Supplementary Materials: D1-Dopamine Receptor Availability in First-Episode Neuroleptic Naive Psychosis Patients

Stenkrona, P.[1]
Matheson, G.J.[1]
Halldin, C.[1]
Cervenka, S.[1]
Farde, L.[1,2]

Contents

Supplementary Materials S1: Choice of Priors for Bayesian Analysis

Priors were defined for the effects of both age and disease status on $[^{11}C]$SCH23390 BP_{ND}. The prior for age was defined as a normal distribution centred around proportional changes estimated in previous studies. This has the effect of informing the model about the approximate size of the effect of age, and assigning the majority of the probability mass around the expected values. The priors for disease status were defined as a zero-centred halfnormal distribution, with standard deviation equal to the proportional changes estimated in previous studies. This has the effect of informing the model that extreme values are unlikely thereby preventing overfitting, as well as by defining a diffuse prior probability density (i.e. which informs the estimation of the parameter minimally) for the parameter in which approximately two thirds of the probability density is lower, and approximately one third is higher than the effect size estimated by previous studies.

Age

We selected previous studies in which the effects of age on $[^{11}C]$SCH23390 BP_{ND} were explicitly examined. We chose the regions most similar to the regions examined in this paper and extracted as a percentage change per decade. We then calculated weighted mean and standard deviation of the proportional changes for each region.

DLPFC

In each case, the decrease in BP_{ND} per decade was calculated relative to the mean BP_{ND} for the study.

- Wang 1998: OCC - 8% per decade (n=18)
- Jucaite 2010: DLPFC - 15.5% per decade (n=30)
  - This estimate comes from predictions of the nonlinear curve for the youngest and oldest age in our sample, and calculating the average decrease as if it were linear over this period.
- Backman 2011: DLPFC - 24% per decade (n=40)
- de Boer 2017: DLPFC - 12.2% per decade (n=56)

The weighted summary statistics were as follows: mean = -1.56% per year, s.d. = 0.66%.

Striatum

- Wang 1998: Caudate - 6.9% per decade
- Jucaite 2010: Caudate - 7.7% per decade
  - This estimate comes from predictions of the nonlinear curve for the youngest and oldest age in our sample, and calculating the average decrease as if it were linear over this period.
- Backman 2011: Caudate - 8% per decade
- de Boer 2017: Caudate - 7.4% per decade (n=56)

The weighted summary statistics were as follows: mean = -0.76% per year, s.d. = 0.04%.

Patient-Control Status

Because we would be calculating Bayes Factors, assessing the relative credibility of the hypotheses of increased and decreased $[^{11}C]$SCH23390 BP_{ND} in patients compared to controls, relative to the null hypothesis of no effect, we decided that the width of the priors for each hypothesis should be equal such that more inflated estimates in one direction might not disadvantage that hypothesis over the other.

DLPFC

- Increase hypothesis: Poels, Girgis, Thompson, Slifstein, & Abi-Dargham (2013) showed a 35% increase in patients compared to controls.
- Decrease hypothesis: Kosaka et al. (2010) showed a 27%, and Hirvonen et al. (2006) showed a 26%, reduction in patients compared to controls,

The standard deviation of the prior will be set to 29% for the DLPFC.

Striatum

- Increase hypothesis: Poels et al. (2013) did not show a difference between patients and controls.
- Decrease hypothesis: Kosaka et al. (2010) showed a 20.9%, and Hirvonen et al. (2006) showed a 20%, reduction in patients compared to controls,

The standard deviation of the prior will be set to 21% for the striatum.

Supplementary Materials S2: Confounder Analysis

In order to assess the potential confounding effects of differences between individuals, we systematically assessed the impact of each confounder to determine whether it should be accounted for in the statistical model. For all binary confounders, we visualise the effects, and test for their effects using both a t-test as well as an equivalence test. The smallest effect size of interest (SESOI) for the equivalence tests was set to a Cohen's D of 0.98 in both the positive and negative directions. This is based on the effect size which a two sample test of 18 measurements in each sample can detect with 80 power. While this is a large effect size, and is not ideal, this selection can be partially justified by the fact that the confounding effects of systematic differences in acquisition are likely to be large if they are present. Further, testing for equivalence with a large equivalence range is statistically sound, while attempting to conclude equivalence based on an insignificant p value in a test for differences between groups is not.

Table 1: Settings of the PET and MRI examinations

Subject no	File format	Neuro-insert	2D/3D Acquisition	R-Z Filter cutoff frequency	R-Z Filter Resolution	MRI modality
HC 1	ecat6	no	2D	0.5	0	T2
HC 2	ecat6	no	2D	0.5	0	T2
HC 3	ecat6	no	2D	0	0	T2
HC 4	ecat6	no	2D	0	0	T2
HC 5	ecat6	no	2D	0	0	T2
HC 6	ecat6	no	2D	0	0	T2
HC 7	ecat6	no	2D	0	0	T2
HC 8	ecat7	yes	2D	0	0.2	T1
HC 9	ecat7	yes	2D	0	0.2	T1
HC 10	ecat7	yes	2D	0	0	T1
HC 11	ecat7	yes	2D	0	0.2	T1
HC 12	ecat7	yes	2D	0	0.2	T1
HC 13	ecat7	yes	2D	0	0.2	T1
HC 14	ecat7	yes	2D	0	0	T1
HC 15	ecat7	yes	2D	0	0.2	T1
HC 16	ecat7	yes	2D	0	0.2	T1
HC 17	ecat7	yes	2D	0	0.2	T1
SCZ 1	ecat6	no	2D	0	0	T2
SCZ 2	ecat6	no	2D	0	0	T2
SCZ 3	ecat6	no	2D	0	0	T2
SCZ 4	ecat6	no	2D	0	0	T2
SCZ 5	ecat6	no	2D	0	0	T2
SCZ 6	ecat6	no	2D	0	0	T2
SCZ 7	ecat6	no	2D	0	0	T2
SCZ 8	ecat6	no	2D	0	0	T2
SCZ 9	ecat6	no	2D	0	0	T2
SCZ 10	ecat6	no	2D	0	0	T2
SCZ 11	ecat7	yes	3D	0	0.2	T1
SCZ 12	ecat7	yes	2D	0	0.2	T1
SCZ 13	ecat7	yes	3D	0	0.2	T1
SCZ 14	ecat7	yes	2D	0	0.2	T1
SCZ 15	ecat7	yes	2D	0	0.2	T1
SCZ 16	ecat7	yes	2D	0	0.2	T1
SCZ 17	ecat7	yes	2D	0	0.2	T1
SCZ 18	ecat7	yes	2D	0	0.2	T1

Note:
HC=healthy controls; SCZ=patients.

Acquisition confounders

R-Z Filter Resolution

There did not appear to be any substantial difference in BP_{ND} between groups whose R-Z filter resolution differed.

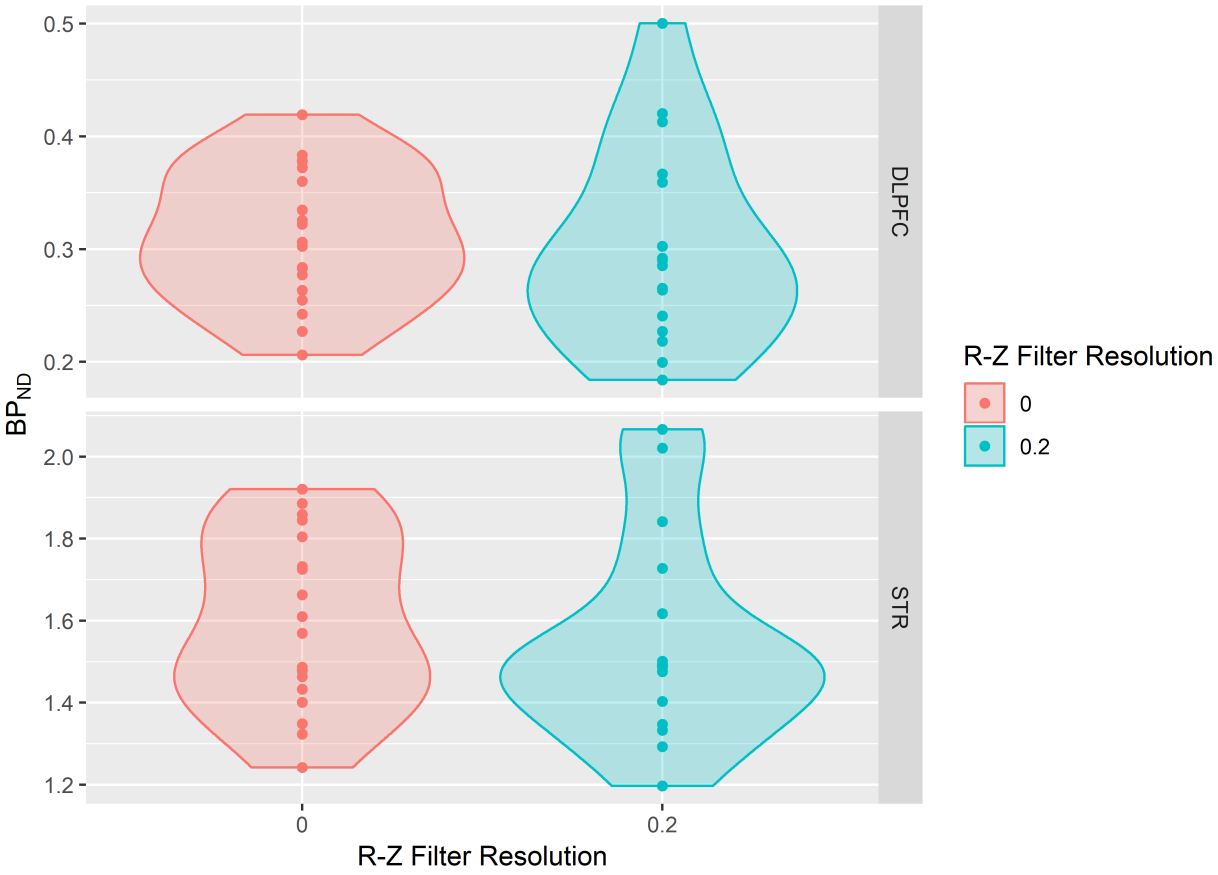

Figure 1: R-Z filter resolution comparison

There were no significant differences between the groups.

ROI	estimate	estimate1	estimate2	statistic	parameter	p.value	method
STR	0.044	1.593	1.549	0.559	29.428	0.581	Welch Two Sample t-test
DLPFC	0.010	0.312	0.302	0.378	25.568	0.709	Welch Two Sample t-test

There was significant equivalence between the groups.

ROI	diff	tost_p.value
STR	0.044	0.014
DLPFC	0.010	0.010

It is therefore concluded that R-Z Filter Resolution is not likely to have been a confounding factor and will not be included in the statistical model.

R-Z Filter Cutoff

There were only two measurements whose R-Z filter cutoff was 0.5 as opposed to 0. This means that the statistical tests will be underpowered to reliably detect differences or equivalence between groups. However, from visual inspection, it does not appear that R-Z filter cutoff affected the results greatly.

