Association of trajectories of depressive symptoms with vascular risk, cognitive function and adverse brain outcomes: The Whitehall II MRI sub-study

Naiara Demnitz DPhil,[1,2] Melis Anatürk DPhil,[1] Charlotte L. Allan MD,[3] Nicola Filippini DPhil,[4] Ludovica Griffanti PhD,[4] Clare E. Mackay PhD,[1,4] Abda Mahmood PhD,[5] Claire E. Sexton DPhil,[1,4] Sana Suri DPhil,[1,4] Anya G. Topiwala DPhil,[1] Enikő Zsoldos DPhil,[1,4] Mika Kivimäki PhD,[6] Archana Singh-Manoux PhD,[6,7] Klaus P. Ebmeier MD[1]

[1] Department of Psychiatry, University of Oxford, Oxford, UK

[2] Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland

[3] Institute of Translational and Clinical Research, Newcastle University, and Tyne and Wear NHS Foundation Trust Cumbria, Northumberland UK

[4] Wellcome Centre for Integrative Neuroimaging (including Oxford Centre for Human Brain Activity and Functional Magnetic Resonance Imaging of the Brain), University of Oxford, UK

[5] Department of Population Health, London School of Hygiene and Tropical Medicine, London, UK

[6] Department of Epidemiology and Public Health, University College London, London, UK

[7] Université de Paris, INSERM U1153, Paris, France

Short Title: Trajectories of Depressive Symptoms

Corresponding Author:

Prof. Klaus P. Ebmeier
Department of Psychiatry, University of Oxford,
Warneford Hospital, Oxford OX3 7JX, UK
Tel.: 0044 1865 613100; email: klaus.ebmeier@psych.ox.ac.uk

Abstract

Background: Trajectories of depressive symptoms over the lifespan vary between people, but it is unclear whether these differences exhibit distinct characteristics in brain structure and function.
Methods: In order to compare indices of white matter microstructure and cognitive characteristics of groups with different trajectories of depressive symptoms, we examined 774 participants of the Whitehall II Imaging Sub-study, who had completed the depressive subscale of the General Health Questionnaire up to nine times over 25 years. Twenty-seven years after the first examination, participants underwent magnetic resonance imaging to characterize white matter hyperintensities (WMH) and microstructure and completed neuropsychological tests to assess cognition. Twenty-nine years after the first examination, participants completed a further cognitive screening test.
Outcomes: Using K-means cluster modelling, we identified five trajectory groups of depressive symptoms: consistently low scorers ("low"; n=505, 62·5%), a subgroup with an early peak in depression scores ("early"; n=123, 15·9%), intermediate scorers ("middle"; n=89, 11·5%), a late symptom subgroup with an increase in symptoms towards the end of the follow-up period ("late"; n=29, 3·7%), and consistently high scorers ("high"; n=28, 3·6%). The late, but not the consistently high scorers, showed higher mean diffusivity, larger volumes of WMH and impaired executive function. In addition, the late subgroup had higher Framingham Stroke Risk scores throughout the follow-up period, indicating a higher load of vascular risk factors.
Interpretation: Our findings suggest that tracking depressive symptoms in the community over time may be a useful tool to identify phenotypes that show different etiologies and cognitive and brain outcomes.

Keywords: Depressive symptoms; aging; white matter hyperintensities; cognition; late onset, vascular

Introduction

Depressive symptoms, tracked longitudinally in the general population, show heterogeneous patterns of change. Although most individuals experience few or no symptoms, some experience them transiently, while others consistently experience depressive symptoms (Musliner et al., 2016). These different trajectory patterns are hypothesized to have different implications for brain structure and function in late adulthood, but evidence for this hypothesis is scarce.

