Compulsivity and impulsivity traits linked to attenuated developmental fronto-striatal myelination trajectories

Supplementary Information

Content:

Supplementary Tables 1-5

Supplementary Math and Notes on Modelling

NSPN Consortium List

Supplementary Table 1. SPM Results testing for positive effects of time/visit or mean age on myelin-sensitive MT.

brain region	cluster size	p(FDR) voxelwise parametric	Z	p(FDR) voxelwise wild bootstrap	p(FWE) clusterwise wild bootstrap	p(unc)	x	y	z
cortical gray matter									
right angular gyrus	593628	3.51E-07	6.78	<0.002	<0.001	6.08E-12	51	-46	44
right posterior cingulate gyrus		3.51E-07	6.67			1.24E-11	16	-44	35
left orbitofrontal cortex		3.51E-07	6.58			2.30E-11	-18	-37	-17
right precuneus		3.51E-07	6.53			3.25E-11	8	-65	54
right supramarginal gyrus		3.51E-07	6.52			3.58E-11	61	-33	25
left transverse temporal gyrus		3.51E-07	6.39			8.43E-11	-52	-20	2
cortical white matter									
left posterior cingulate/precuneus	268806	4.1E-06	6.18	<0.004	0.002	3.23E-10	5	-58	56
left superior temporal gyrus		4.1E-06	6.11			5.05E-10	-58	-25	1
left supramarginal gyrus		4.11E-05	6.04			4.39E-10	-52	-56	26
right transverse temporal gyrus		4.24E-06	5.93			1.47E-09	49	-13	6
right angular gyrus		4.4E-06	5.86			2.3E-09	50	-48	50
left superior parietal lobule		5.05E-06	5.73			5.15E-09	-21	-51	70
left posterior superior frontal gyrus	1277	0.0013	3.48			2.55E-04	-8	-2	70
superior superior frontal gyrus		0.027	2.18			0.015	-20	18	61

SPM longitudinal SwE results table testing for positive time or age effects on MT in cortical gray and adjacent white matter accounting for covariates and confounds (cf. methods and supplementary notes on modelling, n=497/288 scans/subjects). Voxel resolution 1mm isotropic. Voxelwise FDR corrected ($p<0.05$) reporting peaks and clusters with up to 6 local maxima more than 24 mm apart, applied extent threshold k=500 voxel. We report p-values from non-parametric voxelwise FDR and p-values from non-parametric clusterwise FWE corrected inference (with cluster forming threshold 0.001). Notably, the number of permutations limits the precision of the lowest p-values which can be assessed. Using SwE covariance type 'modified', effective degrees of freedom per subject were found to be 0.9896.

Supplementary Table 2. SPM Results testing for negative (positive) effects of time/visit or mean age on local gray matter (white matter) volumes.

brain region	cluster size	p(FDR) voxelwise parametric	Z	p(unc)	x	y	z
cortical gray matter							
right middle temporal gyrus	572245	<0.001	9.02	9.4E-20	65	-42	4
left middle temporal gyrus		<0.001	8.89	9.4E-20	-66	-37	1
left middle frontal gyrus		<0.001	8.77	9.4E-20	-35	44	33
right superior temporal gyrus		<0.001	8.36	5.5E-17	61	-14	-7
right middle frontal gyrus		<0.001	8.35	5.5E-17	52	22	27
left angular gyrus		<0.001	8.00	6.1E-16	-55	-57	25
subcortical gray matter							
right cerebellum exterior	155308	7.06E-06	6.00	10.0E-10	20	-50	-13
left ventral striatum		7.06E-06	5.64	8.5E-09	-7	13	-10
right putamen		3.13E-05	4.94	3.99E-07	32	-1	8
thalamus		3.45E-05	4.90	4.6E-07	-2	-20	10
left putamen		4.1E-05	4.83	6.6E-07	-35	-14	-2
cortical white matter							
left upper pyramidal tract	13078	0.001	4.95	3.7E-07	-19	-19	46
left suppl. motor area		0.006	3.77	8.3E-05	-15	6	44
left post. insular		0.012	3.51	2.3E-04	-32	-11	12
left medial prec. gyrus		0.028	3.07	0.0011	-5	-18	67
right ant. medial prefrontal	3059	0.001	4.65	1.7E-06	13	42	-5
right precuneus	541	0.002	4.24	1.1E-05	14	-61	50
right upper pyramidal tract	6708	0.002	4.20	1.4E-05	23	-14	44
left lingual gyrus	811	0.007	3.7	1.1E-04	-19	-71	3
core white matter							
right pyramidal tract	5753	0.006	4.87	5.7E-07	19	-16	26
left pyramidal tract	6011	0.006	4.48	3.7E-06	-18	-20	46

SPM longitudinal SwE results table testing for negative time or age effects on local volumes in cortical and subcortical gray matter and positive effects in cortex adjacent and core white matter volume accounting for covariates and confounds (cf. methods and supplementary notes on modelling, n=494/285 scans/subjects). Voxel resolution 1mm isotropic. Voxelwise FDR corrected ($p<0.05$) reporting peaks and clusters with up to 6 local maxima more than 24 mm apart, applied extent threshold k=500 voxel. We report p-values from parametric voxelwise FDR. Using SwE covariance type 'modified', effective degrees of freedom per subject were found to be 0.9858.

