White Matter Microstructure and Its Relation to Longitudinal Measures of Depressive Symptoms in Mid-late Life

Supplementary Information

Table of Contents

Supplementary Methods

Supplementary Results

Figures S1-S7

Tables S1-S8

Supplementary References

Supplementary Methods

PHQ-4 questionnaire

PHQ-4 (Patient Health Questionaire-4) questions include: "Frequency of depressed mood in last 2 weeks", "Frequency of unenthusiasm/disinterest in last 2 weeks", "Frequency of tenseness/restlessness in last 2 weeks" and "Frequency of tiredness/lethargy in last 2 weeks". This questionnaire assesses depression-related symptoms within a 2-week timeframe. The sum of the scores was calculated to indicate depressive symptoms.

The mean time lag between the first and second occasion was 4.29 years with a standard deviation of 0.94 years. Between the second and third occasion, mean time lag was 3.32 years with a standard deviation of 1.19 years. Between the third and the final occasion, mean time lag was 0.003 years with a standard deviation of 1.10 years.

For each measure of depressive symptoms derived from cross-sectional assessments, the mean time lag of mean level of depressive symptoms was 8.04 years (sd=1.22 years), the variability of depressive symptoms was 8.09 years (sd=1.14 years), and slope of longitudinal trajectory was 8.26 years (sd=1.11 years). The correlations between longitudinal measures and time lag was very small and therefore this variable was not included in the main model (r ranged from 2.41×10^{-4} to 0.027).

dMRI measures

All imaging data was acquired using a 3T Siemens Skyra (software platform VD13) machine, using a standard ("monopolar") Stejskal-Tanner pulse sequence. FSL packages were used for data preprocessing and microstructure estimation (1). Pre-processing included correction for eddy

currents, head-motion, and gradient distortion, using the Eddy tool (2). The processed tracts included 12 bilateral tracts that has a value for each brain hemisphere (acoustic radiation, anterior thalamic radiation, cingulate gyrus part of cingulum, corticospinal tract, inferior fronto-occipital fasciculus, inferior longitudinal fasciculus, medial lemniscus, parahippocampal part of cingulum, posterior thalamic radiation, superior longitudinal fasciculus, superior thalamic radiation and uncinate fasciculus) and 3 unilateral tracts (forceps major, forceps minor and middle cerebellar peduncle).

Depressive symptoms

For the growth curve model, we used the 'growth' function from the 'lavaan' package [\(http://lavaan.ugent.be/tutorial/growth.html\)](http://lavaan.ugent.be/tutorial/growth.html) in R (3). Scaled age at each assessment was controlled for. The growth curve model showed good fit to the data (CFI = 0.988 , TLI = 0.984 , RMSEA = 0.031 , SRMR = 0.014 , Chi-square (17) = 103.563 with a p<0.001). Longitudinal change within the whole population was in a negative direction but did not reach to significance (β = -0.094, p = 0.164). Both the intercept (β= 7.366, p < 0.001) and variance (β=1.521, p < 0.001) of the mean slope of growth curve model were significant. Each individual's slope of longitudinal trajectory was estimated for further analysis by using the 'predict' function in 'lavaan'.

Covariates

In addition to age, age² and gender, we also included scanner positions for all three axis, alcohol consumption, smoking status and stressful life events. The covariates except for age and gender

will be explained in detail below. All the covariates described here were acquired with the imaging assessments. See also Table S1, S6 and S7.

The scanner position was used for controlling for systematic change in the static magnetic field (the last four fields in [http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=110\)](http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=110). These proxies for scanner position showed minimal correlation with our white matter phenotypes, but in order to achieve a better estimated model, we chose to include them in our models.

Alcohol consumption was self-reported weekly consumption which was used in a published paper on the overall UK Biobank sample of about 500k people (4). We used a slightly different approach to exclude impossible numbers. In the referenced study, they excluded values over 5 standard deviations from mean, and we employed their values as upper and low thresholds instead of calculating our own standard deviations and mean in the subsample with imaging data, because we have a much smaller sample, which may introduce more noise and exclude an excessive amount of people.

