tissue Parametric Imaging Methods: Application to [11C]DASB Positron Emission Tomography Studies of the Serotonin Transporter in Human B rain. *Journal of Cerebral Blood Flow & Metabolism.* 2003;23:1096-1112.

32. Logan J. Graphical analysis of PET data applied to reversible and irreversible tracers. *Nuclear medicine and biology.* 2000;27:661-670.

33. Ogden RT, Tarpey T. Estimation in regression models with externally estimated parameters. *Biostatistics.* 2005;7:115-129.

34. Wiggins JS. A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology.* 1979;37:395-412.

35. Kiesler DJ. Contemporary interpersonal theory and research: Personality, psychopathology, and psychotherapy. NY: John Wiley & Sons; 1996.

36. Pincus AL, Wiggins JS. Interpersonal problems and conceptions of personality

disorders. *Journal of Personality Disorders.* 1990;4:342-352.

37. Huang CL, Yang YK, Chu CL, et al. The association between the Lie scale of the Maudsley personality inventory and striatal dopamine D2/D3 receptor availability of healthy Chinese community subjects. *European Psychiatry: The Journal of the Association of European Psychiatrists.* 2006;21:62-65.

38. Reeves SJ, Mehta MA, Montgomery AJ, et al. Striatal dopamine (D2) receptor availability predicts socially desirable responding. *NeuroImage.* 2007;34:1782-1789.

39. Egerton A, Rees E, Bose SK, et al. Truth, lies or self-deception? Striatal D(2/3) receptor availability predicts individual differences in social conformity. *NeuroImage.* 2010;53:777-781.

40. Cervenka S, Gustavsson P, Halldin C, Farde L. Association between striatal and extrastriatal dopamine D2-receptor binding and social desirability. *Neuroimage.* 2010;50:323-8.

41. Caravaggio F, Fervaha G, Chung JK, et al. Exploring personality traits related to dopamine D2/3 receptor availability in striatal subregions of humans. *European Neuropsychopharmacology.* 2016;26:644-652.

42. Paulhus DL. Two-component models of socially desirable responding. *Journal of personality and social psychology.* 1984;46:598.

43. Couppis MH, Kennedy CH, Stanwood GD. Differences in aggressive behavior and in the mesocorticolimbic DA system between A/J and BALB/cJ mice. *Synapse.* 2008;62:715-724.

44. Zuckerman M, Michael Kuhlman D, Thornquist M, Kiers H. Five (or three) robust questionnaire scale factors of personality without culture. *Personality and Individual Differences.* 1991;12:929-941.

45. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry.* 2005;62:617-627.

46. Clark DM, Wells A. A cognitive model of social phobia. *Social Phobia: Diagnosis, Assessment, and Treatment.* 1995;41:69-93.

47. Bruce LC, Heimberg RG. Social anxiety disorder. *The Wiley Handbook of*

Cognitive Behavioral Therapy. 2014.

48. Morrow B, Elsworth J, Rasmusson A, Roth R. The role of mesoprefrontal dopamine neurons in the acquisition and expression of conditioned fear in the rat. *Neuroscience.* 1999;92:553-564.

49. McCullough LD, Sokolowski JD, Salamone JD. A neurochemical and behavioral investigation of the involvement of nucleus accumbens dopamine in instrumental avoidance. *Neuroscience.* 1993;52:919-925.

50. Schneier FR, Liebowitz MR, Abi-Dargham A, Zea-Ponce Y, Lin SH, Laruelle M. Low dopamine D(2) receptor binding potential in social phobia. *Am J Psychiatry.* 2000;157:457-9.

51. Schneier FR, Abi-Dargham A, Martinez D, et al. Dopamine transporters, D2 receptors, and dopamine release in generalized social anxiety disorder. *Depression and Anxiety.* 2009;26:411-418.

52. Freitas-Ferrari MC, Hallak JEC, Trzesniak C, et al. Neuroimaging in social anxiety disorder: a systematic review of the literature. *Progress in Neuro-Psychopharmacology & Biological Psychiatry.* 2010;34:565-580.

53. Brühl AB, Delsignore A, Komossa K, Weidt S. Neuroimaging in social anxiety disorder—a meta-analytic review resulting in a new neurofunctional model. *Neuroscience & Biobehavioral Reviews.* 2014;47:260-280.

54. American-Psychiatric-Association-Task Force on-DSM-IV. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. American Psychiatric Publishing, Inc.; 2000.

55. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS medicine.* 2005;2:e141.

56. Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014;511:421-427.

57. Arias I, Sorlozano A, Villegas E, et al. Infectious agents associated with schizophrenia: a meta-analysis. *Schizophrenia research.* 2012;136:128-136.

58. Dalman C, Allebeck P, Gunnell D, et al. Infections in the CNS during childhood and the risk of subsequent psychotic illness: a cohort study of more than one million

Swedish subjects. *American Journal of Psychiatry.* 2008;165:59-65.

59. Müller N. The role of anti-inflammatory treatment in psychiatric disorders. *Psychiatria Danubina.* 2013;25:0-298.

60. Miller BJ, Buckley P, Seabolt W, Mellor A, Kirkpatrick B. Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects. *Biological psychiatry.* 2011;70:663-671.

61. Upthegrove R, Manzanares-Teson N, Barnes NM. Cytokine function in medication-naive first episode psychosis: a systematic review and meta-analysis. *Schizophrenia research.* 2014;155:101-108.

62. Mizrahi R. Social stress and psychosis risk: common neurochemical substrates? *Neuropsychopharmacology.* 2015.

63. Howes OD, McCutcheon R. Inflammation and the neural diathesis-stress hypothesis of schizophrenia: a reconceptualization. *Translational psychiatry.* 2017;7:e1024.

64. Bayer TA, Buslei R, Havas L, Falkai P. Evidence for activation of microglia in patients with psychiatric illnesses. *Neuroscience letters.* 1999;271:126-128.

65. Radewicz K, Garey LJ, Gentleman SM, Reynolds R. Increase in HLA-DR immunoreactive microglia in frontal and temporal cortex of chronic schizophrenics. *Journal of Neuropathology & Experimental Neurology.* 2000;59:137-150.

66. Kellom M, Basselin M, Chen M, Rapoport SI, Rao JS. Increased Neuroinflammatory and Arachidonic Acid Cascade Markers with Synaptic Marker Loss in Lipopolysaccharide Infused Rats. *The FASEB Journal.* 2011;25:615-650.

67. Wierzba-Bobrowicz T, Lewandowska E, Lechowicz W, Stepień T, Pasennik E. Quantitative analysis of activated microglia, ramified and damage of processes in the frontal and temporal lobes of chronic schizophrenics. *Folia Neuropathologica.* 2005;43:81-89.

68. Trepanier MO, Hopperton KE, Mizrahi R, Mechawar N, Bazinet RP. Postmortem evidence of cerebral inflammation in schizophrenia: a systematic review. *Molecular psychiatry.* 2016;21:1009.

69. Kipnis J. Multifaceted interactions between adaptive immunity and the central

nervous system. *Science.* 2016;353:766-771.

70. Kesteren C van, Gremmels H, Witte LD de, et al. Immune involvement in the pathogenesis of schizophrenia: a meta-analysis on postmortem brain studies. *Translational Psychiatry.* 2017;7:e1075.

71. Venneti S, Lopresti BJ, Wiley CA. The peripheral benzodiazepine receptor (Translocator protein 18kDa) in microglia: from pathology to imaging. *Progress in neurobiology.* 2006;80:308-322.

72. Toth M, Little P, Arnberg F, et al. Acute neuroinflammation in a clinically relevant focal cortical ischemic stroke model in rat: longitudinal positron emission tomography and immunofluorescent tracking. *Brain Structure and Function.* 2016;221:1279-1290.

73. Notter T, Coughlin JM, Sawa A, Meyer U. Reconceptualization of translocator protein as a biomarker of neuroinflammation in psychiatry. *Molecular psychiatry.* 2018;23:36.

74. Edison P, Archer HA, Gerhard A, et al. Microglia, amyloid, and cognition in Alzheimer's disease: An [11C](R) PK11195-PET and [11C] PIB-PET study. *Neurobiology of disease.* 2008;32:412-419.

75. Debruyne JC, Versijpt J, Van Laere KJ, et al. PET visualization of microglia in multiple sclerosis patients using [11C] PK11195. *European journal of neurology.* 2003;10:257-264.

76. Holmes SE, Hinz R, Conen S, et al. Elevated translocator protein in anterior cingulate in major depression and a role for inflammation in suicidal thinking: a positron emission tomography study. *Biological psychiatry.* 2018;83:61-69.

77. Van Berckel BN, Bossong MG, Boellaard R, et al. Microglia activation in recent-onset schizophrenia: a quantitative (R)-[11 C] PK11195 positron emission tomography study. *Biological psychiatry.* 2008;64:820-822.