Supplementary Table 3. SPM results for negative compulsivity by time/visit interaction on myelin-sensitive MT.

brain region	cluster size	p(FDR) voxelwise parametric	Z	p(FWE) tfce wild bootstrap	p(FWE) clusterwise wild bootstrap	p(unc)	x	y	z
cortical gray matter									
left superior frontal gyrus	55570	0.009	4.87	0.019	0.015	5.6E-07	-23	34	49
left precentral gyrus		0.009	4.73			1.1E-06	-58	11	27
left frontal operculum/ant. insula		0.009	4.59	0.018	0.015	2.3E-06	-55	10	-3
left postcentral gyrus		0.009	4.51			3.2E-06	-31	-31	73
left middle frontal gyrus		0.009	4.36			6.3E-06	-35	52	27
left ant. superior frontal gyrus		0.009	4.35	0.019		6.8E-06	-11	61	31
right postcentral gyrus	1789	0.009	4.13			1.8E-05	17	-41	77
right temporal pole	547	0.009	4.1			2E-05	40	7	-38
right superior frontal gyrus	11259	0.009	4.1	0.027		2.1E-05	18	1	58
right precentral gyrus		0.011	3.85			6.0E-05	26	-19	74
right ant. sup. frontal gyrus		0.010	4.07			2.4E-05	12	54	30
left post. middle temporal gyrus		0.010	3.88			5.1E-05	-50	-68	-9
cortical white matter									
right ant. superior frontal gyrus	700	0.05	4.66			1.6E-06	11	54	31
left superior frontal gyrus	12996	0.05	4.28	0.011		9.5E-06	-24	-4	64
left middle frontal gyrus		0.05	4.03	0.011		2.7E-05	-29	-21	53
left ant. midcingulate		0.05	3.74	0.011	0.02	9.2E-05	-25	1	37
left ant. superior frontal gyrus		0.05	3.58	0.011		1.7E-04	-7	59	30
right ant. middle frontal gyrus	726	0.05	4.01	0.011		3.0E-05	42	41	24
right posterior superior frontal gyrus	4331	0.05	3.63	0.011		1.4E-04	13	1	68
right ant. midcingulate		0.05	3.14	0.011		8.3E-04	10	13	54
right temporal pole	582	0.05	3.5			2.3E-04	43	9	-36
right post. middle frontal gyrus	731	0.05	3.29			5.1E-04	34	15	40

SPM longitudinal SwE results table testing for negative time by compulsivity interaction effects on MT in cortical gray and adjacent white matter accounting for covariates and confounds (cf. methods and supplementary notes on modelling, n=454/246 scans/subjects). Voxel resolution 1mm isotropic. Voxelwise FDR corrected ($p<0.05$) reporting peaks and clusters with up to 6 local maxima more than 24 mm apart, applied extent threshold k=500 voxel. We report FDR on whole-brain level as requested during revision and additionally performed Wild Bootstrapping using 999 straps for non-parametric inference using same model for originally submitted fronto-striatal analysis due to substantial speedup compared to whole-brain sampling in this large sample. We report p-values from non-parametric tfce FWE (Smith & Nichols, 2009 with E=0.5, H=2) and p-values from non-parametric clusterwise FWE corrected inference (with cluster forming threshold 0.001) for corresponding regions. Notably, the number of permutations limits the precision of the lowest p-values which can be assessed. Using SwE covariance type 'modified', effective degrees of freedom per subject were found to be 0.9861.

Supplementary Table 4. SPM results for negative impulsivity by time/visit interaction on myelin-sensitive MT.

brain region	cluster size	p(FDR) voxelwise parametric	Z	p(FWE) tfce wild bootstrap	p(FWE) clusterwise wild bootstrap	p(unc)	x	y	z
cortical gray matter									
left frontal operculum/ant. insula/IFG	79829	0.031	4.65	0.005	0.006	1.6E-06	-48	13	-4
right ant. medial PFC		0.031	4.13	0.005	0.041	1.8E-05	15	58	18
left precentral gyrus		0.031	4.08			2.2E-05	-24	-13	71
left midcingulate cortex		0.031	3.86	0.005	0.014	5.6E-05	-4	-1	42
left ant. medial PFC		0.031	3.64	0.005	0.046	1.3E-05	-14	44	15
right temporal pole	1862	0.031	3.93			4.3E-05	49	11	-35
right ant. insula/IFG	13978	0.031	3.89	0.005	0.016	4.9E-05	40	13	-10
right central operculum		0.032	3.25			5.7E-04	46	-2	15
right middle frontal gyrus		0.036	2.86	0.005		0.002	47	23	30
right precentral gyrus	1768	0.031	3.77			8.2E-05	40	7	-38
right superior parietal lobe	1817	0.031	3.73			9.6E-05	29	-39	62
left precentral gyrus	1047	0.031	3.65			1.3E-04	-45	-2	40
left inf. temporal gyrus	809	0.031	3.52			2.1E-05	-47	-55	-17
left superior parietal lobe	1746	0.031	3.5			2.3E-05	-26	-41	58
cortical white matter									
left inferior frontal/midcingulate cortex	162984	0.015	4.38	0.002	0.002	6.05E-06	-27	39	-2
left precentral sulcus/pyramidal		0.015	4.21	0.002		1.27E-05	-27	-17	44
right postcentral gyrus		0.015	3.84			6.2E-05	17	-43	72
left supramarginal gyrus		0.015	3.83			6.3E-05	-48	-34	44
left postcentral gyrus		0.015	3.7			1.1E-04	-18	-41	64
left ant. medial PFC		0.015	3.69	0.002	0.002	1.1E-05	-12	47	20
right post. middle frontal gyrus		0.015	3.59			1.6E-04	35	4	46
right ant. insula/IFG		0.015	3.58	0.002		1.7E-04	42	49	4
left ant. sup. temporal gyrus	750	0.015	3.08			3.8E-04	-50	5	-11
left middle temporal gyrus	991	0.05	3.14			0.001	-60	-43	-16
right temporal pole	582	0.05	3.5			2.3E-04	43	9	-36
right middle frontal gyrus	731	0.05	3.29	0.002		5.1E-04	34	15	40