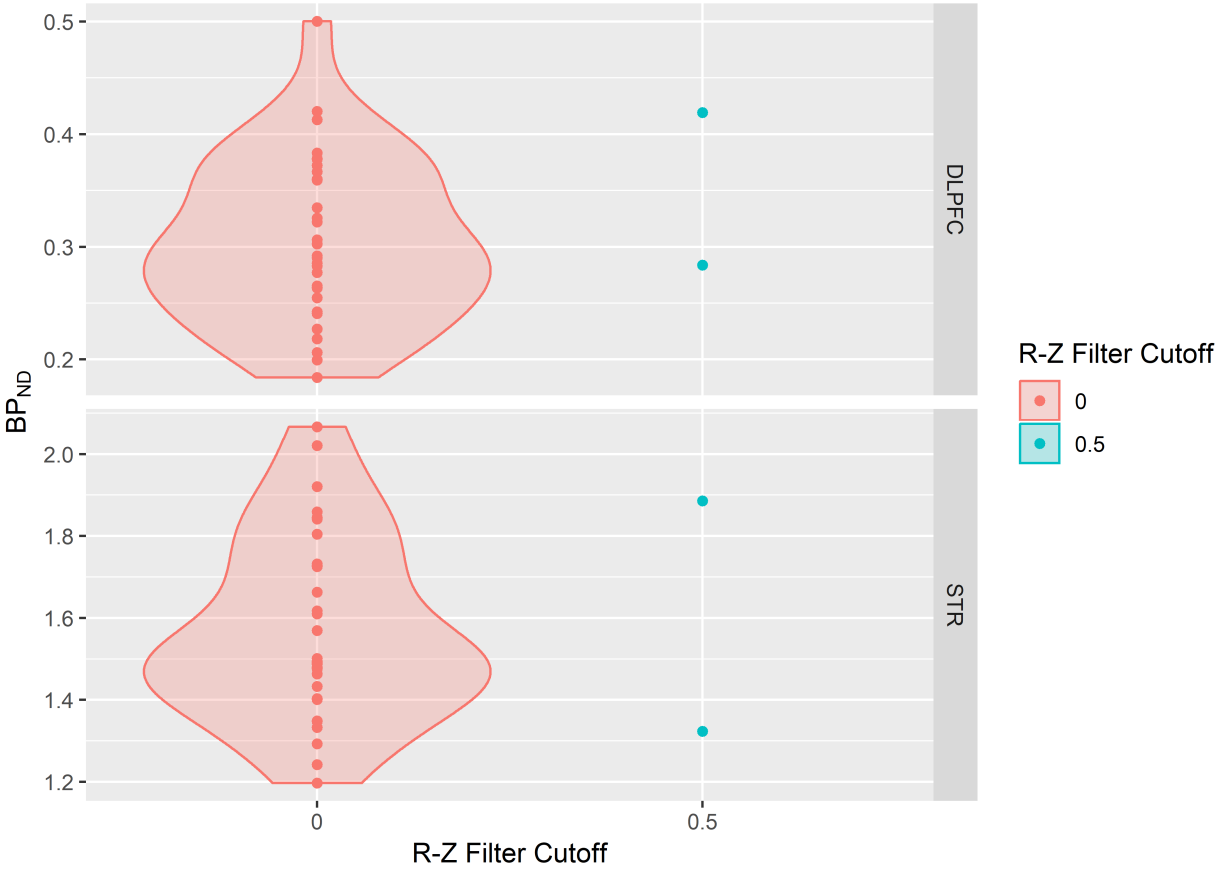

Figure 2: R-Z filter cutoff comparison

There were no significant differences between the groups.

ROI	estimate	estimate1	estimate2	statistic	parameter	p.value	method
STR	-0.033	1.571	1.605	-0.118	1.038	0.925	Welch Two Sample t-test
DLPFC	-0.047	0.304	0.352	-0.686	1.071	0.611	Welch Two Sample t-test

There was no significant equivalence between the groups.

ROI	diff	tost_p.value
STR	-0.033	0.249
DLPFC	-0.047	0.344

Because there were only two individuals with different R-Z filter cutoff, the statistical tests were unable to provide a statistically significant answer regarding the influence of the R-Z filter cutoff. However, from visual inspection of the figure, it does not appear to have made a large difference, and theoretically, we would not have expected this to play a large role in the outcome. It is therefore concluded that R-Z filter cutoff is unlikely to have been a confounding factor and will not be included in the statistical model.

File format

Ecat6 file format stores a 2D representation of the data, and Ecat7 file format stores a 3D representation of the data. A previous observation was made within our research group of differences between the formats, although the conversion software has changed.

There did not appear to be any substantial difference in BP_{ND} between groups whose PET measurements were saved in ecat6 and ecat7 file formats.

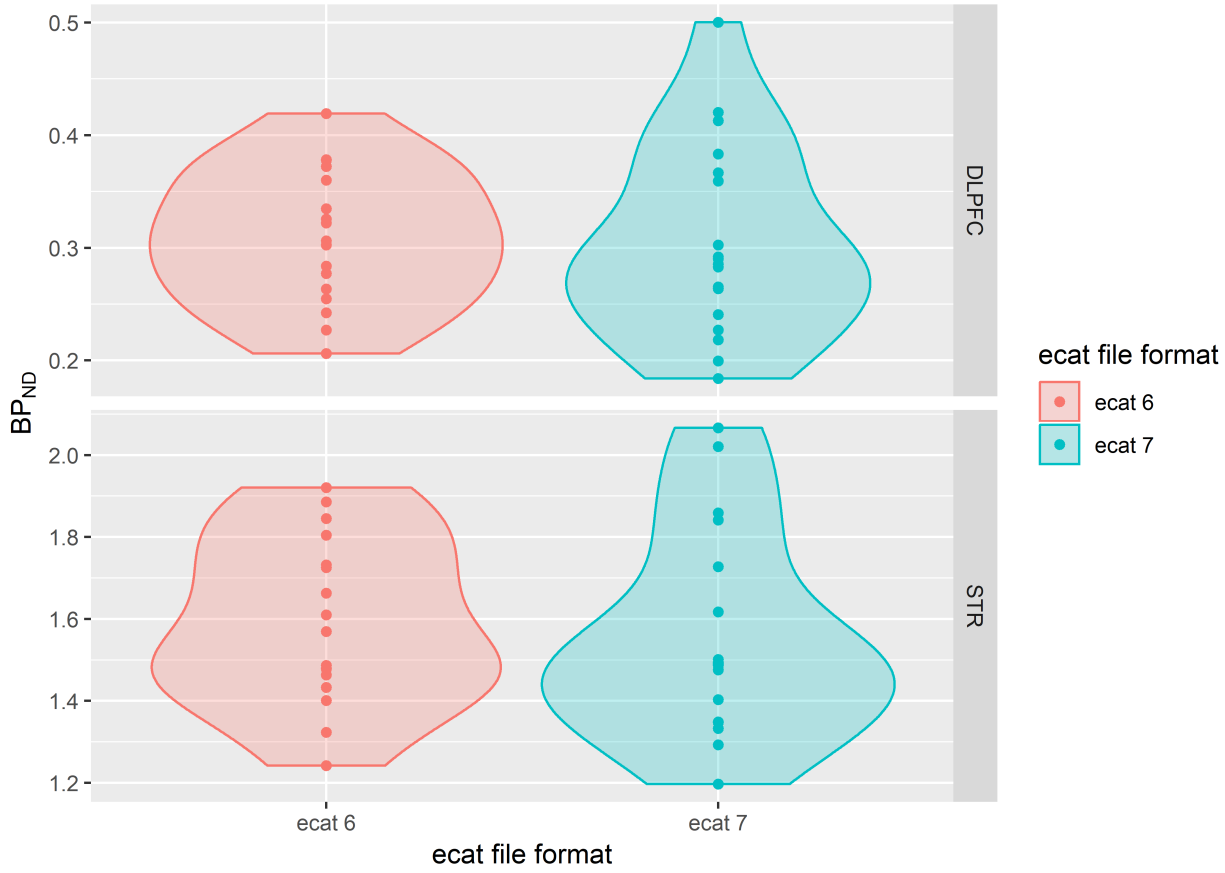

Figure 3: File format comparison

There were no significant differences between the groups.

ROI	estimate	estimate1	estimate2	statistic	parameter	p.value	method
STR	0.037	1.592	1.555	0.477	32.255	0.637	Welch Two Sample t-test
DLPFC	0.004	0.309	0.305	0.153	30.587	0.880	Welch Two Sample t-test

There was significant equivalence between the groups.

ROI	diff	tost_p.value
STR	0.037	0.010
DLPFC	0.004	0.005

It is therefore concluded that file format is not likely to have been a confounding factor and will not be included in the statistical model.

2D and 3D PET

Two of the patients' PET measurements were acquired in 3D instead of 2D. We did not have any a priori expectations of whether this might result in higher, lower or unchanged BP_{ND} values. Since there were only two measurements recorded in 3D, this means that the statistical tests will be underpowered to reliably detect differences or equivalence between groups.

There did not appear to be any large difference in BP_{ND} between groups whose PET measurements were recorded in 2D and 3D, although there may be a small negative bias.

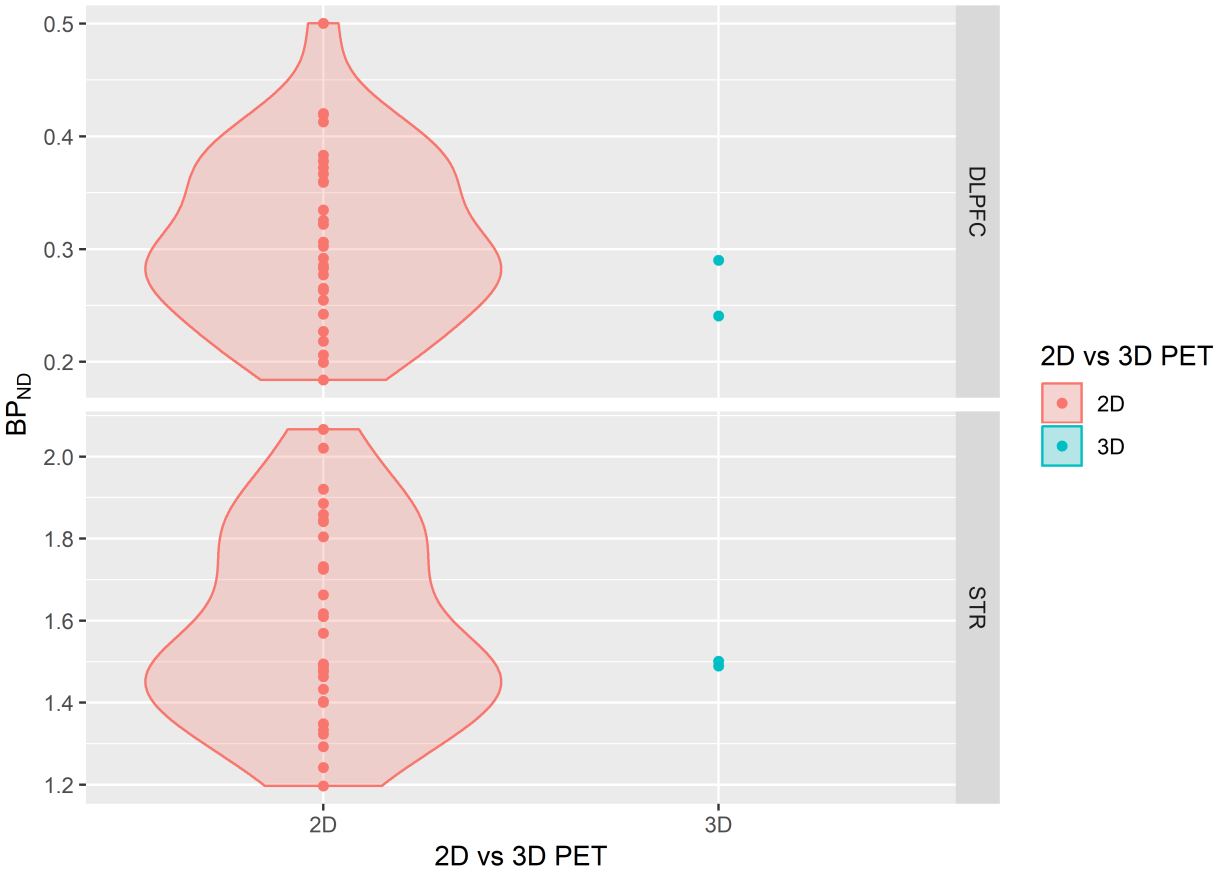

Figure 4: 2D and 3D PET comparison

There were no significant differences between the groups, although the difference was nearly at the significance level for the striatum.

ROI	estimate	estimate1	estimate2	statistic	parameter	p.value	method
STR	0.082	1.578	1.495	2.019	32.892	0.052	Welch Two Sample t-test
DLPFC	0.044	0.310	0.265	1.587	1.626	0.280	Welch Two Sample t-test

There was significant equivalence in the striatum, but not in the the DLPFC.

ROI	diff	tost_p.value
STR	0.082	0.032
DLPFC	0.044	0.350

Because there were only two individuals with 3D PET acquisition, the results of the statistical tests cannot be relied upon regarding the influence of this factor. Due to the fact that some bias might be evident from the figure above, the influence of the two 3D PET acquisitions must be assessed in the final statistical model. Because there are only two measurements acquired in this way, this cannot be simply added to the regression model, but the effect sizes resulting from the final model should be tested with both the inclusion and exclusion of these two individuals to assess whether this has any systematic influence on the results.

NeuroInsert

This is a shield which shields radiation from the body. Its influence is likely to be substantial for 3D scans, but not for 2D scans. Firstly we checked whether any of the measurements which were acquired in 3D were also not acquired with the NeuroInsert. This was not the case: both 3D measurements were acquired with the NeuroInsert in place.

There did not appear to be any substantial difference in BP_{ND} between groups whose PET measurements were acquired with and without the NeuroInsert.