78. Doorduin J, De Vries EFJ, Willemsen ATM, De Groot JC, Dierckx RA, Klein HC. Neuroinflammation in schizophrenia-related psychosis: a PET study. *Journal of Nuclear Medicine.* 2009;50:1801-1807.

79. Holmes SE, Hinz R, Drake RJ, et al. In vivo imaging of brain microglial activity in antipsychotic-free and medicated schizophrenia: a [11C](R)-PK11195 positron

emission tomography study. *Molecular psychiatry.* 2016;21:1672-1679.

80. Van Der Doef TF, De Witte LD, Sutterland AL, et al. In vivo (R)-[11C]PK11195 PET imaging of 18kDa translocator protein in recent onset psychosis. *NPJ schizophrenia.* 2016;2:16031.

81. Di Biase MA, Zalesky A, O'keefe G, et al. PET imaging of putative microglial activation in individuals at ultra-high risk for psychosis, recently diagnosed and chronically ill with schizophrenia. *Translational psychiatry.* 2017;7:e1225.

82. Kreisl WC, Fujita M, Fujimura Y, et al. Comparison of [11C]-(R)-PK 11195 and [11C]PBR28, two radioligands for translocator protein (18 kDa) in human and monkey: Implications for positron emission tomographic imaging of this inflammation biomarker. *Neuroimage.* 2010;49:2924-2932.

83. Kobayashi M, Jiang T, Telu S, et al. 11C-DPA-713 has much greater specific binding to translocator protein 18 kDa (TSPO) in human brain than 11C-(R)-PK11195. *Journal of Cerebral Blood Flow & Metabolism.* 2017:0271678X17699223.

84. Wilson AA, Garcia A, Parkes J, et al. Radiosynthesis and initial evaluation of [18F]-FEPPA for PET imaging of peripheral benzodiazepine receptors. *Nuclear Medicine and Biology.* 2008;35:305-314.

85. Jučaite A, Cselényi Z, Arvidsson A, et al. Kinetic analysis and test-retest variability of the radioligand [11C](R)-PK11195 binding to TSPO in the human brain - a PET study in control subjects. *EJNMMI research.* 2012;2:1.

86. Fujita M, Kobayashi M, Ikawa M, et al. Comparison of four 11C-labeled PET ligands to quantify translocator protein 18 kDa (TSPO) in human brain: (R)-PK11195, PBR28, DPA-713, and ER176—based on recent publications that measured specific-to-non-displaceable ratios. *EJNMMI Research.* 2017;7:84.

87. Collste K, Forsberg A, Varrone A, et al. Test–retest reproducibility of [11C]PBR28 binding to TSPO in healthy control subjects. *European Journal of Nuclear Medicine and Molecular Imaging.* 2016;43:173-183.

88. Owen DRJ, Gunn RN, Rabiner EA, et al. Mixed-affinity binding in humans with 18-kDa translocator protein ligands. *Journal of Nuclear Medicine.* 2011;52:24-32.

89. Owen DR, Yeo AJ, Gunn RN, et al. An 18-kDa translocator protein (TSPO) polymorphism explains differences in binding affinity of the PET radioligand PBR28.

Journal of Cerebral Blood Flow & Metabolism. 2012;32:1-5.

90. Kreisl WC, Jenko KJ, Hines CS, et al. A genetic polymorphism for translocator protein 18 kDa affects both in vitro and in vivo radioligand binding in human brain to this putative biomarker of neuroinflammation. *Journal of Cerebral Blood Flow & Metabolism.* 2013;33:53-58.

91. Nair A, Veronese M, Xu X, et al. Test-retest analysis of a non-invasive method of quantifying [11C]-PBR28 binding in Alzheimer's disease. *EJNMMI research.* 2016;6:72.

92. Albrecht DS, Normandin MD, Shcherbinin S, et al. Pseudoreference Regions for Glial Imaging with 11C-PBR28: Investigation in 2 Clinical Cohorts. *Journal of Nuclear Medicine.* 2018;59:107-114.

93. Bloomfield PS, Selvaraj S, Veronese M, et al. Microglial activity in people at ultra high risk of psychosis and in schizophrenia: an [11C] PBR28 PET brain imaging study. *American Journal of Psychiatry.* 2015.