SPM longitudinal SwE results table testing for negative time by impulsivity interaction effects on MT in cortical gray and adjacent white matter accounting for covariates and confounds (cf. methods and supplementary notes on modelling, n=497/288 scans/subjects). Voxel resolution 1mm isotropic. Voxelwise FDR corrected ($p<0.05$) reporting peaks and clusters with up to 8 local maxima more than 24 mm apart, applied extent threshold k=500 voxel. We report FDR on whole-brain level as requested during revision and additionally performed Wild Bootstrapping using 999 straps for non-parametric inference using same model for originally submitted fronto-striatal analysis due to substantial speedup compared to whole-brain sampling in this large sample. We report p-

values from non-parametric tfce FWE (Smith & Nichols, 2009 with E=0.5, H=2) and p-values from non-parametric clusterwise FWE corrected inference (with cluster forming threshold 0.001) for corresponding regions. Notably, the number of permutations limits the precision of the lowest p-values which can be assessed. Using SwE covariance type ‘modified’, effective degrees of freedom per subject were found to be 0.9792.

Supplementary Table 5. SPM Results testing for negative main effects of impulsivity on myelin-sensitive MT

brain region	cluster size	Z	p(FDR) voxelwise wild bootstrap	p(FWE) clusterwise wild bootstrap	p(unc)	x	y	z
cortical white matter								
right medial orbitofrontal cortex	7286	4.96	0.037	0.004	3.5E-7	13	44	-22
right orbital inferior frontal gyrus		4.15	0.037	0.004	1.7E-5	26	26	-8
right superior frontal gyrus		3.24	0.057	0.004	0.001	20	59	-2
left inferior temporal sulcus	910	4.44	0.037	0.042	4.5E-6	-51	-28	-20
left medial frontal cortex	2799	4.34	0.037	0.012	7.2E-6	-8	42	-16
right middle frontal gyrus	868	4.29	0.037	0.043	8.7E-6	41	42	17
right superior temporal gyrus	801	3.74	0.037	0.045	9.2E-5	44	-16	-11
left posterior inf. temp. gyrus	626	3.52	0.037	0.048	2.2E-4	-37	-61	-3
core white matter								
left external capsule	1726	4.10	0.016	0.021	2.0E-05	-29	16	0
right external capsule	2264	4.09	0.016	0.019	2.1E-05	24	22	-10

SPM longitudinal SwE results table testing for negative effects of impulsivity on MT in cortex adjacent white matter accounting for covariates and confounds (cf. methods and supplementary notes on modelling, n=497/288 scans/subjects). Voxel resolution 1mm isotropic. Voxelwise FDR corrected ($p<0.05$) reporting peaks and clusters with up to 6 local maxima more than 24 mm apart, applied extent threshold k=500 voxel. We report p-values from non-parametric voxelwise FDR and p-values from non-parametric clusterwise FWE corrected inference (with cluster forming threshold 0.001). Notably, the number of permutations limits the precision of the lowest p-values which can be assessed. Using SwE covariance type 'modified', effective degrees of freedom per subject were found to be 0.9792.

Supplementary Math and Notes on Modelling.

Linear-mixed effects (LME) modelling for questionnaire scores and global brain parameters

LME is a widely used analysis technique for univariate (scalar) or mass-univariate neuroimaging data with available repeated measurements[1,2]. This typically assumes modelling subject *i*'s data as $y_i = X_i\beta + Z_i b_i + \varepsilon_i$, with fixed effects design X_i, random effects design Z_i, and residuals ε_i. Residuals were assumed to have mean zero expectation and error covariance matric $\sigma^2 I$. Moreover, random effects follow $b_i \sim N(0,\sigma^2 D)$. Throughout this paper, we used LME as implemented in MATLAB (R2016b; function 'fitlmematrix') using Restricted Maximum Likelihood (ReML) optimization of the above model under the full covariance with Cholesky parametrization, i.e. $D=LL^T$, with lower triangular L. Notably, implementing linear trajectory models in this framework, allows for intercept slope correlations ($d_{12}=d_{21}\neq 0$). Thus, inclusion of random slopes results in two additional hyperparameters, which might result in over-parametrization for a given dataset. We therefore compared this model to one with uncorrelated random effects using simulated likelihood-tests (using MATLAB's 'compare' function). Comparison of multiple models with varying numbers of random effects was performed using likelihood ratio tests ($p<0.05$). For inference on fixed effects about main effects, interactions, quadratic components, we used linear contrasts.