A number of studies have used trajectory analyses to identify groups based on longitudinal change patterns of depressive symptoms - for review, see Musliner et al. (2016). For example, Mirza et al. (2016) tracked depressive symptoms over the course of 11 years in a sample of 3,325 older adults and identified 5 trajectories: “low”, “increasing”, “decreasing”, “remitting” and “high”. The study showed that the trajectory characterized by increasing depressive symptoms was associated with a higher risk of developing dementia (Mirza et al., 2016). Similarly, in an analysis of depressive trajectories over 28 years, a late-onset, but not early-onset of depressive symptoms was predictive of an increased dementia risk (Singh-Manoux et al., 2017). Almeida et al. (2017) described a greater risk of dementia over 14 years follow-up in 4,922 cognitively healthy 71 to 89-year-olds with current depression, as opposed to those with a past history of depression or no history of depression. These findings have led to the suggestion that depressive symptoms may be a prodromal feature of dementia, or that the two share common causes, as opposed to depression being a causal risk factor for dementia (Herrmann et al., 2008; Mirza et al., 2016; Singh-Manoux et al., 2017). However, it remains unclear whether the groups that emerge directly from trajectory modelling based on depressive symptoms can also be predictive of cognitive impairment and structural brain differences.

The vascular depression hypothesis poses that comorbid late onset depression (LOD) and cerebrovascular disease disrupt fiber tracts within prefrontal systems (Alexopoulos et al., 1997). Meta-analyses of studies investigating cognitive and white matter changes in late life depression found that LOD was characterized by reduced measures of executive function and more frequent and intense white matter abnormalities (Herrmann et al., 2007; Herrmann et al., 2008). Longitudinal studies are required to test the relationship between depressive symptom trajectories and structural abnormalities directly, as well as their role in cognitive function and deterioration.

We hypothesize that temporal patterns of depressive symptoms have implications for cognitive function. While greater cognitive impairment has been reported in patients with LOD compared to those with early onset depression (EOD) (Eraydin et al., 2019; Herrmann et al., 2007), the reverse

pattern (Castaneda et al., 2008), as well as no group differences (Brodaty et al., 2001), have also been noted. In addition, the group-difference between LOD and EOD may be specific to cognitive domains. Interestingly, a review of studies comparing cognitive outcomes between EOD and LOD groups found that patients with LOD showed greater reductions in executive function than patients with EOD and controls, while memory impairments were equivalent in the two groups (Herrmann et al., 2007).

We hypothesize that participants showing a trajectory of late-onset depressive symptoms have a larger volume of white matter hyperintensities and indices of impairments in white matter microstructure, such as greater mean diffusivity and reduced fractional anisotropy. In line with our previous findings (Herrmann et al., 2007), we expect the late-onset group to have reduced executive function and be at greater risk of further cognitive deterioration.

Methods

Study sample

At baseline, the Whitehall II Study recruited, and continues to follow-up, a cohort of 10,308 participants who worked in the British civil service in 1985 (Marmot and Brunner, 2005). Eight hundred of these were randomly selected from Wave 11 (2012/13, see Figure 1) and recruited for the Whitehall II Imaging sub-study (Filippini et al., 2014). We included participants who completed at least four of nine instances of the General Health Questionnaire (GHQ) depressive subscale during the follow-up period (Figure 1), as well as a 3T MRI brain scan and a battery of cognitive tests on average 27·3 (SD = 1·5) years after the first GHQ assessment. We excluded participants with a history of neurological illness (participants with a history of transient ischemic attacks were included) or significant structural MRI abnormalities. Participants who had no available DTI or WMH data were excluded only from the analyses of that outcome. Thus, the total analytic sample from which the trajectories were derived included 774 participants, while analyses of WMH and DTI outcomes had 750 and 728 participants, respectively.

The Whitehall II Study and the Whitehall II Imaging sub-study received ethical approval from the University College London Medical School Committee on the Ethics of Human Research (Reference: 85/093) and the Oxford Central University Research Ethics Committee (Reference: MS IDREC-C1-2011-71), respectively. Informed written consent was obtained from all participants at all waves.

Depressive symptoms

Depressive symptoms were measured as total score, ranging from 0 to 12, on the depressive subscale of the General Health Questionnaire (GHQ). The GHQ has previously been validated for Whitehall II data and shown to have good criterion validity for minor psychiatric disorders (Head et al., 2013). Participants completed the GHQ at nine waves from 1985-2013 (Figure 1).

Neuropsychological Tests

At the time of the MRI scan, participants completed a battery of cognitive tests (Filippini et al., 2014). In line with our hypothesis, and to limit multiple comparisons, only tests of memory and executive function were included here in addition to a global cognitive screening test. For the measurement of executive function, we selected the Trail-Making Test (Completion Time Part B – Part A) (Reitan, 1955), Memory performance was measured using the delayed score and the total learning score over three trials of the Hopkins Verbal Learning Test Revised (HVLT-R) (Shapiro et al., 1999). The Mini Mental State Examination (MMSE) was used to assess global cognitive function in Waves 11 and 12 (Folstein et al., 1975).