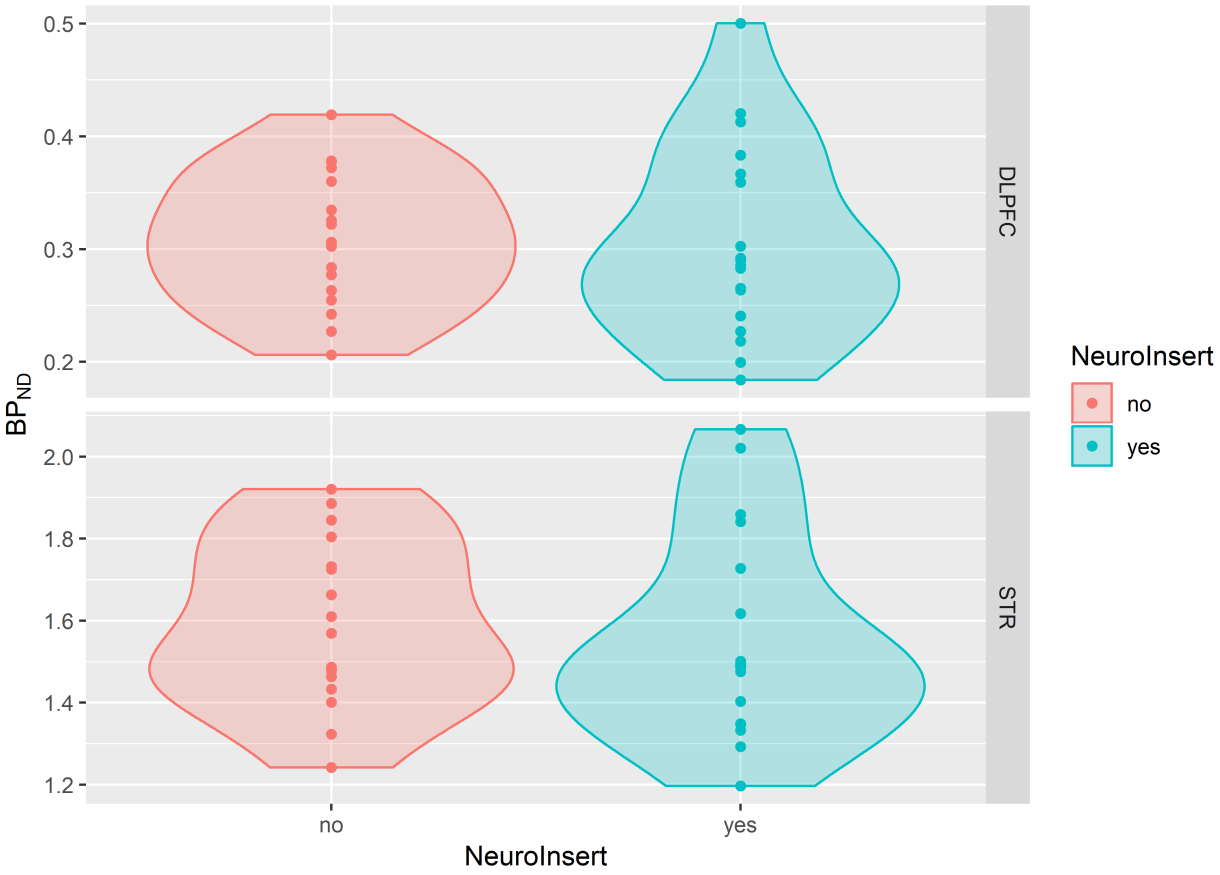

Figure 5: NeuroInsert comparison

There were no significant differences between the groups.

estimate	estimate1	estimate2	statistic	parameter	p.value	method
0.037	1.592	1.555	0.477	32.255	0.637	Welch Two Sample t-test
0.004	0.309	0.305	0.153	30.587	0.880	Welch Two Sample t-test

There was significant equivalence between the groups.

ROI	diff	tost_p.value
STR	0.037	0.010

ROI	diff	tost_p.value
DLPFC	0.004	0.005

It is therefore concluded that the presence or absence of the NeuroInsert is not likely to have been a confounding factor and will not be included in the statistical model.

MR modality

Some measurements only had T2 MR measurements. These measurements therefore have lower spatial accuracy for ROI delineation and may affect BP_{ND} outcomes.

There did not appear to be any substantial difference in BP_{ND} between groups and the modality of the MR images.

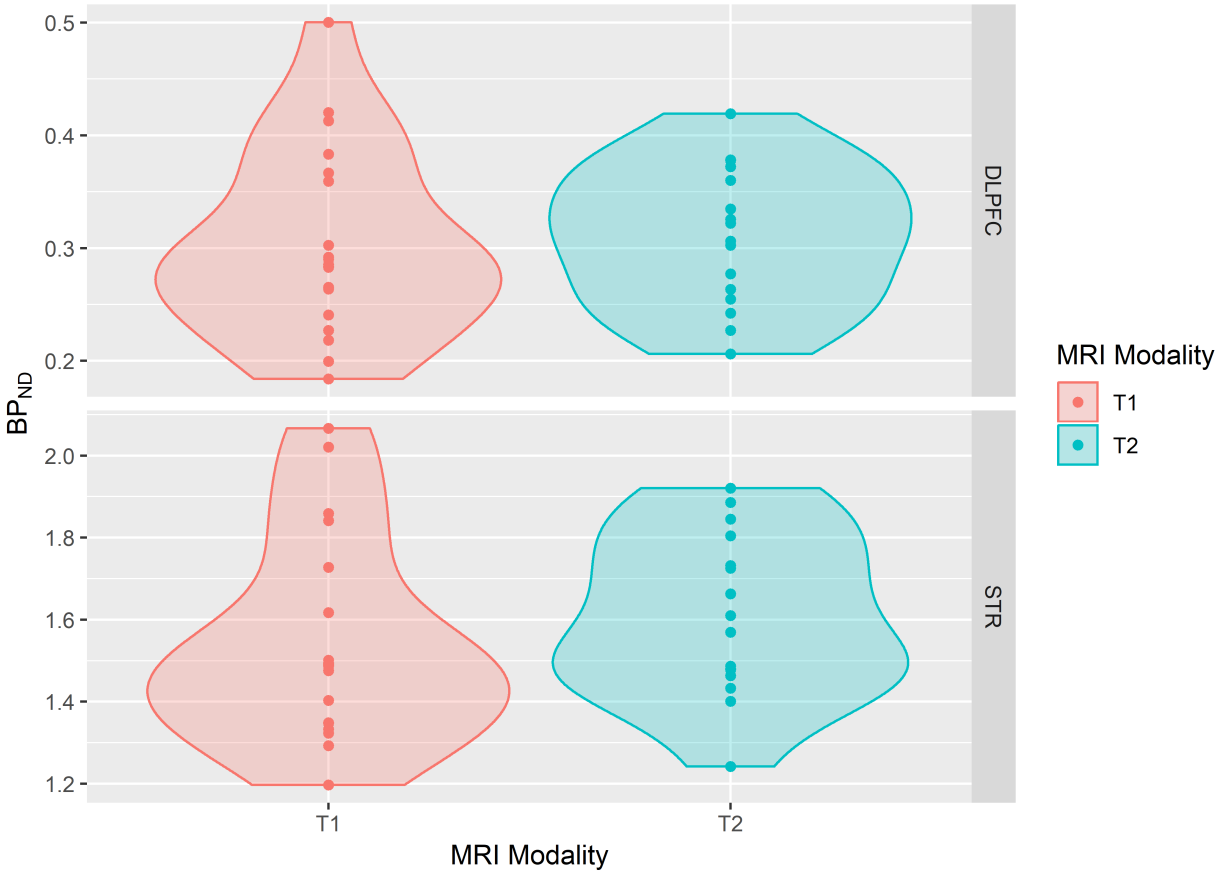

Figure 6: MR modality comparison

There were no significant differences between the groups.

ROI	estimate	estimate1	estimate2	statistic	parameter	p.value	method
STR	0.066	1.609	1.543	0.870	32.883	0.391	Welch Two Sample t-test
DLPFC	0.006	0.311	0.304	0.264	32.550	0.793	Welch Two Sample t-test

There was significant equivalence between the groups.

ROI	diff	tost_p.value
STR	0.066	0.024
DLPFC	0.006	0.006

It is therefore concluded that MR modality is not likely to have been a confounding factor and will not be included in the statistical model.

ROI delineation

The manual delineation of ROIs was performed blind to patient/control status. However, this does not rule out the possibility of bias introduced in ROI delineation due to visible differences between groups. In order to test for this, we also delineated automatic regions of interest to test whether there was any systematic bias between automatic and manual ROIs. Due to the fact that T1 and T2 MR images were used, manual delineation was considered the gold standard, and automated ROI delineation was only performed to validate that there had not been any bias in the delineation between groups.

Automatic delineation was performed by warping masks from MNI space into individual space. MR images were first pre-masked and skull-stripped using BET (Smith, 2002). MR images were then registered to ICBM 2009c Nonlinear Asymmetric $1 \times 1 \times 1$ mm templates (Fonov et al., 2011), using the T1w template for T1w MR measurements, and using the T2w template for T2w MR measurements. This was achieved by first applying an affine transformation using FLIRT (Jenkinson, Bannister, Brady, & Smith, 2002) using the skull-stripped image (registering to the skull-stripped template), followed by a non-linear registration of the original unstripped image to MNI space using FNIRT (Andersson, Jenkinson, & Smith, 2007). The FNIRT warp parameters were inverted, such that they could be used to warp ROIs in MNI space to individual MR space. These ROIs were then resliced to PET space using the inverse of the SPM coregistration parameters.

For the striatum, we combined all three ROIs from the the 1mm FSL (FMRIB Software Library) maximum probability Oxford-Imanova Striatal Connectivity Atlas with three subdivisions thresholded at 50% (Tziortzi et al., 2014). For the DLPFC, we used the MFG ROI from the 1mm FSL maximum probability Harvard-Oxford Cortical Atlas thresholded at 25%. The cerebellum was defined using the FSL 1mm maximum probability MNIfnirt Probabilistic cerebellar atlas (Diedrichsen, Balsters, Flavell, Cussans, & Ramnani, 2009), from which the cerebellar ROI consisted of cerebellar regions VI, Crus I and Crus II (indices [5, 7, 8, 10, 11, 13]). All three ROIs were multiplied by the FreeSurfer grey matter segmentation mask, to obtain ROIs consisting only of voxels identified as belonging to the grey matter. The cerebellar ROI was additionally trimmed to improve its properties as a reference region: 8mm from the cortex, 8mm from the vermis (defined using the same atlas), 4mm from the edge of the brain mask (using the FSL brain mask), and finally voxels belonging to the two most inferior planes of the PET image were excluded from the ROI. This method of cerebellar reference region delineation has previously been shown to exhibit good reliability (Matheson et al., 2017; Stenkrona et al., 2018).

Manual and automated ROI delineation methods yielded highly correlated BP_{ND} outcomes, with one outlier in the striatum for automated methods. Because the automated delineation was performed as a validation measure, this measurement was excluded from the following comparisons of delineation strategy.