94. Narendran R, Frankle WG. Comment on analyses and conclusions of "microglial activity in people at ULTRA high risk of psychosis and in schizophrenia: An [11C] PBR28 pet brain imaging study". *American Journal of Psychiatry.* 2016;173:536-537.

95. Turkheimer FE, Edison P, Pavese N, et al. Reference and target region modeling of [11C]-(R)-PK11195 brain studies. *Journal of Nuclear Medicine.* 2007;48:158-167.

96. Yaqub M, Van Berckel BN, Schuitemaker A, et al. Optimization of supervised cluster analysis for extracting reference tissue input curves in (R)-[11C] PK11195 brain PET studies. *Journal of Cerebral Blood Flow & Metabolism.* 2012;32:1600-1608.

97. Kenk M, Selvanathan T, Rao N, et al. Imaging neuroinflammation in gray and white matter in schizophrenia: an in-vivo PET study with [18F]-FEPPA. *Schizophrenia bulletin.* 2015;41:85-93.

98. Coughlin JM, Wang Y, Ambinder EB, et al. In vivo markers of inflammatory response in recent-onset schizophrenia : a combined study using [11 C] DPA-713 PET and analysis of CSF and plasma. *Translational Psychiatry.* 2016;6.

99. Hafizi S, Tseng HH, Rao N, et al. Imaging microglial activation in untreated first-episode psychosis: A PET study with [18F]FEPPA. *American Journal of Psychiatry.* 2017;174:118-124.

100. Ottoy J, De Picker L, Verhaeghe J, et al. [18F] PBR111 PET Imaging in

Healthy Controls and Schizophrenia: Test–Retest Reproducibility and Quantification of Neuroinflammation. *Journal of Nuclear Medicine.* 2018:jnumed-117.

101. Fisher RA. The design of experiments. 1935.

102. Fisher RA. The logic of inductive inference. *Journal of the Royal Statistical Society.* 1935;98:39-82.

103. Neyman J, Pearson ES. IX. On the problem of the most efficient tests of statistical hypotheses. *Phil Trans R Soc Lond A.* 1933;231:289-337.

104. Lee MD, Wagenmakers E-J. Bayesian cognitive modeling: A practical course. Cambridge university press; 2014.

105. McElreath R. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. CRC Press; 2018.

106. Fletcher PC, Grafton ST. Repeat after me: replication in clinical neuroimaging is critical. *NeuroImage: Clinical.* 2013;2:247.

107. Editorial-staff. Go forth and replicate. *Nature.* 2016;536:373.

108. Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nature Human Behaviour.* 2017;1:0021.

109. Kerr NL. HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review.* 1998;2:196-217.

110. Schmidt S. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology.* 2009;13:90.

111. Begley CG, Ioannidis JPA. Reproducibility in science: improving the standard for basic and preclinical research. *Circulation research.* 2015;116:116-126.

112. Poldrack RA, Baker CI, Durnez J, et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience.* 2017;18:115.

113. Matheson GJ, Stenkrona P, Cselényi Z, et al. Reliability of volumetric and surface-based normalisation and smoothing techniques for PET analysis of the cortex: A test-retest analysis using [11C]SCH-23390. *NeuroImage.* 2017;155:344-353.

114. Hedman E, Andersson G, Ljótsson B, et al. Internet-based cognitive behavior therapy vs. cognitive behavioral group therapy for social anxiety disorder: a

randomized controlled non-inferiority trial. *PloS one.* 2011;6:e18001.

115. Coughlin JM, Wang Y, Ambinder EB, et al. In vivo markers of inflammatory response in recent-onset schizophrenia: a combined study using [11C]DPA-713 PET and analysis of CSF and plasma. *Translational Psychiatry.* 2016;6.

116. Owen DR, Guo Q, Kalk NJ, et al. Determination of [11C]PBR28 binding potential in vivo: a first human TSPO blocking study. *Journal of Cerebral Blood Flow & Metabolism.* 2014;34:989-994.

117. Gustavsson JP, Bergman H, Edman G, Ekselius L, Knorring L von, Linder J. Swedish universities Scales of Personality (SSP): construction, internal consistency and normative data. *Acta psychiatrica Scandinavica.* 2000;102:217-225.

118. Schalling D, Edman G. Personality and vulnerability to psychopathology: the development of the Karolinska Scales of Personality (KSP). *Stockholm: Karolinska Institutet.* 1987.

119. Nordström P, Schalling D, Asberg M. Temperamental vulnerability in attempted suicide. *Acta Psychiatrica Scandinavica.* 1995;92:155-160.