For smoking status, we used the self-reported smoking status information [\(http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20116\)](http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20116). Participants could choose from one of the four options: (a) current smoker, (b) previous smoker, (c) non-smoker and (d) prefer not to answer. There were NAs for those did not answer. As the number of people who chose 'prefer not to answer' was very small, we did not transfer this into NA so to maximize our sample size. We treated this covariate as a categorical variable in our model. For the sensitivity analysis shown in Table S6 and S7, in order to make it easier for demonstration, we transferred it into a numeric variable (current smoker = 2, previous smoker = 1, non-smoker = 0, prefer not to answer = NA). This still represents the effects of smoking to depressive symptoms and white matter

microstructure, as it is generally believed that there should be a gradient effect from current smoker to non-smoker.

Stressful life events described the number of events happened within 2 years before scanning session [\(http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6145\)](http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6145). Items include: serious illness, injury or assault to oneself, death of a close relative, death of a spouse or partner, marital separation/divorce and financial difficulties.

Other behavioural measures included in the present study

All the behavioural variables were collected along with the imaging assessment unless it was collected within other categories such as online-follow up questionnaires that were not collected at any of the UK Biobank assessment centres.

Variables collected with the imaging assessment are listed below (relevant URL in the data showcase website shown in the brackets). For all the answers for the items listed below, "Prefer not to answer" and "Do not know" were recoded as NA.

- 1. Insomnia [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=1200\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=1200), as indicated by 1=Never/rarely, 2=Sometimes, and 3=Usually.
- 2. Smoking status [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20116\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20116), with 0=Never, 1=Previous, and 2=Current.
- 3. Recent pains in last month [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6159\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6159), with 1=any type of pain reported and 0=None of the above. This variable was set as binary and logistic regression was used for its association tests. Therefore log odds ratios were reported (see Table 1).

- 4. Hand grip strength was derived as the mean of left- and right-hand grip strength [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=46,](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=46) [http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=47\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=47).
- 5. g.Cognition. A selection of variables for cognitive functions wasincluded to generate a g factor that represents the variance of general cognitive ability. Cognitive tasks include: Reaction time [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20023\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20023), Verbal-numeric reasoning (also referred to as fluid intelligence in UK Biobank data dictionary: [http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20016\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20016), Trail making (derived by Trail B – Trail A: [http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6350,](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6350) [http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6348\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6348), Pair matching [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=399\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=399), Digit symbol substitution [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=23324\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=23324), Matrix pattern completion [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6373\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6373) and Tower rearranging [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=21004\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=21004). A principal component analysis (PCA) was performed and the score of the first unrotated principal component was extracted as a measure of g factor for cognitive functions. The first component explained 34.5% of the total variance. Absolute correlation loadings for each item ranged from 0.350 to 0.691 (loadings for reaction time and pair matching were negative, as for these items a smaller reaction time is better).
- **6.** g.Cognition for processing speed. An independent PCA was conducted on a selected range of the above tasks that has explicated stated measuring processing speed. The tasks included were: Reaction time, Trail making which tests visual processing speed, Digit symbol substitution which tests complex processing speed, and Pair matching, Matrix pattern completion, Tower rearranging and Verbal-numeric reasoning as participants were instructed to give answers in a limited time frame. Similar as g.Cognition, these measures used the score of the first unrotated principal component. The first principal component explained 42.2% of total variance. Absolute correlation loadings ranged from 0.553 to 0.734 (loading for reaction time was negative and all the others positive).

7. Neuroticism. Fields from 1920 to 2030 in the URL <http://biobank.ctsu.ox.ac.uk/showcase/label.cgi?id=100060> were used, and a total number of items with a 'Yes' answer for each participant was calculated as the neuroticism score.

Other variables that were not acquired with the imaging assessment were also often considered as lifetime variables that largely remain stable over time. These include the age of onset for depression [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20433\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20433), household income [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=738\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=738), educational attainment [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6138,](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6138) 1=College or University degree, 0=others that are less than College or University degree), Townsend Index [\(http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=189,](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=189) 1=the least deprived tertile, 2=the second tertile, 3=the most deprived tertile).

Supplementary Results

Associations between measures of depressive symptoms and NODDI measures

In general, results for ISOVF showed the largest resemblance with results for MD (see Figure S3). Here we report the results for all NODDI measures for each measure of depressive symptoms.