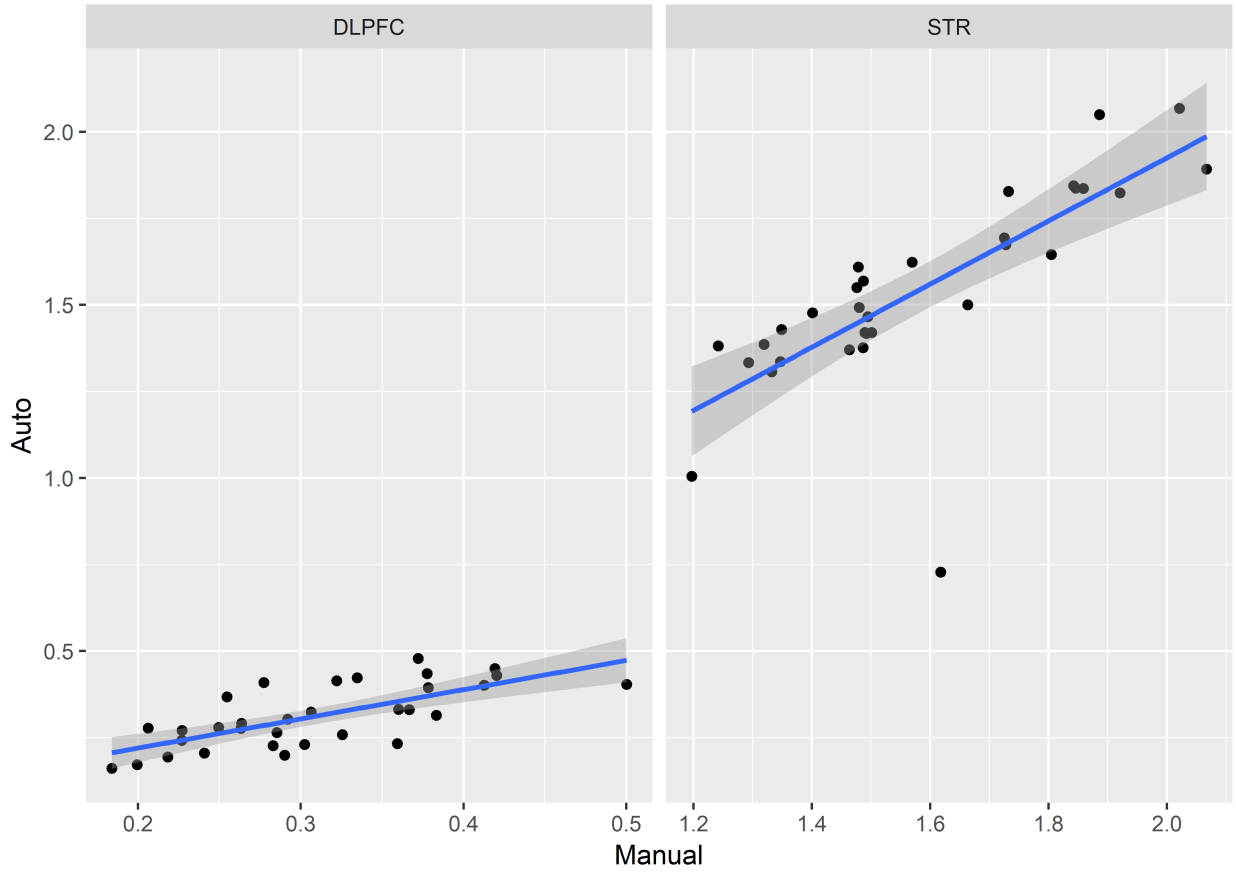

Figure 7: Correlations between BP_{ND} using automated and manual ROI delineation

Following the removal of the outlier individual from both groups, the correlations between BP_{ND} values obtained using automated and manual ROI delineation methods were high (DLPFC: r=0.71, STR: r=0.92).

Bias was calculated as the BP_{ND} resulting from manual delineation divided by the BP_{ND} resulting from automated delineation. Bias was compared between patient and control groups, in which there appeared to be no substantial differences between groups.

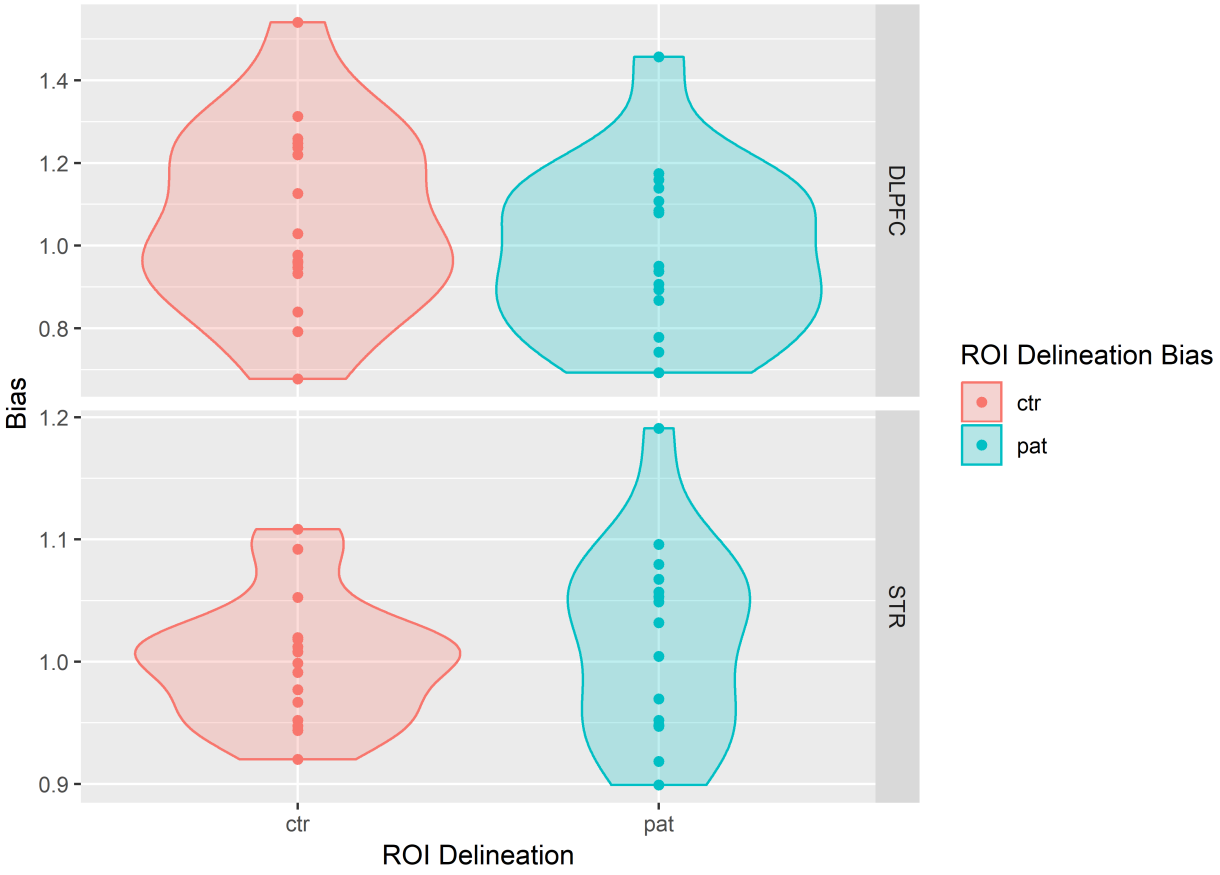

Figure 8: ROI delineation bias comparison

There were no significant differences between the groups.

ROI	estimate	estimate1	estimate2	statistic	parameter	p.value	method
DLPFC	0.082	1.066	0.984	1.086	29.663	0.286	Welch Two Sample t-test
STR	-0.016	1.002	1.018	-0.657	24.021	0.518	Welch Two Sample t-test

There was significant equivalence in the striatum, and close to significant equivalence in the DLPFC.

ROI	diff	tost_p.value
DLPFC	0.082	0.051
STR	-0.016	0.026

In conclusion, there do not appear to be any systematic differences in BP_{ND} between groups induced by the manual ROI delineation.

Date of Measurement

To assess whether there was any drift in BP_{ND}, we checked whether there were any systematic changes in BP_{ND} values with the year of measurement.

There did not appear to be any substantial drift in BP_{ND} over time.

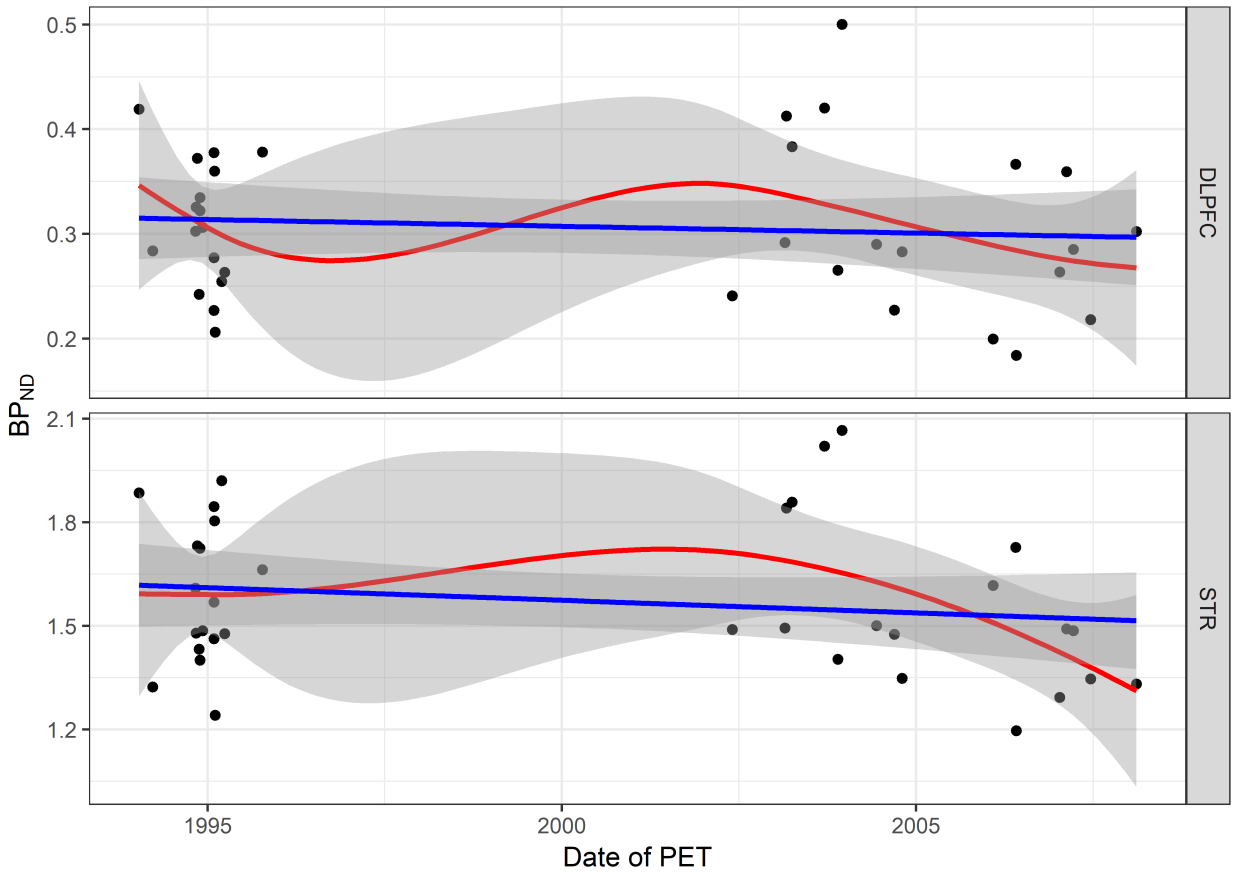

Figure 9: Neither a linear model (blue) nor a spline (red) reveal any substantial drift over time in BP_{ND} values

Due to cluestering of measurements by date, BP_{ND} values were also plotted by year as below, similarly not suggesting any systematic change over time.

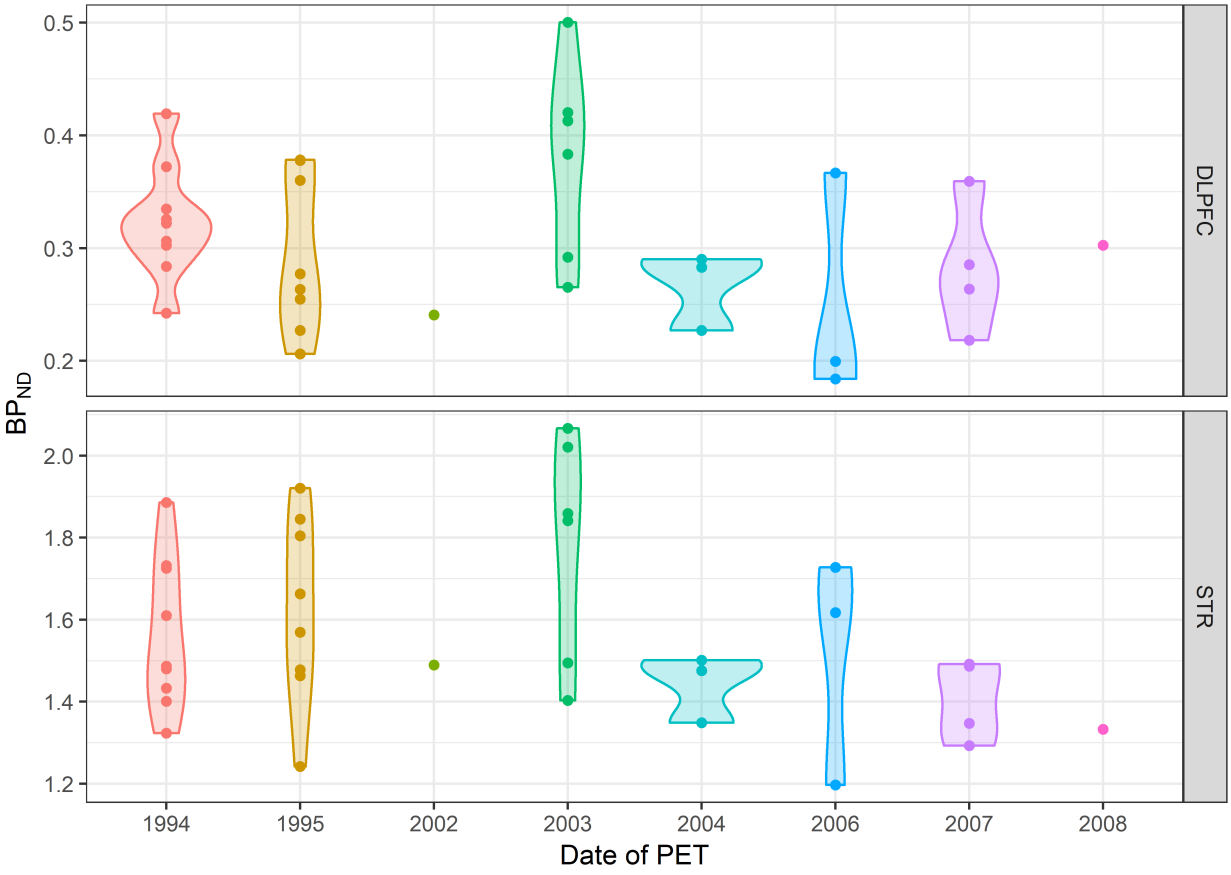

Figure 10: Yearly distributions of BP_{ND} values

It is therefore concluded that the date of measurement is not likely to have been a confounding factor and will not be included in the statistical model.