Longitudinal image modelling using the Sandwich Estimator (SwE)

The prevailing longitudinal image analysis method (in the sense of mass-univariate approaches) in context of brain development is LME. However, this approach often makes restrictive or unrealistic assumptions, e.g. compound symmetry[3]. Moreover, LME is based on iterative algorithms, which are not guaranteed to converge in all voxels of the search space. In this study, we therefore performed modelling of the image data using a state-of-the-art analysis method recently introduced as the longitudinal Sandwich Estimator (SwE)[4], http://www.nisox.org/Software/SwE, SPM toolbox). Using this so-called marginal model one describes individual *i*'s data as $y_i = X_i\beta + \varepsilon_i^*$, i.e. based on only fixed effects design matrix X_i. The randomness is treated as nuisance and modelled by marginal error terms ε_i^* with mean 0 and positive semi-definite covariance V_i. Marginal models do not require specification of random-effects and allow unbiased population-average inference and predictions about brain change in certain sub-groups or in relation to covariates[4].

The modelling approach first estimates the parameters of interest with a simple ordinary least squares $\hat{\beta} = (\sum_{i=1}^{m} X_i' W_i X_i)^{-1} \sum_{i=1}^{m} X_i' W_i y_i$ while working matrix W_i is assumed to be identity in this application. Second, the approach estimates variances/covariances with the specific estimator, which accounts for the within-subject correlation existing in the longitudinal image observations. The covariance of the parameter estimate is $Cov(\hat{\beta}) = (\sum_{i=1}^{m} X_i' W_i X_i)^{-1}(\sum_{i=1}^{m} X_i' W_i V_i W_i X_i)(\sum_{i=1}^{m} X_i' W_i X_i)^{-1}$ with separate estimates of subject *i*'s covariance V_i accounting for the fact that we have 2 or 3 scans per person available in our application within the accelerated longitudinal design. Due to our large available sample, no small sample size corrections were necessary. The SwE method has been shown to allow unbiased estimation of within- and between-subject effects as columns of the same design matrix using longitudinal image observations. Inference on linear hypothesises $\mathcal{H}_0: C\beta = 0$ are based on Wald tests (for further details on inference see Guillaume et al.[4]). More specifically, the sandwich estimator type (in SwE toolbox) was set to 'modified' using approximately homogenized subgroups based on age, gender, and follow-up-time. Parametric inference results were found to be comparable to using estimator type 'classic'. Small sample adjustment was set to 'C2' and degrees of freedom approximation type 'III'.

Because of its advantages and efficiency for the *mass-univariate* voxel-based modelling scenario, we here apply SwE instead of LME. One limitation of SwE, however, is that it does not provide explicit access to change parameters on the individual level, e.g. each subject's rate of change. However, it does allow for analysis of individual differences of change by assessing group-level effects of within- or between-subject covariates. Therefore, we applied classic LME when modelling *univariate* frontal lobe aggregate (i.e. global as opposed to voxel-based) MT (Supplementary Fig. 7a-b) and for assessing differences across subjects and changes over time/visit of the PI-WSUR and BIS scores (using available follow ups, see section above).

Assessing compulsivity and impulsivity traits

To assess whether there are considerable developmental changes of traits, LME modelling was applied to total scores of PI-WSUR and BIS for which longitudinal follow-ups were available. We modelled total scores with fixed effects *X=[intercept, time/visit, age_mean, sex, socioeconomic status]* and individual random effects Z_i*=[intercept]* for both domains. The rationale behind separation of effects of study time/visit and mean age of subjects is outlined below. Alternative models including random slopes did not provide significant improvements for PI-WSUR (BIS) (assuming uncorrelated p-value=.76 (.80) or correlated p-value=.52 (.27) random effects). Testing for effects of study time/visit (in years) did not demonstrate maturational effects for PI-WSUR (BIS) (p=.35 (.23)). The powerful questionnaire sample did not reveal indications for substantial group-level change of both traits over the study visits. Moreover, models including additional random slopes (to capture individual differences of change) did perform worse than random intercept only models in likelihood ratio tests. More follow-

ups with longer time-intervals and/or larger sample sizes might provide access to group- and individual level trajectories of compulsivity and impulsivity in future work. In contrast to ongoing change observed over course of the longitudinal study, between-subject differences across subjects at baseline could have emerged during previous development, from conception onwards. Given our particular sample, longitudinal analysis suggested that major portions of the observed variability of both scores was sufficiently captured by between-subject differences, and that the remaining variability might be largely attributable to unstructured noise.

In order to further operationalize the compulsivity differences across subjects, in addition to PI-WSUR, we also collected OCI-R scores (primarily cross-sectionally). Both total scores correlated highly (Supplementary Fig. 1a, r=.764, p<.001). Having no evidence for compulsivity change over study visits, a more reliable characterization of subject's trait-compulsivity should be accomplished using all items in both questionnaires. Defining the compulsivity trait in terms of a latent variable (based on shared variance) generalizing across two questionnaires is expected to reduce (a) influence by test-specific biases and (b) influence of measurement errors[5]. To define a common trait score for compulsivity, we performed a principal component analysis on all available items (Supplemental Fig. 1b-c, as implemented in MATLAB R2016b, using an alternating least-squares algorithm to account for missingness). In all analyses, the first component (PC1, items' contribution shown in Supplementary Fig. 1c) is subsequently named 'compulsivity' and correlated highly with both, OCI-R (Supplementary Fig. 1d, r=0.982, p<0.001) and PI-WSUR (Supplementary Fig. 1e, r=0.810, p<0.001). The fact that the derived compulsivity trait correlated higher with both questionnaires than the two with each other (Supplementary Fig. 1a) suggests that factor very well captures the main commonalities behind them.