Vascular risk

Vascular risk was assessed using the Framingham stroke risk scores (in percent risk over the following 10 years) collected during Waves 3, 5, 7, 9, and 11 and log10-transformed for analysis. Mean arterial BP at the time of scan was computed as the weighted mean (1:2) of systolic and diastolic BP.

MRI acquisition and processing

As previously described (Filippini et al., 2014), magnetic resonance imaging data were acquired on a 3T Siemens Magnetom Verio scanner in the first 550 participants (between April 2012 – December 2014), using a 32-channel receive head coil, at the Oxford Centre for Functional MRI of the Brain. T1-weighted structural images were acquired using a gradient echo sequence (TR = 2530ms, TE= 1·79/3·65/5·51/7·37ms, flip angle = 7°, FOV = 256mm, voxel dimension = 1·0 mm isotropic, acquisition time = 6m12s). Diffusion-weighted images were acquired using an echoplanar sequence, with 60 diffusion-weighted directions (*b*-value = 1500 s/mm2), 5 non-diffusion weighted images (*b*-value = 0s/mm2) and one b0 volume in the reversed phase-encoded direction (TR = 8900ms, TE = 91·2ms, FOV= 192 mm, voxel dimension = 2·0mm isotropic). FLAIR images were also acquired (TR = 9000ms, TE = 73, voxel dimension=0·9x0·9x3mm^3, FOV = 220, acquisition time = 4m14s). The last 250

participants were scanned on a 3T Siemens Magnetom Prisma scanner with 64-channel receive head-neck coil in the same center (between July 2015 – December 2016). T1-weighted structural images were acquired using a gradient echo sequence (TR = 1900ms, TE= 3·97ms, flip angle = 8°, FOV = 192mm, voxel dimension = 1·0 mm isotropic, acquisition time = 5m31s). FLAIR (TR = 9000ms, TE = 73, voxel dimension=0·4x0·4x3mm^3, FOV = 220, acquisition time = 4m14s) and diffusion-weighted images, with a matched protocol except for a change in echo time (TE = 91ms), were also acquired. Analysis of MRI data was conducted using tools from the FMRIB Software Library (FSL). Partial-volume tissue segmentation was performed using the FMRIB Automated Segmentation Tool (FAST) (Zhang et al., 2001). Diffusion-weighted images were first corrected for head movement and eddy currents. Global measures of fractional anisotropy (FA) and mean (MD) diffusivity were then extracted using DTIFit (http://fsl.fmrib.ox.ac.uk/fsl/fdt). First, FA maps were created by fitting a tensor model to the raw diffusion data. This fits a diffusion tensor model to the raw diffusion data and then applies BET to remove non-brain voxels (Smith, 2002). All participants' FA data were aligned into standard space using FMRIB's Nonlinear Registration Tool (FNIRT) (Andersson et al., 2007). The mean FA image was then calculated and thinned to create a mean FA skeleton representing the centers of all tracts common to the group. This method was repeated for mean diffusivity (MD). Measures of FA and MD were then extracted from the TBSS skeleton.

To acquire an estimate of white matter hyperintensity volume, FLAIR images were automatically segmented using the Brain Intensity AbNormality Classification Algorithm (BIANCA), a fully-automated, supervised method for WMH detection, based on the k-nearest neighbor algorithm (Griffanti et al., 2016). Briefly, BIANCA classifies the image voxels based on their intensity and spatial features. The output image represents the probability per voxel of being WMH. In order to adjust for brain volume, total WMH volume was then divided by total intracranial volume (ICV) and multiplied by 100 (WMH%). The application of this processing pipeline to this dataset has been described elsewhere (Griffanti et al., 2018). All analyses of imaging data included a binary covariate for scanner.