Cross-sectional assessment of depressive symptoms

For general variances of NODDI measures, higher cross-sectional depressive symptoms were associated with higher global ISOVF, higher ISOVF in gAF, gTR and gPF (β ranged from 0.018 to 0.029, p_{corr} ranged from 0.012 to 1.83 \times 10⁻⁵). Higher cross-sectional measure was also associated

with lower OD in gTR (β = -0.016, $p_{corr} = 0.047$).

Tract-specific associations were mainly found in ISOVF. Higher cross-sectional depressive symptoms was associated with higher ISOVF in anterior thalamic radiations, cingulate gyrus part of cingulum, parahippocampal part of cingulum, corticospinal tract, inferior fronto-occipital fasciculus, inferior longitudinal fasciculus, superior longitudinal fasciculus, superior thalamic radiation, uncinate fasciculus and middle cerebellar peduncle (β ranged from 0.015 to 0.039, p_{corr} ranged from 0.048 to 6.45×10^{-8}). For other NODDI measures, higher cross-sectional measure was positively associated with OD in medial lemniscus (β = 0.018, p_{corr} = 0.050) and negatively associated with OD in superior thalamic radiation ($β = -0.018$, $p_{corr} = 0.050$).

No general or tract-specific associations were found for ICVF (p_{corr} >0.314).

Longitudinal trajectory of depressive symptoms

For general variance in NODDI measures, a higher slope of longitudinal increase of depressive symptoms was associated with lower OD in global variance and gPF (β ranged from-0.035 to -0.039, p_{corr} ranged from 0.046 to 0.045). No general variance in ISOVF (p_{corr} >0.052) or ICVF (p_{corr} >0.580) showed a significant association with longitudinal slope.

For specific tracts, a steeper slope of longitudinal trajectory was associated with higher ISOVF in corticospinal tract and superior thalamic radiation (β ranged from 0.029 to 0.036, p_{corr} ranged from 0.040 to 0.025), as well as lower OD in inferior longitudinal fasciculus (β = -0.051, p_{corr} = 0.004). No tract measure was found associated with the slope of the longitudinal growth curve for ICVF measures (all $p_{corr} > 0.314$).

The mean of depressive symptoms derived from multiple assessments

Associations regarding mean depressive symptoms were mainly shown in ISOVF. Higher mean depressive symptoms were associated with higher ISOVF in global variance, gAF, gTR and gPF (β ranged from 0.014 to 0.027, p_{corr} ranged from 0.046 to 3.29 \times 10⁻⁴). It is also associated with higher gAF OD (β = 0.016, p_{corr} = 0.047) and lower gTR OD (β = -0.016, p_{corr} = 0.049). Tract-wise analysis showed that ISOVF of anterior thalamic radiations, cingulate gyrus part of cingulum, corticospinal tract, superior longitudinal fasciculus, superior thalamic radiation, uncinate fasciculus and middle cerebellar peduncle (β ranged from 0.017 to 0.029, p_{corr} ranged from 0.021 to 5.32×10⁻⁹) were associated with higher mean depressive symptoms. Other than the associations with mean depressive symptoms shown in ISOVF, other association was found in lower OD in parahippocampal part of cingulum (β = 0.016, pcorr = 0.050), medial lemniscus (β = 0.030, pcorr = 6.83×10⁻⁴), and superior thalamic radiation (β = -0.019, p_{corr} = 0.050).

The variability of depressive symptoms derived from multiple assessments

For general variance, positive associations were shown in all g measures including global variance, gAF, gTR and gPF in ISOVF (β ranged from 0.016 to 0.033, p_{corr} ranged from 0.042 to 1.83×10⁻⁴). Negative associations in OD were found in global variance and gAF (β ranged from 0.019 to 0.023, p_{corr} ranged from 0.046 to 0.040).

Tract-wise analysis showed that variability of depressive symptoms were positively associated with ISOVF in anterior thalamic radiations, superior thalamic radiation, uncinate fasciculus and middle cerebellar peduncle (β ranged from 0.015 to 0.037, p_{corr} ranged from 0.048 to 5.64×10⁻⁶). Other than measures of ISOVF, other association was found between variability of depressive symptoms and OD in parahippocampal part of cingulum and medial lemniscus (β ranged from 0.020 to 0.021,

p_{corr} ranged from 0.047 to 0.050).