To operationalize impulsivity differences based on BIS scores (available longitudinally), each subject's average of all available measurements of a person could be used. Alternatively, each person's random-intercept from the LME model (as introduced above) is (a) less susceptible to influence of noise than simple averages and (b) shrinks towards the group mean for subjects with less observations. LME-based estimates have been shown to outperform noisier unbiased estimators[6] and thus are expected to be more powerful for explorative brain-behaviour analysis. If not stated otherwise (e.g. Supplementary Fig. 7c), in this study 'impulsivity' refers to individual random-intercepts from LME modelling of the BIS score.

Longitudinal design specification: Separation of study time/visits and mean age

Following common guidelines for longitudinal design specification (freesurfer: https://surfer.nmr.mgh.harvard.edu/fswiki/LinearMixedEffectsModels, Bernal-Rusiel et al.[13], SPM analysis practice http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/ and Guillaume et al.[4]) the longitudinal design matrix $X = [intercept, X_{eoi}, X_{eni}, X_{conf}]$ (obtained from stacked individual design matrices X_i) was carefully specified to include effects of interest X_{eoi}, covariates of no interest X_{eni} and potentially confounding effects X_{conf}. The particular choice of these matrices is motivated and specified as follows. The clear distinction between variables of interest and no interest is partially arbitrary and rather following naming conventions since both sets of factors are similarly included, estimated and its contribution modelled explicitly, the variables of interest are close to our main hypothesis focused on in this paper and will be reported within given space constraints of main and supplemental results. Explicit focus on covariates of no interest (e.g. sex and socioeconomic status effects) might be addressed in follow-up papers using similar models.

As outlined in detail in Guillaume et al.[4], for a given longitudinal observational design, the participants' covariate '*age at scan*' (e.g. available at *0, 6, 12, 18* months after baseline) has two separable components, a between-subject component and a within-subject component, indistinguishable with cross-sectional samples. The between-subject component (further referred to as *age_mean*) is purely cross-sectional and can be obtained by considering only each participant's mean age across all their acquired scans. In contrast, the within-subject component (further referred to as *time/visit*) is purely longitudinal and considers the actual effects of study time/visit on each individual, and follows by subtracting the mean age of a participant from its age covariate. In this study, we consistently decomposed a centred '*age at scan*' variable (subtracting its overall mean $\overline{age}$) using this idea, and the separate (within- and between-) subject components for subject i (*=1, ..., 288*) and visit/timepoint j (*=1, 2, 3*) were obtained as follows

$$age_{ij} - \overline{age} = (age_{ij} - \overline{age}_i) + (\overline{age}_i - \overline{age}) = time_{ij} + age\_mean_i$$

with centred study time/visit variable $time_{ij}$ and the mean age of each participant age_mean_i. If we would only use the original '*age at scan*' covariate in the longitudinal design matrix X, we would implicitly assume that the effects on the images are the same for both above components. However, the effects of within- and between subjects component can be very different as shown by Neuhaus and Kalbfleisch[7]. Cross-sectional components might be affected by cohort effects (i.e. different populations) which then would lead to biased estimates. More specifically, (1) overestimation of ageing-related effects due to e.g. cohort differences[8,9] (2) underestimation e.g. due to selective attrition, training effects and (3) even sign reversals of between- and within-subjects effects have been observed in previous analyses[11]. Furthermore, (4) cross-sectional data complicate addressing some of the fundamental questions in the field: How do brain changes (over time) differ across individuals- or groups, typically in relation to a third variable[12] e.g. trait dimensions?

Thus, here we follow the recommendation of Guillaume et al.[4] and systematically split the age covariate into its between- and within-subject components and consistently include both in the design matrix *X*. As shown by Guillaume et al., this separation also improves the efficiency of the longitudinal Sandwich Estimator (SwE) modelling method (applied below) when assuming an identity working covariance matrix, showing that the SwE is nearly as efficient as Generalized Least Squares estimates. This finding also suggests the importance of centring covariates when inference is made on the intercepts, time etc. by avoiding increasing variance due to correlated regressors such as the intercept, time and other covariates of no interest. In addition to the presented voxel-based neuroimaging analysis in this paper, the outlined splitting of *age_mean* and *time/visit* was also applied in all linear-mixed effects (LME) models of presented in this study.

Finally, our findings indicated coarse consistency of longitudinal *time/visit* (Supplementary Fig. 3a) and cross-sectional *age_mean* (Supplementary Fig. 3b) effects of brain maturation, with noticeable regional differences with respect to topography and statistics. In this sample, statistical *age_mean* effects were found to be stronger, while actual effect sizes of growth were comparable. A detailed discussion of observed discrepancies goes beyond the scope of this paper (for more details see [11]). Some differences might be explained using various statistical arguments, ranging from (a) local noise level; the (b) ground truth ratio of within- and between subject variability; and the (c) presence of sampling biases or cohort effects for the specific marker of interest. Moreover, we expect a crucial contribution of applied image processing techniques, likely to introduce affect between- and within subject variability differently.