Movement

Despite our use of frame-to-frame realignment for movement correction, movement during frames, as well as movement away from the position in which the transmission scan was recorded, can result in biased outcome measures. Systematic differences in movement between groups is a plausible hypothesis in patient studies, and should be investigated. Here we depict the translation and rotation parameters returned during frame-by-frame realignment with their corresponding 10% and 90% quantiles.

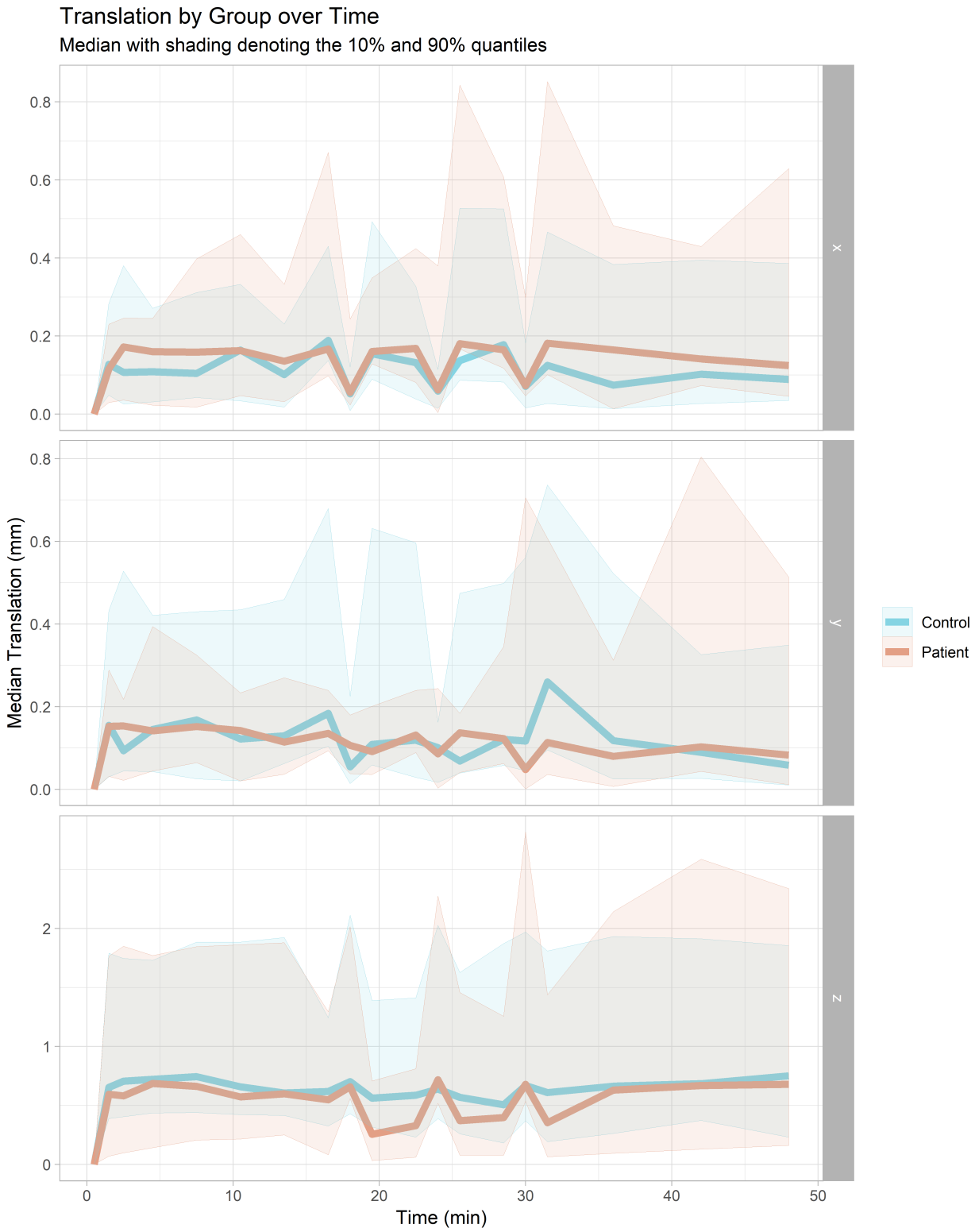

Figure 11: Translation parameters compared between patient and control groups

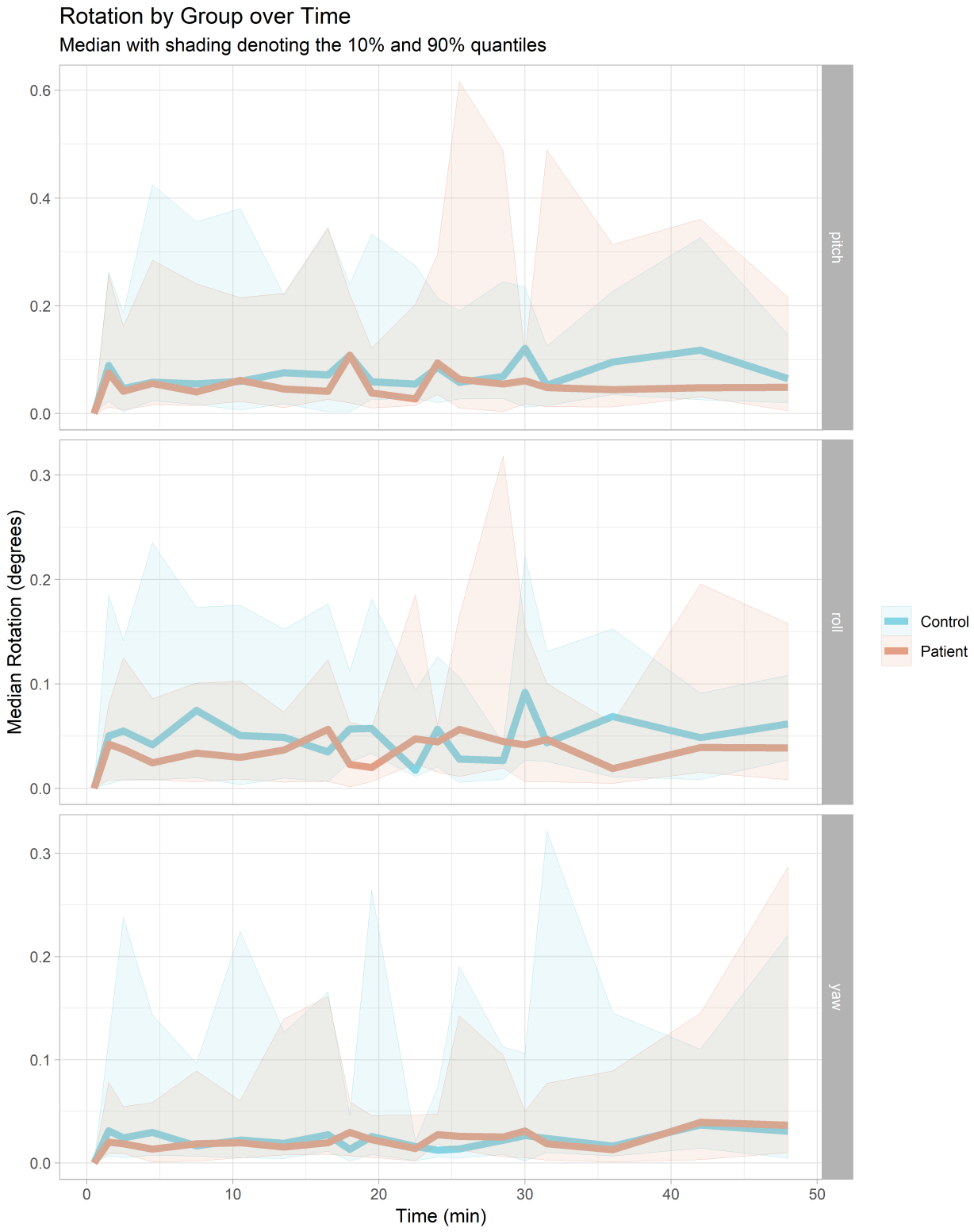

Figure 12: Rotation parameters compared between patient and control groups

In conclusion, there do not appear to be any systematic differences in translation or rotation parameters between groups.

Biological confounders

Age

As expected from the literature, we observed strong negative associations between D1R BP_{ND} and age in both regions. Age will, as planned, be included in the statistical model.

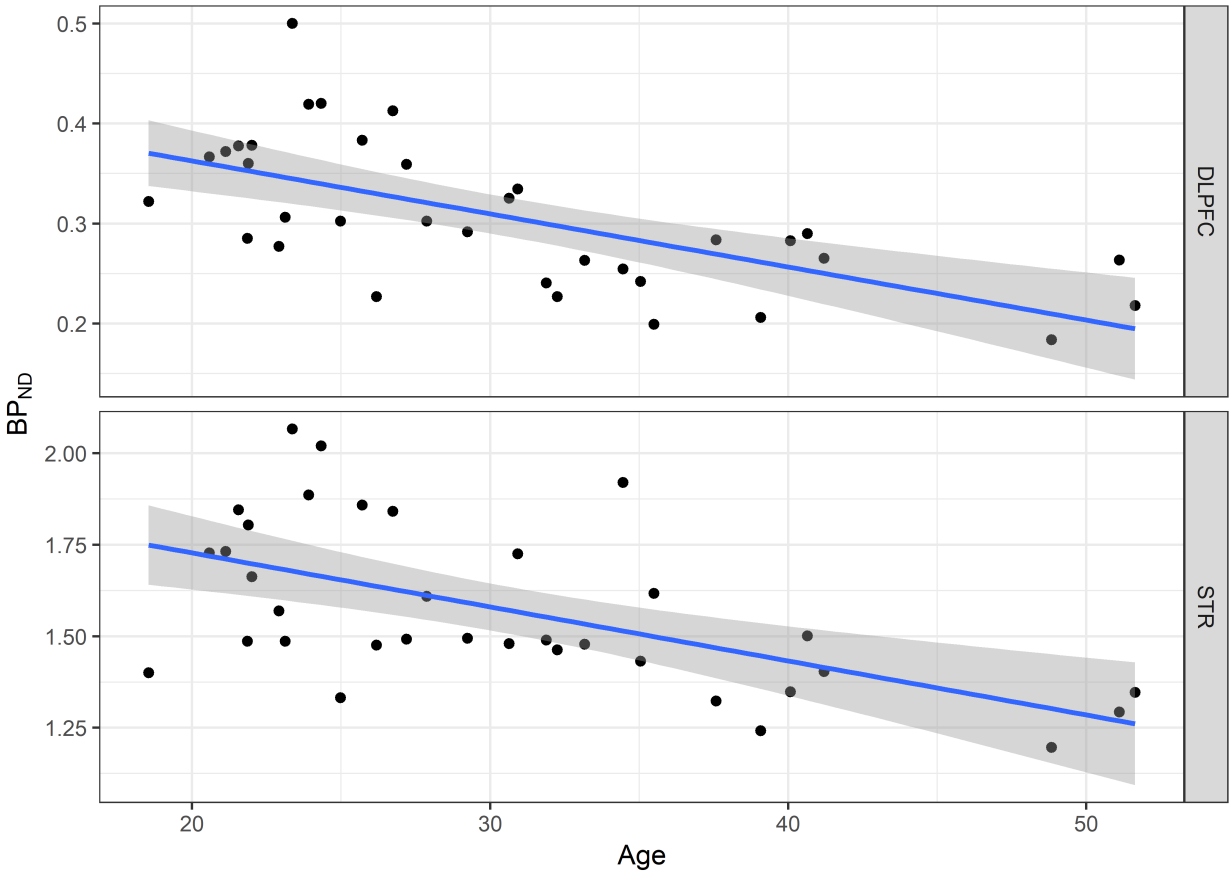

Figure 13: BP_{ND} was negatively associated with age, corresponding to previous studies

Sex

Unexpectedly, we observed what appeared to be a difference between D1R BP_{ND} between males and females in the sample.