For the above associations, we ran additional models to test if the results would change when smoking status, SLE and alcohol consumption were not corrected for. No association found in the main model where the three covariates were included turned insignificant after the covariates were removed (see Figure S5).

Effect of partial-volume contamination to FA and MD

We observed differences of results for FA and MD. A possible explanation may be the different effect of partial volume contamination on FA and MD. To control for possible effects of partialvolume contamination related to structural atrophy, we have included age and age² as covariates(5). We also conducted an additional analysis including brain size (in UK Biobank data dictionary: f.25010.2.0, described as "Volume of Brain, grey+white matter") as one of the covariates and the results remained significant for both FA and MD except for one association turned null for MD in anterior thalamic radiation after controlling for brain size, however reached nominal significance before multiple correction applied (β = 0.030, $p_{uncorr} = 0.020$, $p_{corr} = 0.053$) (Figure S6).

Effect of recent antidepressant intake on white matter microstructure

We have conducted a sensitivity analysis, removing participants who had medication only around the imaging assessment but not on other occasions. This was to test if removing the effect of recent intake of antidepressants would change the results.

First, we extracted a list of drug names from the British National Formulary-70 (BNF 70) [\(https://www.bnf.org/products/books/\)](https://www.bnf.org/products/books/) under the category for depression. Drug names that

matched the entries in treatment/medication (field name: f.20003, URL: [https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20003\)](https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20003) were identified as reported antidepressant intake. For the last instance of online follow-up questionnaires, the information of antidepressant intake was assessed in a different way by asking whether there is any substance taken for depression (field name: f.20546, URL: [http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20546\)](http://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=20546). Answers of "medication prescribed to you (for at least two weeks)" were identified as antidepressant intakes.

Participants who reported intake of antidepressant only at the imaging assessment (instance = 2), but not on other occasions (instance = 0/1/3) were removed from analysis, in order to remove the effect of antidepressant taken recently. Data from a total of 323 people were removed from analysis, and 19,345 remained.

The same regression models were conducted on this subsample, results are shown in Figure S7. Regional patterns for this subsample are very similar with the main findings (in main text).

Figure S1. Description of sample sizes and changes due to each step of data merging or outlier removal. Cross-sectional = Cross-sectional measure at the imaging assessment, Mean = mean level of depressive symptoms based on multiple assessments for at least two times, Variability = standard deviation of depressive level of multiple assessments for at least three times, and Longitudinal slope = slope of longitudinal changes over all four times of assessments.

Merged with distress measures

Merged with covariates

Figure S2. Illustration of WM tracts. The tracts were defined by tractography mapping on FA (fractional anisotropy) data using AutoPtx [\(https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/AutoPtx\)](https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/AutoPtx). They were categorised into three subsets as shown in the figure. Forceps major, forceps minor and uncinate fasciculus are unilateral structures and the rest are bilateral. For the purpose of clear illustration, bilateral structures were shown identically in both hemispheres.

Figure S3. Results for all dMRI measures including DTI and NODDI. More descriptive statistics were shown in Table S3 and S8, results in the main text and supplementary results. In the heatmaps, each colour theme represents one dMRI measure. For the measures of depressive symptoms: Cross-sectional = depressive level at the imaging assessment, Longitudinal slope = slope of longitudinal changes over three time points until the imaging assessment, Mean = mean level of depressive symptoms based on multiple assessments for at least two times, Variability = standard deviation of depressive level of multiple assessments for at least three times. As the measures of FA, ICVF and OD have negative direction with MD and ISOVF, here in this figure, the effect sizes for FA, ICVF and OD were reversed (×-1). Significant associations after FDR correction were marked with a single asterisk.

Figure S5. Results for a secondary model without controlling for stressful life events, smoking status or alcohol consumption. For the abbreviations and multiple correction methods, see the legend of Figure S3.

Figure S6. Results for adding brain size as a covariate.

Figure S7. Results for removing data from participants who had recent medication intake around the imaging assessment.