Longitudinal image analysis: Modelling effects of interest

The considered effects of interest of this particular study are (a) the study *time/visit* and *age_mean* (b) the effects of compulsivity and impulsivity *traits* and their interaction with study *time/visit*, i.e. X_{eoi} = *[time, age_mean, time by trait interaction, trait]*. For example, time interactions allow testing for different rates of change (over study) in subjects with higher or lower expression of the traits. In this study, we use *time/visit* to refer to the individually-centred study *time* as introduced above (after separation from *age_mean* differences, i.e. time of scan in years relative to each subject's mean age over all visits), and *trait* refers to subject-specific scores characterizing the average impulsivity and compulsivity of participants (introduced and derived in a previous section) and all modelled interactions used products based on centred variables. Using LME analysis of the longitudinal questionnaire data (and accounting for covariates X_{eni}, see above) we observed that in our sample, most of the variance of compulsivity and impulsivity scores was found on the between- rather than the within-subjects level. Thus, in our main mass-univariate VBQ analysis presented in this paper (Fig. 3 & 4), *traits* are understood as a time-independent variable characterizing either compulsivity or impulsivity between-subject differences observed in the available questionnaire data (including follow-ups if available). To avoid biases (and sensitivity loss) due to potential shared variance with *age_mean*, the *trait* regressors were corrected for *age_mean* trends before entering the design. Notably, the main effect of *trait* was also included as effects of interest for analysis of *trait*-related brain differences (shown Fig. 3c and 4b) independent of development (in terms of *age_mean* and *time*). More intuitively, these main effects of *trait* might be expected to be observed in a purely cross-sectional study of a large sample of (>300) individuals with *trait* as regressor. Additionally, this renders the observed *time/visit by trait interaction* effects independent from existing *trait* differences across individuals at baseline, i.e. statistically decoupling past differences (which might be related to *trait*) from ongoing change (which also might be related to *trait* but potentially in same or even different brain networks).

Longitudinal image analysis: Modelling non-linearities

The applied longitudinal design matrix X = *[intercept,* X_{eoi}, X_{eni}, X_{conf}*]* enables focus on effects of development as well as effects of *traits* and their interactions with time/visit or *mean_age* (in terms of purely correlative associations). One might argue that developmental brain growth at different ends of the NSPN study age range (14-26 years) is likely to be different. Consequently, a similar effect of study time/visit (assuming 'homogeneity of change' of younger and older participants) is therefore not likely. This homogeneity would be violated if (a) there are severe non-linearities (e.g. a deceleration of growth) of trajectories in networks of interest; or if (b) population differences across the age domain have emerged through biased sampling. If there is evidence for (a) for the chosen myelin marker (MT), similarly to other interactions with study time/visit, systematic effects can be modelled, tested, and (if undesired) also corrected within the same design matrix.

Using our large sample, we therefore tested for indications of non-linear brain changes implying *time/visit* by *age_mean* interactions or quadratic/cubic effects of *age_mean* in late adolescence. For example, observing a positive *time by age_mean* interaction would reveal evidence for deceleration of change in early adulthood (as illustrated in Supplementary Fig. 3d & 4b). Very mild negative quadratic effects were seen for cross-sectional white matter volume expansion (Supplementary Fig. 4c). Findings indicated very minor influence of non-linearities (given the level of residual noise), even much weaker for MT than for volumes. This suggested a rather consistent direction of growth/shrinkage (MT/volume) across the age range, which can be sufficiently captured by linear and quadratic trajectories for MT and volumes respectively. Similar to traditional morphometric brain

markers, future longitudinal studies with more power and especially wider age range (e.g. full second and third decade) might be promising to explore long-term trajectory shapes of novel quantitative imaging markers and repeated measures.

Longitudinal image analysis: Modelling effects of covariates and confounders

Given that quantitative MRI measures during development are expected to vary with further covariates known by design, known variability across subjects, undesired effects of intra-scan motion, scanning site and global brain variables, we explicitly model variability due to other covariates (or 'effects of no interest' for the main hypothesis) X_{eni} *= [sex, socioeconomic status]* as well as confounders X_{conf} *= [total intracranial volume, site1, site2, motion proxy]* using centred variables to minimize biases induced by correlated regressors.

There is strong prior expectations about the sexual-dimorphism in adolescent brain development[13–15], which resulted in always including *sex* differences as a covariate in all models. Similarly, we account for potential effects due to differences in socioeconomic status across individuals in all analyses using the official neighbourhood poverty index that is based on the proportion of neighbourhood households that live below official poverty income in the area of residence[16]. Additionally, we included a regressor accounting for ethnicity (self-reported 'white' vs 'other' origin; correlation with impulsivity r=.2, compulsivity r=-.11 across entire NSPN cohort). As outlined above, in order to model potential in-homogeneity of change across individuals we tested for (and if substantial additionally included) all first order interactions of *time/visit* or *age_mean* with *trait, sex* and *socioeconomic status* respectively.

A well-known confounder for analysis of local structural change and neuroanatomic correlates of cognitive or other traits is that results might be induced by substantial differences in brain size. We therefore accounted for variations in a subject-specific estimate of total intracranial volume *(ticv)* in all models. Although not intended per construction, quantitative MRI features might be partially influenced by normalization and local tissue morphometry as well. This resulted in including the *ticv* consistently as covariate for both VBM and VBQ. Finally, unintended measurement variations due to site-specific acquisition differences were modelled using two centred indicator variables encoding each participant's scan location (two in Cambridge, one in London).

To account for differences between scanners, we used scanner sites as an additional covariate with (Wolfson Brain Imaging Centre) WBIC as reference site (377 out of 500 scans were acquired at WBIC site). Thus, two additional offsets to describe potential mean deviations from WBIC were included if a scan took place at UCL and CBU sites. This procedure also accounts for change of scanners over visits, because these covariates were implemented in a fully time-varying design correcting the estimated offset for each scan independently.