Sample characteristics and covariates

Age, sex and years of full-time education were recorded for all participants at the time of the MRI. Socio-economic stratum was determined according to occupation in 1985–88 (Wave 1). At the time of the scan (2012-16), premorbid full-scale intelligence quotient (FSIQ) was estimated from the Test of Premorbid Function (TOPF), and the Pittsburgh Sleep Quality Inventory (PSQI) was used to measure sleep quality (Buysse et al., 1989). To describe the diagnostic relevance of the depressive

symptom trajectories, the following measures were used: ICD10 History of depressive episode determined from CIS-R in 2012-13 (Head et al., 2013), DSM IV-TR criteria for lifetime history of minor and major depressive episodes diagnosed by SCID in 2012-16 (First et al., 2002), self-reported clinical diagnosis of depression and current treatment with antidepressant drugs at the same time, as well as the Centre for Epidemiological Studies Depression Scale (CES-D) (Radloff, 1977).

Statistical analysis

K-means cluster modelling (kml) was used to identify distinct cluster trajectories of GHQ depression scores. Kml is a non-parametric hill-climbing algorithm which does not impose assumptions regarding the shape of the trajectories (Genolini and Falissard, 2011). The kml-analysis was specified to allow between 2 and 5 clusters (trajectories), each obtained by running 1000 permutations. The Calinski and Harabasz criteria, along with consideration of clinical relevance, were used to select the optimum number of clusters. Kml offers 3 versions of the Calinski and Harabasz criteria, described in detail elsewhere (Genolini, 2015), and all were considered. While the traditional and Kryszczuk variants of the Calinski and Harabasz criterion indicated that the optimum number of clusters was 2, the Genolini variant suggested a 5-group result. The 2-cluster response grouped participants into those who consistently reported minimal depressive symptoms (81%), and those who showed some symptoms (19%). Although this clustering resulted in a higher Calinski and Harabasz criterion in 2 variants, we opted for the 5-cluster response, because (a) the 2-groups solution is of limited clinical relevance as it combines all trajectories presenting some depressive symptoms, thus losing important information about their trajectories of change, and (b) our research question addresses a group showing a later emergence of depressive symptoms. K-means cluster modelling for longitudinal data and statistical analyses were performed in R version 3.5.1 with RStudio version 1.1.463 (RStudio Team, 2020) using the psych (Revelle, 2018), kml (Genolini and Falissard, 2011), and ggplot2 packages (Wickham and Sievert, 2016).

Data were inspected by plotting variables by trajectory group, in particular for variables that replicate and validate the depression measures, i.e. CES-D, as well as symptoms scales related to clinical depression, such as the Pittsburgh Sleep Quality Inventory, and physiological measures, such as mean arterial blood pressure. Finally, following the vascular hypothesis of late onset depression, we plotted group means of Framingham Stroke risk scores computed in Waves 3, 5, 7, 9, and 11. All measures were adjusted for age and sex differences between trajectory groups. Group differences in cognitive outcomes were also adjusted for premorbid IQ.

Cognitive and brain outcomes

Hypothesis testing was carried out for brain measures at the MRI phase, 27·3 (SD = 1·5) years after the first GHQ assessment. ANCOVAs were used to evaluate whether the identified trajectory groups significantly differed on cognitive and brain structure outcomes. In addition, we tested whether group membership was associated with follow-up cognitive variables that we predicted to be abnormal in a late-onset trajectory: MMSE, HVLT-R, and the difference measure between Trails A and B (TMT; representing executive function), measured 28·8 years (SD: 0·69) after the first examination. Age, sex, education and premorbid IQ were included as covariates in the analysis of cognitive data. Age, sex, education, premorbid IQ, and MRI scanner were included as covariates in the analysis of MRI data. Planned post-hoc tests between the late onset and all other trajectory groups were carried out following our primary hypothesis.

Outcomes

A total of 774 participants were included in the analyses. They were mostly male (81%), had a mean age of 69·8 ± 5·2 years and had, on average, 14·8 ± 3·3 years of full-time education.

The K-means cluster analysis identified five distinct trajectories of depressive symptoms in these 774 participants (Figure 2a). Most participants maintained a low score of depressive symptoms throughout the study (“low” depressive symptoms; n = 505, 65·2%). The remaining four trajectories were characterized by early depressive symptoms that then decreased (“early” depressive symptoms; n = 123, 15·9%); moderate levels of depressive symptoms gradually decreasing after Wave 7 (“middle” depressive symptoms; n = 89, 11·5%); low starting scores that steadily increased throughout follow-up (“late” depressive symptoms; n = 29, 3·7%); or maintained high scores throughout (consistent “high” depressive symptoms n = 28, 3·6%).