120. Schalling D, Edman G. The Karolinska Scales of Personality (KSP) Manual: an inventory for assessing temperament dimensions associated with vulnerability for psychosocial deviance. *Stockholm, Sweden: Department of Psychiatry, Karolinska Institutet.* 1993.

121. Ekelund J, Slifstein M, Narendran R, et al. In vivo DA D1 receptor selectivity of NNC 112 and SCH 23390. *Molecular Imaging and Biology.* 2007;9:117-125.

122. Halldin C, Farde L, Hogberg T, et al. Carbon-11-FLB 457: a radioligand for extrastriatal D2 dopamine receptors. *J Nucl Med.* 1995;36:1275-81.

123. Briard E, Zoghbi SS, Imaizumi M, et al. Synthesis and evaluation in monkey of two sensitive 11C-labeled aryloxyanilide ligands for imaging brain peripheral benzodiazepine receptors in vivo. *Journal of medicinal chemistry.* 2007;51:17-30.

124. Shah F, Hume SP, Pike VW, Ashworth S, McDermott J. Synthesis of the enantiomers of [N-methyl-11C] PK 11195 and comparison of their behaviours as radioligands for PK binding sites in rats. *Nuclear medicine and biology.* 1994;21:573-581.

125. Fujita M, Imaizumi M, Zoghbi SS, Fujimura Y, Farris AG, Suhara T. Kinetic analysis in healthy humans of a novel positron emission tomography radioligand to

image the peripheral benzodiazepine receptor, a potential biomarker for inflammation. *NeuroImage.* 2008;40.

126. Boutin H, Chauveau F, Thominiaux C, et al. 11C-DPA-713: a novel peripheral benzodiazepine receptor PET ligand for in vivo imaging of neuroinflammation. *Journal of Nuclear Medicine.* 2007;48:573-581.

127. Wienhard K, Dahlbom M, Eriksson L, et al. The ECAT EXACT HR: performance of a new high resolution positron scanner. *Journal of Computer Assisted Tomography.* 1994;18:110-118.

128. Wienhard K, Schmand M, Casey ME, et al. The ECAT HRRT: performance and first clinical application of the new high resolution research tomograph. *IEEE Transactions on Nuclear Science.* 2002;49:104-110.

129. Varrone A, Sjöholm N, Eriksson L, Gulyás B, Halldin C, Farde L. Advancement in PET quantification using 3D-OP-OSEM point spread function reconstruction with the HRRT. *European Journal of Nuclear Medicine and Molecular Imaging.* 2009;36:1639-1650.

130. Bergström M, Boethius J, Eriksson L, Greitz T, Ribbe T, Widen L. Head fixation device for reproducible position alignment in transmission CT and positron emission tomography. *Journal of Computer Assisted Tomography.* 1981;5:136-141.

131. Schain M, Tóth M, Cselényi Z, et al. Quantification of serotonin transporter availability with [11C]MADAM - a comparison between the ECAT HRRT and HR systems. *Neuroimage.* 2012;60:800-807.

132. Martinez D, Slifstein M, Broft A, et al. Imaging human mesolimbic dopamine transmission with positron emission tomography. Part II: Amphetamine-induced dopamine release in the functional subdivisions of the striatum. *Journal of Cerebral Blood Flow and Metabolism.* 2003;23:285-300.

133. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. Fsl. *Neuroimage.* 2012;62:782-790.

134. Cselényi Z, Olsson H, Farde L, Gulyás B. Wavelet-Aided Parametric Mapping of Cerebral Dopamine D-2 Receptors Using the High Affinity PET Radioligand [11C] FLB 457. *Neuroimage.* 2002;17:47-60.

135. Turkheimer FE, Aston JAD, Banati RB, Riddell C, Cunningham VJ. A linear wavelet filter for parametric imaging with dynamic PET. *Medical Imaging, IEEE*

Transactions on. 2003;22:289-301.

136. Slifstein M, Laruelle M. Models and methods for derivation of in vivo neuroreceptor parameters with PET and SPECT reversible radiotracers. *Nuclear Medicine and Biology.* 2001;28:595-608.

137. Varnäs K, Varrone A, Farde L. Modeling of PET data in CNS drug discovery and development. *Journal of pharmacokinetics and pharmacodynamics.* 2013;40:267-279.

138. Asselin M-C, Montgomery AJ, Grasby PM, Hume SP. Quantification of PET studies with the very high-affinity dopamine D2/D3 receptor ligand [11C]FLB 457: re-evaluation of the validity of using a cerebellar reference region. *Journal of Cerebral Blood Flow and Metabolism.* 2007;27:378-392.