Furthermore, we studied time-varying motion proxy as an additional regressor in our longitudinal SwE modelling of MT. We carried out additional analyses to assess the effects of motion and to control for potential influences thereof. Recent work from Castella et al. (2018)[17] studied prospective motion correction in the context of Multi Parameter Mapping (MPMs). More specifically, during MPM generation a multi-echo model is estimated (www.hmri.info). The standard deviation parameter of R2* residuals in white matter areas (SDR2*) has been shown to be an accurate proxy of individual's movement during a scan (Castella: Fig. 2, 4 & 5). We replicate this link in our analysis and observed that SDR2* correlated with motion regressors from a separate resting-state fMRI scan (Supplementary Fig. 8a). We therefore assessed this movement proxy in all scans of the MT-weighted sequence across all subjects and available time-points. As predicted by Castella et al., SDR2* was a useful index to determine motion artefacts (as detected by detailed visual inspection) in our longitudinal sample. We therefore excluded (on top of other artefacts) those scans with the 10% highest values of SDR2* (above value of 5). A threshold of 10% was set as it revealed a steep increase in our motion proxy for subjects exceeding this threshold. Testing for potential SDR2* effects, we observed locally very restricted, but significant, positive effects of this movement proxy on MT ($p<0.05$, voxelwise FDR), which we report in Supplementary Fig. 8b. To control for residual effects of motion-induced variability we included SDR2* in all MT analyses as confounding variable, rendering the presented associations linearly independent of this proxy of absolute motion. Importantly, only few findings changed slightly after inclusion of motion-related effects.

Finally, we examined additional potentially confounding effects using further covariates. Based on a suggested effect of alcohol use[18,19], we used an alcohol consumption index as a potential confounding covariate. This measure was derived from the Drugs, Alcohol and Self Injury questionnaire (DASI[20,21]) and showed only weak associations with our dimensions of interest across the NSPN cohort (impulsivity r=.13, compulsivity r=-.02). Moreover, we used measures of drug use (also derived from DASI; impulsivity r=.04, compulsivity r=-.04), general IQ (as measured using WASI[22]), and ethnicity. More specifically, we examined how general cognitive functioning is related to the association of compulsivity/impulsivity and brain maturation. As an index of cognitive function, we used a measure of general IQ as measured using WASI intelligence test (unstandardized raw scores; matrix and vocabulary subtests) and observed a developmental growth of cognitive abilities with age and over study visits (Supplementary Fig. 9a). Moreover, we aimed to predict cognitive differences and changes using compulsivity/impulsivity traits and did not observe any main effect or trait by visit/time interactions, suggesting independence of these measures (Supplementary Fig. 9b-c). Finally, we also controlled for fully time-varying IQ

scores (either age-corrected or age-uncorrected) of both subtests and did not observe any alteration of the impulsivity and compulsivity MT-effects presented in main results of our study (Supplementary Fig. 9d), suggesting that reported effects are independent of general cognitive ability as captured by IQ. More generally, we did not find substantial effects of any of above covariates on our analyses of interest (cf. Supplementary Fig. 9d).

Longitudinal brain-behavioural analysis using sandwich estimator: correlated change

A key hypothesis of this study is that a higher expression of the considered trait might be reflective of an impaired/altered growth of MT in fronto-striatal areas. These *trait*-associated late-adolescent changes might build upon earlier developmental alterations and even further increase individual differences of *traits* in subsequent development into early adulthood. In the above described analyses on local and frontal global brain changes, we considered *trait* differences to predict brain trajectories (since generative brain models being a widely applied standard in SPM). Notably, this direction of prediction (i.e. brain measures as dependent variables rather than traits) does not reflect our assumptions about causality since this study is focused on associative/correlative effects. The operationalisation of compulsivity and impulsivity used in all above analyses was based on assessing between-subject differences. Although the performed LME analyses suggested that very large portions of the structured (non-residual) variance of compulsivity and impulsivity scores observed in our longitudinal sample was capturing reliable between-subject differences. Our trait approach was reasonable and does not preclude a characterization of impulsivity state changes over time points in more powerful designs or using more reliable assessments. Here we finally aimed to explore whether the within-subject variations of BIS scores available over study *time/visits* might complement and support trait-based analyses.

We pursued two separate analyses. First, we conducted an ROI-based LME analysis to test for indications of 'correlated changes' using IFG MT changes (in terms of individual intercepts and slopes) as predictors for longitudinal impulsivity scores. Bilateral IFG ROI definition was based on the above neuromorphometrics atlas a priori, not making any use of the obtained SwE cluster in previous analysis to avoid any selection biases. The LME model used here directly extends the model (using same estimation procedures) used above. We specified fixed effects *X=[intercept, time, time by IFG_slope interaction, IFG_icpt, age_mean, sex, socioeconomic status]* and random effects Z_i*=[intercept]*. *IFG_icpt* and *IFG_slope* refer to the obtained intercept and slope parameters from the IFG ROI during regional-level summary statistics analysis. We performed a correlated change analysis using LME fixed effects, but for visualisation of the correlation of brain-behaviour rates of change across subjects, individual random effects slopes were obtained from an empty model without brain predictors, i.e. *X=[intercept, time, age_mean, sex, socioeconomic status]*, and Z_i*=[intercept,* $time_i$*]*, and resulting impulsivity slope estimates were plotted over the *IFG_slope* from regional MT summary statistics.