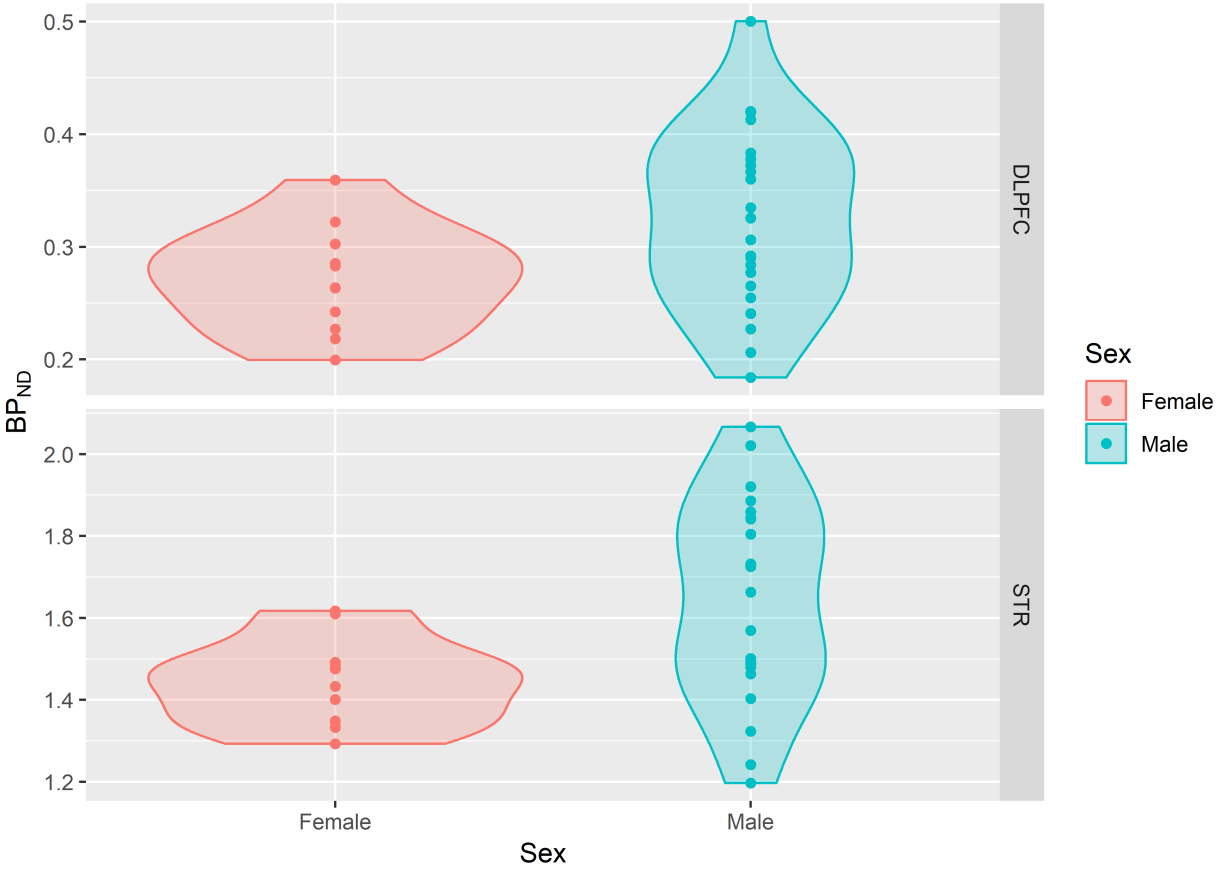

Figure 14: Sex comparison of BP_{ND} values

This difference was significant between groups.

estimate	estimate1	estimate2	statistic	parameter	p.value	method
0.198	1.641	1.443	3.354	32.214	0.002	Welch Two Sample t-test
0.053	0.325	0.272	2.493	32.403	0.018	Welch Two Sample t-test

In order to assess whether this was due to systematic differences in age between the samples of males and females, we assessed the age of the two groups.

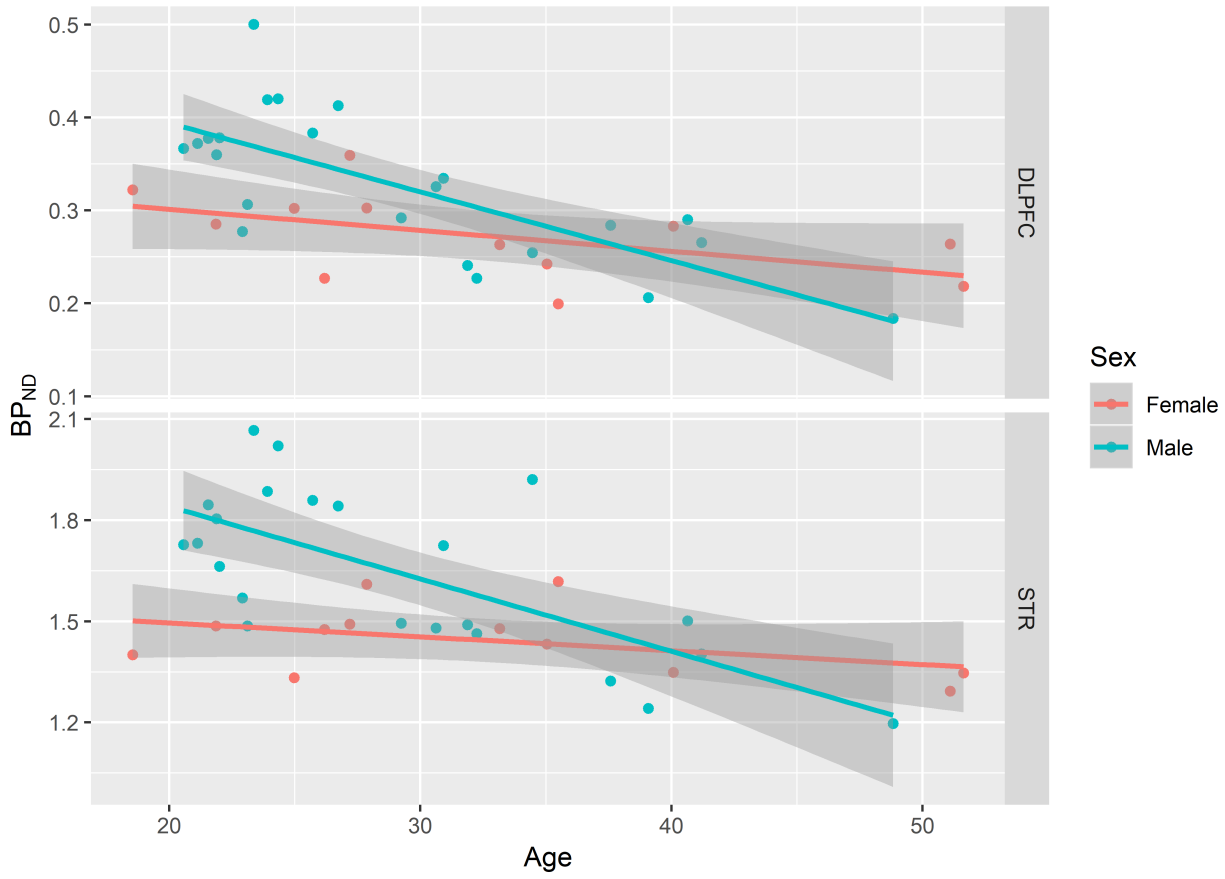

Figure 15: An apparent interaction between sex and age on BP_{ND} values

From this figure, it appeared that there may be an age × sex interaction. We tested this using multiple regression models, finding significant interaction effects in both regions.

Table 17: Multiple regression model to assess the interaction between age and sex for the DLPFC

term	estimate	std.error	statistic	p.value
(Intercept)	0.346	0.049	7.022	0.000
Age	-0.002	0.001	-1.568	0.127
SexMale	0.195	0.064	3.031	0.005
Age:SexMale	-0.005	0.002	-2.585	0.015

Table 18: Multiple regression model to assess the interaction between age and sex for the striatum

term	estimate	std.error	statistic	p.value
(Intercept)	1.578	0.154	10.224	0.000
Age	-0.004	0.004	-0.916	0.367
SexMale	0.692	0.202	3.431	0.002
Age:SexMale	-0.017	0.006	-2.791	0.009

term	estimate	std.error	statistic	p.value

Effects of sex on $[^{11}C]$SCH23390 BP_{ND} have, to our knowledge, not previously been reported. In order to assess whether this is a true effect which must be corrected for in the model, or whether it is a false positive and can be safely ignored, we tested for sex effects and this interaction effect in two other datasets with larger samples and larger age ranges.

Karolinska Behavioural PET Database

In the Karolinska Behavioural PET Database, we had access to a sample of n=40, consisting of 20 males and 20 females between the ages of 23 and 76, processed as described in Matheson et al. (2018), but using the whole frontal cortex instead of the DLPFC.

There was no evident age × sex interaction in this data set when plotted in the same way.

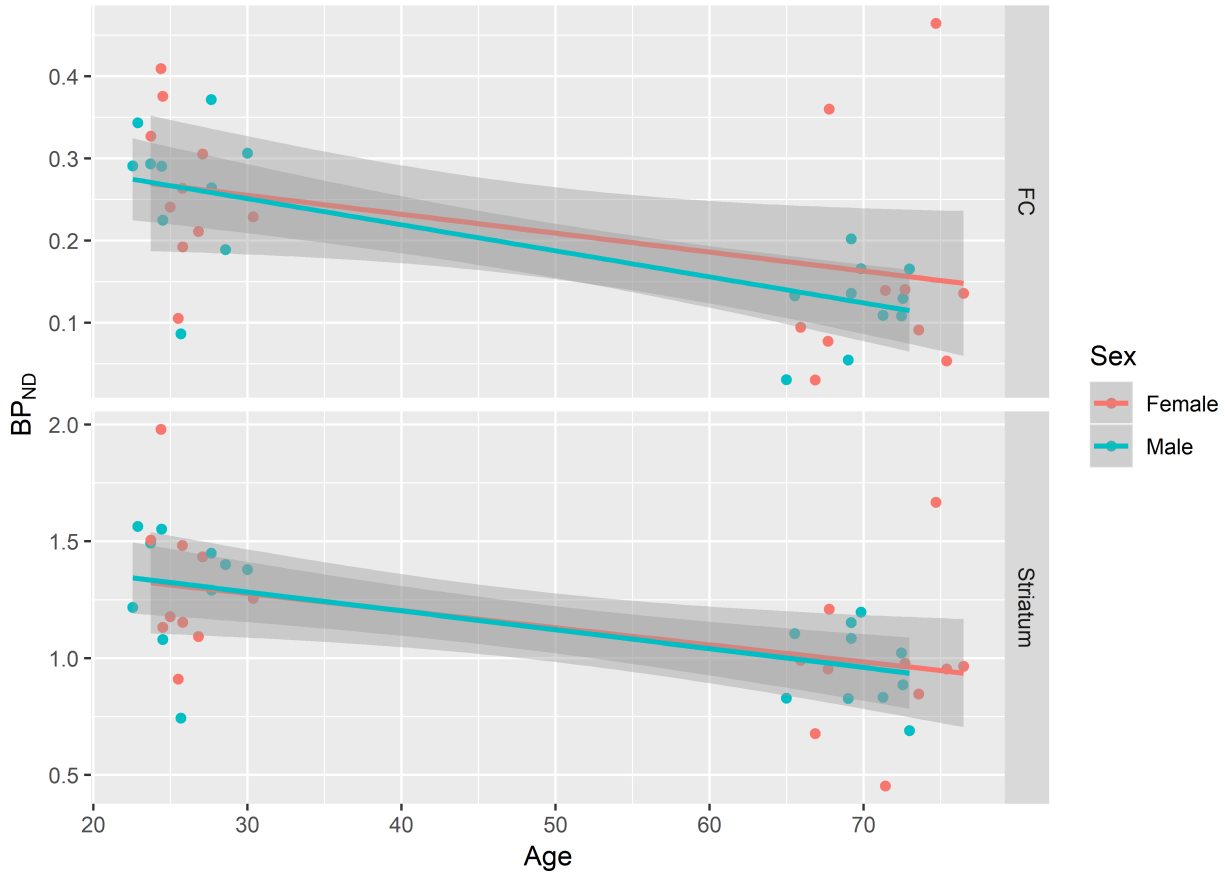

Figure 16: No apparent association between age and sex in the Karolinska Behavioural PET Database

The age × sex interaction was not significant in either the frontal cortex or striatum.