Table S1. Correlation matrix for all measures for depressive symptoms, smoking status, alcohol consumption, stressful life events, neuroticism and age range (time lag) for the multiple assessments used for generating mean level of depressive symptoms. For the measures of depressive symptoms: Cross-sectional measure = depressive level at the imaging assessment, Mean = mean level of depressive symptoms based on multiple assessments for at least two times, Variability = standard deviation of depressive level of multiple assessments for at least three times, and Longitudinal slope = slope of longitudinal changes over all four times of assessments. All r>0.3 are highlighted in bold. Measures for depressive symptoms were correlated with one another.

Table S2. Correlation loadings of each tract of PCA. For each dMRI measure, PCA on all tracts, association fibres, thalamic radiations and projection fibres were performed respectively. The loadings were reported as correlation loadings.

Table S3. Main results for DTI measures (FA and MD). Betas were standardised effect sizes. P values were un-corrected p values. All p_{corrected}<0.05 were marked by asterixis. FDR correction was applied on all association tests for a dMRI measure (n = 15*4 measures of depressive symptoms = 60 for tract analysis, and n = $4*4$ measures of depressive symptoms = 16 for g analysis). For the measures of depressive symptoms: Cross-sectional measure = depressive level at the imaging assessment, Mean = mean level of depressive symptoms based on multiple assessments for at least two times, Variability = standard deviation of depressive level of multiple assessments for at least three times, and Longitudinal slope = slope of longitudinal changes over all four times of assessments.

Table S4. Comparing models with and without the cross-sectional measure. H0 model: Imaging variables~covariates+Mean+Variability, and H1 model: Imaging variables~covariates+Mean+Variability+Cross-sectional measure. ANOVA was utilised to test if adding the cross-sectional measure as an independent variable gives significantly larger variance explained by the model compared to the H0 model. Dependent variables were MD in the tract categories/tracts that were found associated with the cross-sectional measure. Significant p values are highlighted in red.

Table S5. Comparing models with and without the longitudinal measures. H0 model: Imaging variables~covariates+Cross-sectional measure, and H1 model: Imaging variables~covariates+Mean+Variability+Cross-sectional measure. ANOVA was utilised to test if adding the cross-sectional measure as an independent variable gives significantly larger variance explained by the model compared to the H0 model. Dependent variables were MD in the tract categories/tracts that were found associated with the cross-sectional measure. Significant p values are highlighted in red.

Table S6. The effects of stressful life events (SLE), neuroticism, smoking status, and alcohol consumption as covariates for DTI measures. The test was conducted on the full sample of IDP from UK Biobank imaging team, without outlier exclusion or phenotypic data merging. Betas were standardised effect sizes. P values were un-corrected p values. All p_{corrected}<0.05 were marked by asterixis. FDR correction was applied on all tests conducted on one dMRI measure.

Table S7. The effects of stressful life events (SLE), neuroticism, smoking status, and alcohol consumption as covariates for NODDI measures. The test was conducted on the full sample of IDP from UK Biobank imaging team, without outlier exclusion or phenotypic data merging Betas were standardised effect sizes. P values were un-corrected p values. All p_{corrected}<0.05 were marked by asterixis. FDR correction was applied on all the tests for a dMRI measure.

Table S8. Main results for NODDI measures (ICVF, ISOVF, and OD). Betas were standardised effect sizes. For the abbreviations of measures for depressive symptoms, see the legend of Table S3.

Supplemental References

- 1. Andersson JLR, Jenkinson M, Smith SM (2007): Non-linear optimisation. FMRIB technical report TR07JA1. *In Pract*. .
- 2. Andersson JLR, Sotiropoulos SN (2015): An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage*. 125: 1063–1078.
- 3. Rosseel Y (2012): lavaan : an R package for structural equation modeling and more Version 0 . 5-12 (BETA). *J Stat Softw*. 48: 1–36.
- 4. Clarke T-K, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, *et al.* (2017): Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Mol Psychiatry*. 22: 1376–1384.
- 5. Smith SM, Nichols TE (2018): Statistical Challenges in "Big Data" Human Neuroimaging NeuroView. *Neuron*. 97: 263–268.