Second, we ran an exploratory voxel-wise correlated change analysis. As suggested by Guillaume et al.[4], time-varying BIS scores were decomposed in purely within- and between subjects components and entered as regressors in voxel-wise SwE modelling of myelin-sensitive MT (in addition to covariates *time/visits*, *age_mean*, *sex*, interactions and confounds) in fronto-striatal areas. We further tested for negative effects of (within-subject) score changes predicting MT changes for both PI-WSUR (which had more longitudinal data than OCI-R) and BIS scores (with tendencies observed for the latter illustrated in Supplementary Fig. 7c).

1. Pinheiro, J. & Bates, D. *Mixed-Effects Models in S and S-PLUS*. (Springer Science & Business Media, 2000).
2. Bernal-Rusiel, J. L. *et al.* Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *NeuroImage* **66**, 249–260 (2013).
3. Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. *Longitudinal Data Analysis*. (Chapman and Hall/CRC, 2008).
4. Guillaume, B. *et al.* Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. *NeuroImage* **94**, 287–302 (2014).
5. Bollen, K. A. Latent Variables in Psychology and the Social Sciences. *http://dx.doi.org/10.1146/annurev.psych.53.100901.135239* (2003). Available at: http://www.annualreviews.org/doi/10.1146/annurev.psych.53.100901.135239. (Accessed: 14th June 2017)
6. Robinson, G. K. That BLUP is a Good Thing: The Estimation of Random Effects. *Stat. Sci.* **6**, 15–32 (1991).
7. Neuhaus, J. M. & Kalbfleisch, J. D. Between- and Within-Cluster Covariate Effects in the Analysis of Clustered Data. *Biometrics* **54**, 638–645 (1998).
8. Hoffman, L., Hofer, S. M. & Sliwinski, M. J. On the confounds among retest gains and age-cohort differences in the estimation of within-person change in longitudinal studies: a simulation study. *Psychol. Aging* **26**, 778–791 (2011).
9. Sliwinski, M., Hoffman, L. & Hofer, S. M. Evaluating Convergence of Within-Person Change and Between-Person Age Differences in Age-Heterogeneous Longitudinal Studies. *Res. Hum. Dev.* **7**, 45–60 (2010).

10. Willis, S. L. & Schaie, K. W. *Practical Intelligence: Nature and Origins of Competence in the Everyday World*. (Cambridge University Press, 1986).
11. Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J. & Borsboom, D. Simpson's paradox in psychological science: a practical guide. *Front. Psychol.* **4**, 513 (2013).
12. Raz, N. & Lindenberger, U. Only time will tell: cross-sectional studies offer no solution to the age-brain-cognition triangle: comment on Salthouse (2011). *Psychol. Bull.* **137**, 790–795 (2011).
13. Blakemore, S.-J. Imaging brain development: the adolescent brain. *NeuroImage* **61**, 397–406 (2012).
14. Mills, K. L. *et al.* Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. *NeuroImage* **141**, 273–281 (2016).
15. Ziegler, G., Ridgway, G. R., Blakemore, S.-J., Ashburner, J. & Penny, W. Multivariate dynamical modelling of structural change during development. *NeuroImage* **147**, 746–762 (2017).
16. Personal and household finances - Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/. (Accessed: 17th October 2018)
17. Castella, R. *et al.* Controlling motion artefact levels in MR images by suspending data acquisition during periods of head motion. *Magn. Reson. Med.* (2018). doi:10.1002/mrm.27214
18. Whelan, R. *et al.* Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature* **512**, 185–189 (2014).
19. O'Halloran, L., Nymberg, C., Jollans, L., Garavan, H. & Whelan, R. The potential of neuroimaging for identifying predictors of adolescent alcohol use initiation and misuse. *Addict. Abingdon Engl.* **112**, 719–726 (2017).
20. Kiddle, B. *et al.* Cohort profile: The NSPN 2400 Cohort: a developmental sample supporting the Wellcome Trust NeuroScience in Psychiatry Network. *Int. J. Epidemiol.* (2017). doi:10.1093/ije/dyx117
21. Clair, M. C. S. *et al.* Characterising the latent structure and organisation of self-reported thoughts, feelings and behaviours in adolescents and young adults. *PLOS ONE* **12**, e0175381 (2017).
22. Wechsler, D. *Wechsler Abbreviated Scale of Intelligence*. (The Psychological Corporation: Harcourt Brace & Company, 1999).

Neuroscience in Psychiatry Network (NSPN) Consortium author list

Principal investigators:
Edward T. Bullmore (CI from 01/01/2017)
Raymond J. Dolan
Ian Goodyer (CI until 01/01/2017)
Peter Fonagy
Peter B. Jones

NSPN (funded) staff:
Michael Moutoussis
Tobias U. Hauser
Sharon Neufeld
Rafael Romero-Garcia
Michelle St Clair
Petra Vértes
Kirstie Whitaker
Becky Inkster
Gita Prabhu
Cinly Ooi
Umar Toseeb
Barry Widmer
Junaid Bhatti
Laura Villis
Ayesha Alrumaithi
Sarah Birt
Aislinn Bowler
Kalia Cleridou
Hina Dadabhoy
Emma Davies
Ashlyn Firkins
Sian Granville
Elizabeth Harding
Alexandra Hopkins
Daniel Isaacs
Janchai King
Danae Kokorikou
Christina Maurice
Cleo McIntosh
Jessica Memarzia
Harriet Mills
Ciara O'Donnell
Sara Pantaleone
Jenny Scott

Affiliated scientists:
Pasco Fearon
John Suckling
Anne-Laura van Harmelen
Rogier Kievit