Table 19: Multiple regression model to assess the interaction between age and sex for the frontal cortex in the Karolinska Behavioural PET Database

term	estimate	std.error	statistic	p.value
(Intercept)	0.324	0.051	6.343	0.000
Age	-0.002	0.001	-2.419	0.021
SexMale	0.022	0.073	0.300	0.766
Age:SexMale	-0.001	0.001	-0.631	0.532

Table 20: Multiple regression model to assess the interaction between age and sex for the striatum in the Karolinska Behavioural PET Database

term	estimate	std.error	statistic	p.value
(Intercept)	1.497	0.141	10.651	0.000
Age	-0.007	0.003	-2.800	0.008
SexMale	0.029	0.200	0.143	0.887
Age:SexMale	-0.001	0.004	-0.198	0.844

Furthermore, there was no significant increase in BP_{ND} in males compared to females in either the frontal cortex or striatum.

Table 21: Multiple regression model to assess the effect of sex for the frontal cortex in the Karolinska Behavioural PET Database

term	estimate	std.error	statistic	p.value
(Intercept)	0.345	0.039	8.742	0.000
Age	-0.003	0.001	-4.009	0.000
SexMale	-0.020	0.031	-0.648	0.521

Table 22: Multiple regression model to assess the effect of sex for the striatum in the Karolinska Behavioural PET Database

term	estimate	std.error	statistic	p.value
(Intercept)	1.515	0.108	14.047	0.000
Age	-0.008	0.002	-4.141	0.000
SexMale	-0.007	0.084	-0.087	0.931

de Boer et al. (2017)

A recent paper reported $[^{11}C]$SCH23390 BP_{ND} in a sample of 56 individuals between the ages of 19 and 75 (De Boer et al., 2017). BP_{ND} values for the caudate and DLPFC were were shared with the paper, and sex of all participants was obtained through correspondence with the authors.

There was no evident age $\times$ sex interaction in this data set when plotted in the same way.

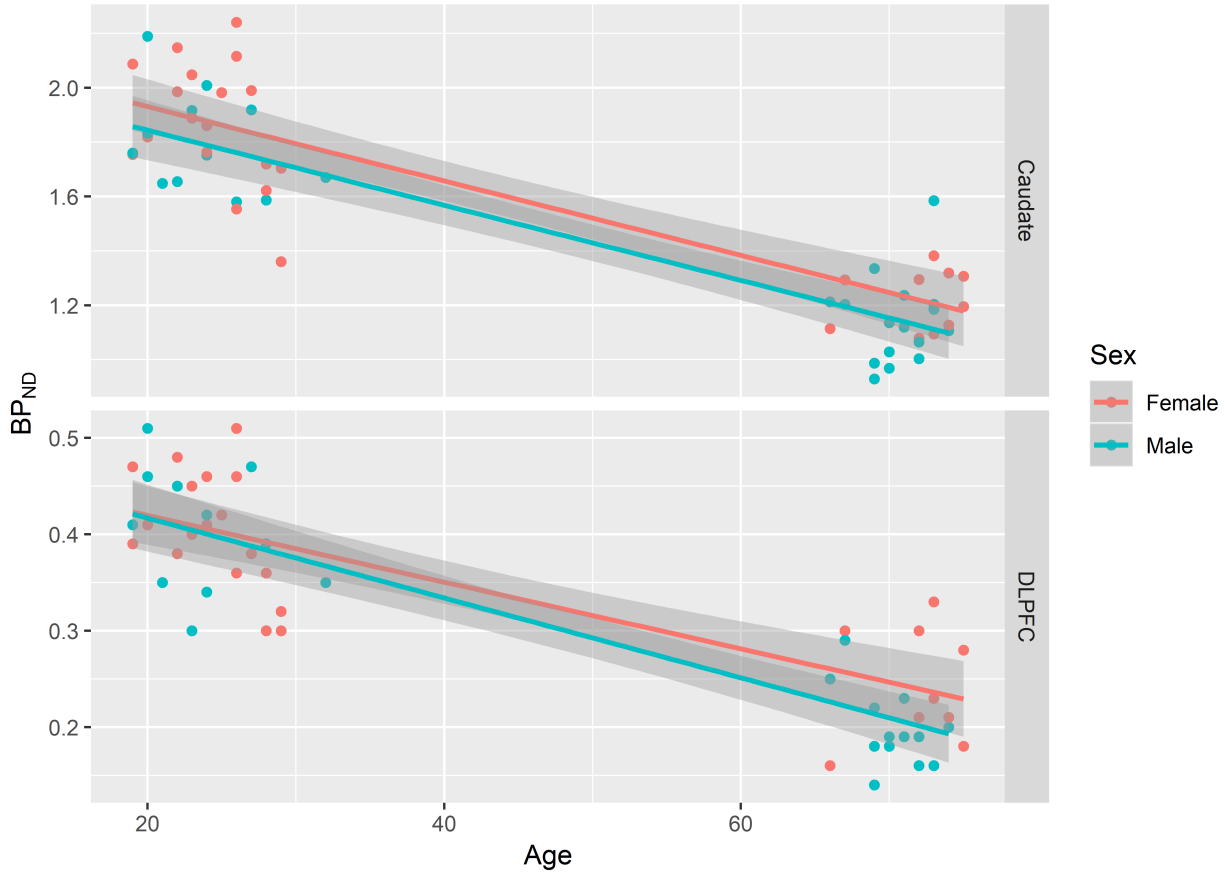

Figure 17: No apparent association between age and sex in de Boer et al., (2017)

The age × sex interaction was not significant in either the frontal cortex or striatum.

Table 23: Multiple regression model to assess the interaction between age and sex for the frontal cortex in de Boer et al., (2017)

term	estimate	std.error	statistic	p.value
(Intercept)	0.489	0.022	22.560	0.000
Age	-0.003	0.000	-7.575	0.000
SexMale	0.011	0.033	0.330	0.743
Age:SexMale	-0.001	0.001	-1.065	0.292

Table 24: Multiple regression model to assess the interaction between age and sex for the striatum in de Boer et al., (2017)

term	estimate	std.error	statistic	p.value
(Intercept)	2.204	0.070	31.427	0.000
Age	-0.014	0.001	-9.246	0.000
SexMale	-0.084	0.107	-0.786	0.436
Age:SexMale	0.000	0.002	-0.061	0.951

Furthermore, there was no significant increase in BP_{ND} in males compared to females in either the frontal cortex or striatum. In fact, the effect of sex was nearly significant in the opposite direction.

Table 25: Multiple regression model to assess the effect of sex for the frontal cortex in de Boer et al., (2017)

term	estimate	std.error	statistic	p.value
(Intercept)	0.503	0.017	29.610	0.000
Age	-0.004	0.000	-11.837	0.000
SexMale	-0.020	0.015	-1.349	0.183

Table 26: Multiple regression model to assess the effect of sex for the striatum in de Boer et al., (2017)

term	estimate	std.error	statistic	p.value
(Intercept)	2.207	0.054	40.553	0.000
Age	-0.014	0.001	-13.341	0.000
SexMale	-0.090	0.049	-1.856	0.069

Summary

In summary, there was no evidence in larger samples with wider age ranges to support either the influence of sex, or an age $\times$ sex interaction, on BP_{ND} estimates. For this reason, we conclude that this effect in our sample was most likely observed as a false positive due to overfitting the data. For this reason, the effects of sex and an age $\times$ sex interaction will not be included in the final regression model.

Supplementary Materials S3: Time Stability Analysis

Due to the different lengths of the PET measurements, we aimed to evaluate whether there were changes in the mean and standard deviation of BP_{ND} values at different PET lengths. Based on the results of this analysis, we planned to perform analysis as follows:

- If there were no change in mean BP_{ND} over time, then measurements with different lengths could be included in the analysis without requiring the application of a correction factor to their outcome values.
- If there would be a systematic change in the mean BP_{ND} value for PET measurements of different lengths, then the measurement length at which the standard deviation would be lowest would be considered to be most likely to be the most accurate. Longer measurements would be shortened to this length, and a correction factor for the systematic change in mean BP_{ND} would be applied to those measurements which were shorter.
- If this selected measurement length were longer than 51 minutes, the length of the majority of the PET measurements, then 51 minutes would be used to avoid applying a correction factor to more than half of the measurements.

First, we selected only those measurements for which 63 minutes of acquisition were recorded, and estimated BP_{ND} for measurement durations from 20 minutes to 63 minutes, and calculated their BP_{ND} proportional to their BP_{ND} for 63 minutes.

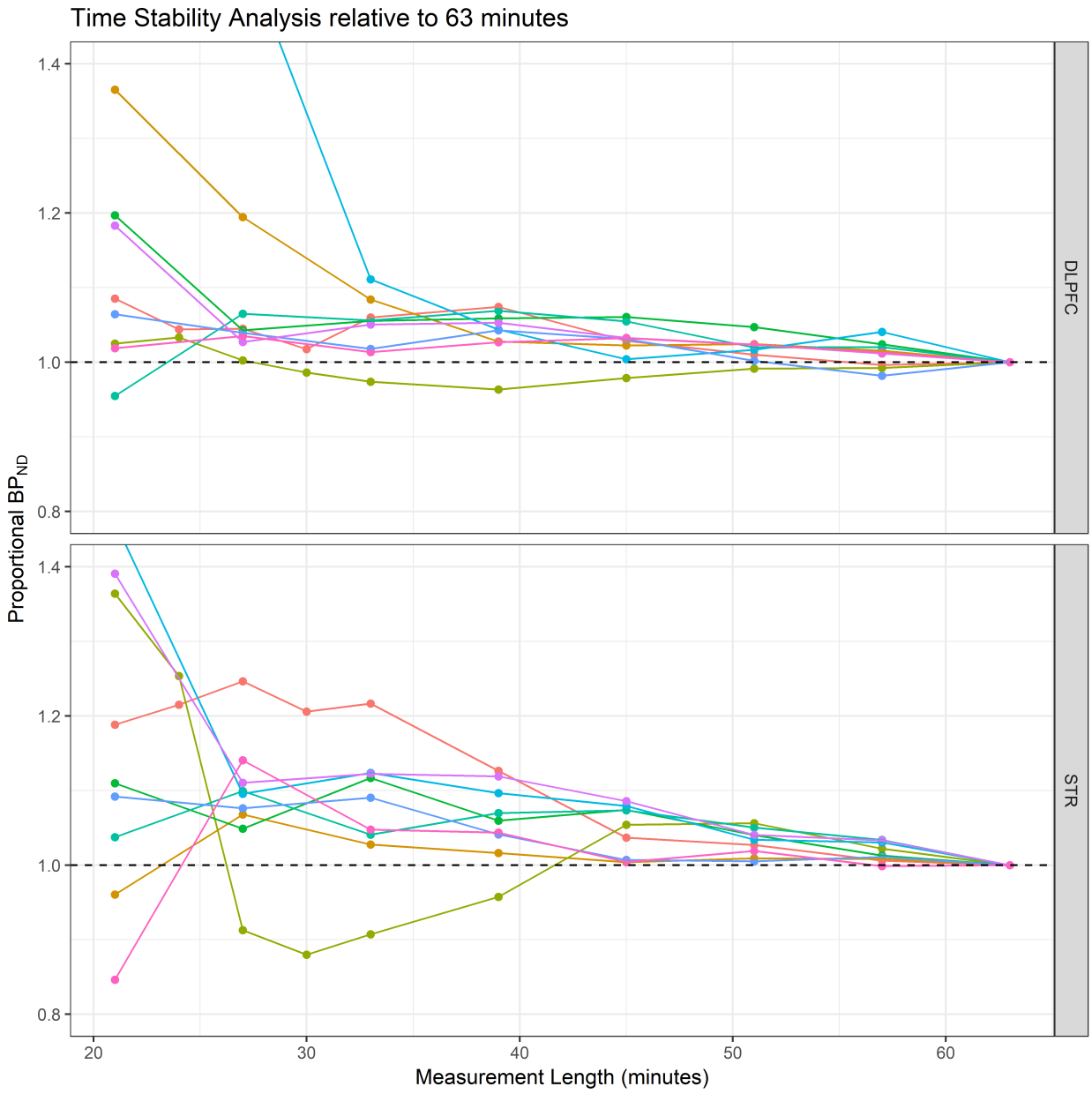

Figure 18: Time stability analysis of BP_{ND} for different measurement lengths relative to BP_{ND} for 63 minutes.