139. Narendran R, Mason NS, Chen C-M, et al. Evaluation of dopamine D2/3 specific binding in the cerebellum for the positron emission tomography radiotracer [11C] FLB 457: Implications for measuring cortical dopamine release. *Synapse.* 2011;65:991-997.

140. Cselényi Z, Olsson H, Halldin C, Gulyás B, Farde L. A comparison of recent parametric neuroreceptor mapping approaches based on measurements with the high affinity PET radioligands [11C] FLB 457 and [11C] WAY 100635. *Neuroimage.* 2006;32:1690-1708.

141. Bloomfield PS, Howes OD, Turkheimer F, Selvaraj S, Veronese M. Response to Narendran and Frankle: the interpretation of PET microglial imaging in schizophrenia. *American Journal of Psychiatry.* 2016;173:537-538.

142. Ogden RT, Zanderigo F, Parsey RV. Estimation of in vivo nonspecific binding in positron emission tomography studies without requiring a reference region. *Neuroimage.* 2015;108:234-242.

143. Cunningham VJ, Rabiner EA, Slifstein M, Laruelle M, Gunn RN. Measuring drug occupancy in the absence of a reference region: the Lassen plot re-visited. *J Cereb Blood Flow Metab.* 2010;30.

144. Schain M, Zanderigo F, Ogden RT. Likelihood estimation of drug occupancy for brain PET studies. *NeuroImage.* 2018.

145. Verhagen J, Wagenmakers E-J. Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General.* 2014;143:1457.

146. Wagenmakers E-J, Verhagen J, Ly A. How to quantify the evidence for the

absence of a correlation. *Behavior Research Methods.* 2016;48:413-426.

147. Collste K, Plavén-Sigray P, Fatouros-Bergman H, et al. Lower levels of the glial cell marker TSPO in drug-naive first-episode psychosis patients as measured using PET and [11C]PBR28. *Molecular Psychiatry.* 2017;22:850-856.

148. Jeffreys H. Theory of probability. 3rd ed. Oxford: Oxford University Press; 1961.

149. Kass RE, Raftery AE. Bayes factors. *Journal of the american statistical association.* 1995;90:773-795.

150. Dickey JM. The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics.* 1971:204-223.

151. Portney LG, Watkins MP. Foundations of clinical research: application to practice. 3rd ed.; 2009.

152. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research.* 2005;19:231-240.

153. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science.* 2012;23:524-532.

154. Turkheimer FE, Smith CB, Schmidt K. Estimation of the number of "true" null hypotheses in multivariate analysis of neuroimaging data. *Neuroimage.* 2001;13:920-930.

155. Chumbley JR, Friston KJ. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage.* 2009;44:62-70.

156. Rytwinski NK, Fresco DM, Heimberg RG, et al. Screening for social anxiety disorder with the self-report version of the Liebowitz Social Anxiety Scale. *Depression and Anxiety.* 2009;26:34-38.

157. Stalnaker TA, Cooch NK, Schoenbaum G. What the orbitofrontal cortex does not do. *Nature Neuroscience.* 2015;18:620-627.

158. Grupe DW, Nitschke JB. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat Rev Neurosci.* 2013;14:488-501.

159. Corlett PR, Fletcher PC. Delusions and prediction error: clarifying the roles of

behavioural and brain responses. *Cognitive neuropsychiatry.* 2015;20:95-105.

160. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature.* 2009;459:837-841.

161. Bromberg-Martin ES, Matsumoto M, Hikosaka O. Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. *Neuron.* 2010;68:815-834.

162. Tudur Smith C, Marcucci M, Nolan SJ, et al. Individual participant data meta-analyses compared with meta-analyses based on aggregate data. *The Cochrane Library.* 2016.

163. Cohen J. Statistical power analysis for the behavioral sciences. 1998.

164. Schain M, Zanderigo F, Mann JJ, Ogden RT. Estimation of the binding potential BPND without a reference region or blood samples for brain PET studies. *NeuroImage.* 2017;146:121-131.

165. Chambers CD, Dienes Z, McIntosh RD, Rotshtein P, Willmes K. Registered reports: realigning incentives in scientific publishing. *Cortex.* 2015;66:A1-A2.

166. Gorgolewski KJ, Auer T, Calhoun VD, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data.* 2016;3:160044.