From this analysis, it was clear that $[^{11}C]$SCH23390 BP_{ND} did not show sufficient time stability to be used without modification for individuals with different lengths of measurement.

Next, we selected only those measurements for which at least 51 minutes of acquisition were recorded and estimated BP_{ND} for measurement durations from 25 minutest to 51 minutes to assess whether shorter measurements were also biased compared to 51 minute acquisitions. The same lack of time stability was apparent, meaning that a correction would need to be used.

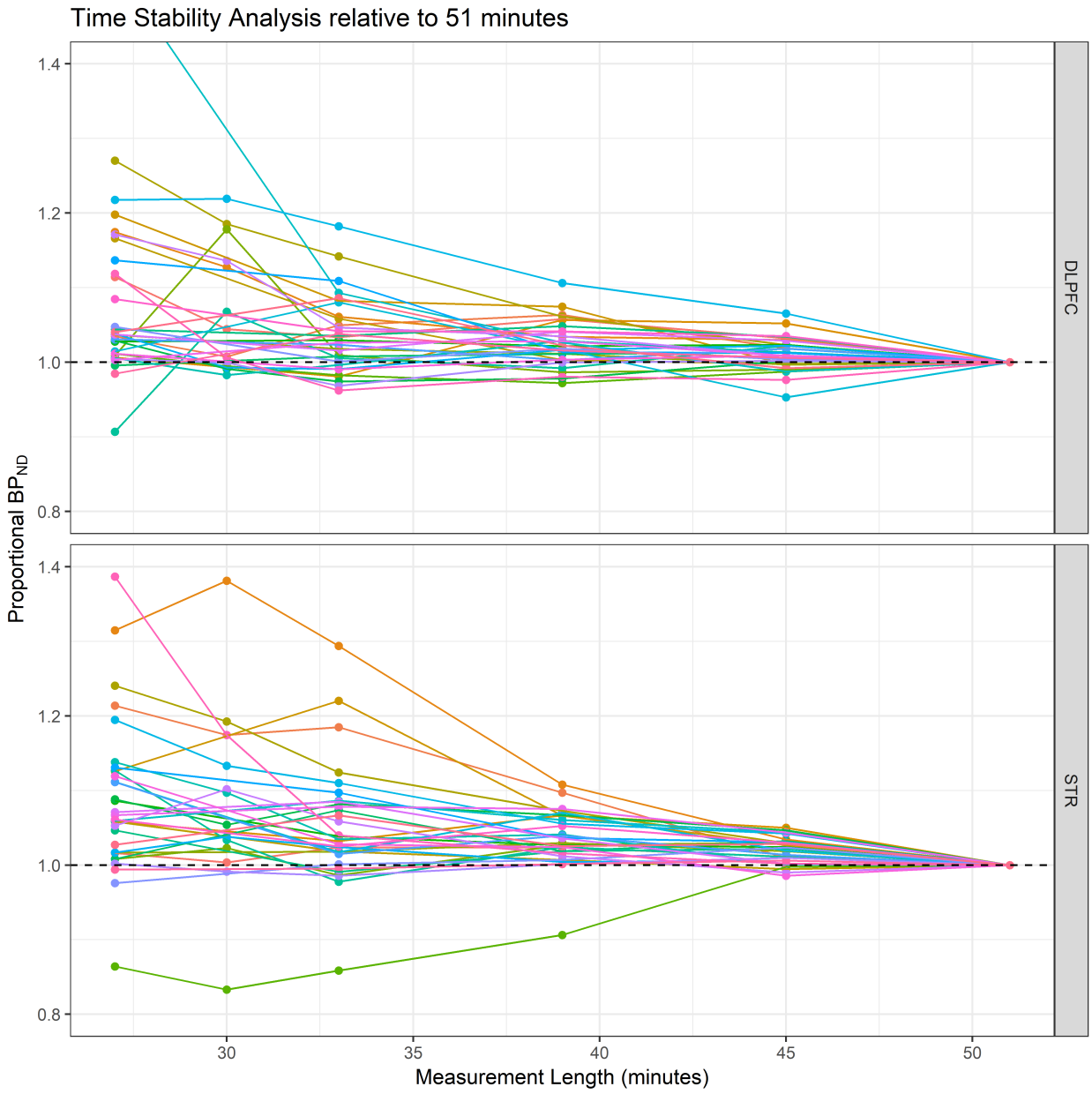

Figure 19: Time stability analysis of BP_{ND} for different measurement lengths relative to BP_{ND} for 51 minutes.

To assess the acquisition length for which the correction factor would be applied, we examined the standard deviation of the estimated BP_{ND} values for measurement durations from 25 minutes to 51 minutes, from which we observed a decreasing standard deviation with measurement duration. We therefore selected 51 minutes as the measurement length to set a correction factor for.

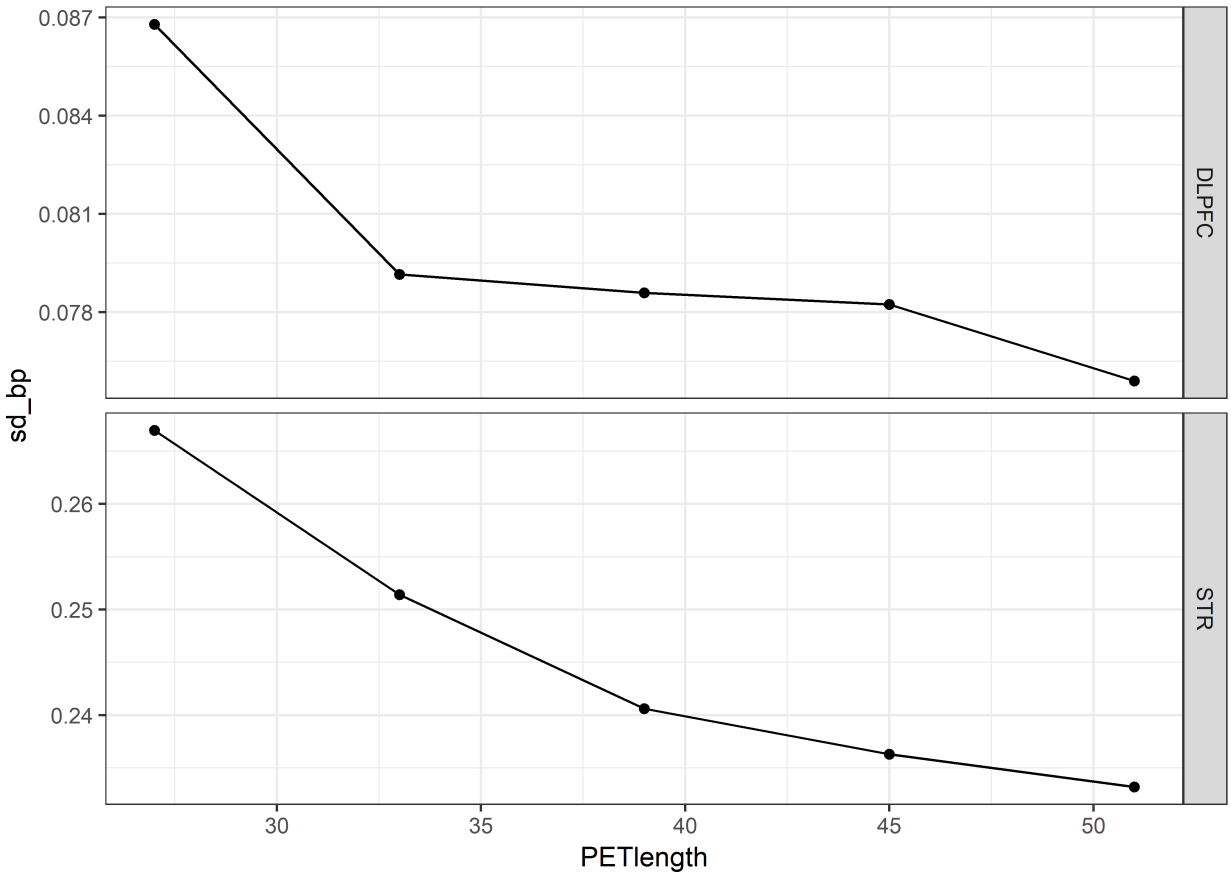

Figure 20: Standard deviation of BP_{ND} values for different measurement lengths

Correction factors were calculated as the mean proportional BP_{ND} value for each measurement duration shorter than 51 minutes. For the three measurements shorter than 51 minutes, corrected BP_{ND} values were calculated by dividing their estimated BP_{ND} values for their full measurement duration by the corresponding correction factor. The calculated correction factors were as follows.

Table 27: Mean and standard deviation of the proportional difference of BP_{ND} compared to BP_{ND} calculated for 51 minutes, as well as mean and standard deviation of BP_{ND} values for each length of PET measurement.

PETlength	ROI	mean_propdif	sd_propdif	mean_bp	sd_bp	n
27	DLPFC	1.078	0.115	0.334	0.087	32
33	DLPFC	1.035	0.051	0.320	0.079	32
39	DLPFC	1.024	0.030	0.317	0.079	32
45	DLPFC	1.011	0.022	0.313	0.078	32
51	DLPFC	1.000	0.000	0.309	0.076	32
27	STR	1.087	0.101	1.715	0.267	32
33	STR	1.052	0.078	1.661	0.251	32
39	STR	1.035	0.037	1.635	0.241	32
45	STR	1.019	0.017	1.611	0.236	32
51	STR	1.000	0.000	1.582	0.233	32

Supplementary Materials S4: Regression Coefficients

Below are presented the unstandardised regression coefficients.

Table 28: Striatum

term	estimate	std.error	statistic	p.value
(Intercept)	2.045	0.510	4.007	0.0
Age	-0.065	0.016	-3.970	0.0
CtrPat	-0.151	0.285	-0.530	0.6

Table 29: DLPFC

term	estimate	std.error	statistic	p.value
(Intercept)	2.344	0.445	5.268	0.000
Age	-0.067	0.014	-4.731	0.000
CtrPat	-0.583	0.249	-2.345	0.025

References

Andersson, J. L. R., Jenkinson, M., & Smith, S. M. (2007). Non-linear registration aka Spatial normalisation FMRIB Technial Report TR07JA2. *In Practice*, (June), 22. Retrieved from http://fmrib.medsci.ox.ac.uk/analysis/techrep/tr07ja2/tr07ja2.pdf

De Boer, L., Axelsson, J., Riklund, K., Nyberg, L., Dayan, P., Bäckman, L., & Guitart-Masip, M. (2017). Attenuation of dopamine-modulated prefrontal value signals underlies probabilistic reward learning deficits in old age. *eLife*, *6*.

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage*, *46*(1), 39–46.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327. Retrieved from http://arxiv.org/abs/NIHMS150003

Hirvonen, J., Erp, T. G. M. van, Huttunen, J., Aalto, S., Nagren, K., Huttunen, M., Lonnqvist, J., et al. (2006). Brain dopamine d1 receptors in twins discordant for schizophrenia. *Am J Psychiatry*, *163*(10), 1747–1753. Retrieved from http://dx.doi.org/10.1176/appi.ajp.163.10.1747{\%}5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/17012685 http://psychiatryonline.org/article.aspx?doi=10.1176/appi.ajp.163.10.1747 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cm

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. M. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841.

Kosaka, J., Takahashi, H., Ito, H., Takano, A., Fujimura, Y., Matsumoto, R., Nozaki, S., et al. (2010). Decreased binding of [11C]NNC112 and [11C]SCH23390 in patients with chronic schizophrenia. *Life Sciences*, *86*(21-22), 814–818. Elsevier Inc. Retrieved from http://dx.doi.org/10.1016/j.lfs.2010.03.018

Matheson, G. J., 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. *bioRxiv*. Retrieved from https://www.biorxiv.org/content/early/2018/05/14/321646

Matheson, G. J., Stenkrona, P., Cselényi, Z., Plavén-Sigray, P., Halldin, C., Farde, L., & Cervenka, S. (2017). 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*, *155*.

Poels, E. M. P., Girgis, R. R., Thompson, J. L., Slifstein, M., & Abi-Dargham, A. (2013). In vivo binding of the dopamine-1 receptor PET tracers [11C]NNC112 and [11C]SCH23390: A comparison study in individuals with schizophrenia. *Psychopharmacology*, *228*(1), 167–174.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155.

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

Tziortzi, A. C., Haber, S. N., Searle, G. E., Tsoumpas, C., Long, C. J., Shotbolt, P., Douaud, G., et al. (2014). Connectivity-based functional analysis of dopamine release in the striatum using diffusion-weighted MRI and positron emission tomography. *Cerebral Cortex*, *24*(5), 1165–1177.