Diagnostic validation

Figure 2b shows the percentage of participants qualifying for GHQ-caseness (GHQ depression sub-score>3) in each group by waves. Table 1 describes the trajectory groups in terms of ICD10 History of a depressive episode determined from CIS-R in 2012-13 (Head et al., 2013), as DSM IV-TR criteria for lifetime history of minor and major depressive episodes diagnosed by SCID in 2012-16 (First et al., 2002), and self-reported clinical diagnosis of depression and current treatment with antidepressant drugs at the same time.

Sample characteristics of trajectory groups

Participants in the five trajectory-based groups did not significantly differ in terms of age at scan, ethnic group (white/other), socio-economic stratum, years of education, or pre-morbid IQ. However, as expected, the "late depressive symptoms" and "consistently high depressive symptoms" groups had a significantly larger proportion of women to men and, reflecting the presence of depressive symptoms at the end of the follow-up period, higher CES-D scores at time of scan compared with individuals in the other trajectory groups (Table 2). Together with the "early" and "middle" symptom groups they reported poorer sleep quality measured by the Pittsburgh Sleep Quality Inventory than the "low" symptom group. As predicted, and consistent with the vascular depression hypothesis (Alexopoulos et al., 1997), the "late" depressive symptoms group had higher vascular risk (Framingham Stroke Risk Score) on average during the preceding 28 years, and individually (Bonferroni-controlled) as far back as 22 years during Wave 3, as well as a higher mean arterial BP at the time of the scan (Table 3).

Cognitive and brain outcomes

ANCOVAs indicated that the "late" depressive symptom group had a poorer performance on the executive component of the Trail Making Test, and a significant deterioration of MMSE score between Waves 11 and 12 (Table 2, Figure 4). Performance on the HVLT-R (total and delayed recall) and MMSE at the time of scan did not differ between groups. The "late" onset trajectory group, but not the consistently "high" scorers, had higher global white matter diffusivity values than "low" symptom scorers; this effect was not significant for FA. The volume of white matter hyperintensities in the "late" onset trajectory group was significantly higher than in the consistently "high" scorers.

Interpretation

In this large group of community dwelling older adults, we identified five trajectories of depressive symptoms over a period of 28 years: those with low, early, middle, late and consistently high depressive symptoms. We hypothesized that a pattern depicting a late increase in depressive symptoms would be associated with decreased cognitive function and measures of adverse brain outcomes. In line with our hypothesis, participants in the late depressive symptoms group had a significantly poorer score on the TMT B-A, indicative of difficulties with executive function, increased white matter hyperintensities and increased vascular risk factors going back up to 22 years. We also

found that participants in this trajectory group had reduced average MMSE scores over the 3·2 years between Waves 11 and 12, suggesting that a pattern of late-onset depressive symptoms might be associated with decreasing cognitive function.

The pattern of trajectories observed in our study, as well as the distribution of participants amongst them, is comparable with other studies of community-dwelling older adults (Mirza et al., 2016). Similarly, our findings add to existing reports of cognitive decline, especially in executive function, as a characteristic of late-onset depression (Herrmann et al., 2007; Taylor et al., 2013). While some studies suggest that older adults may develop cognitive impairment as a result of experiencing depressive symptoms and the pathological correlates of depressive episodes or stress (Marazziti et al., 2010; Sapolsky et al., 1985), others argue that depressive symptoms are a prodromal feature of cognitive decline, reflecting a shared underlying pathology (Herrmann et al., 2007; Herrmann et al., 2008; Singh-Manoux et al., 2017), or that depression may be a risk factor for cognitive impairment (Ownby et al., 2006). The first hypothesis, not supported by our data, would predict the presence of poor cognitive performance in all four groups showing depressive symptoms. For the risk-factor hypothesis, also not supported by our data, one would expect the group with consistent high depressive symptoms to show the largest ('dose dependent') decline in cognitive function. Our findings, in line with others (Mirza et al., 2016), support the hypothesis that late-onset depressive symptoms may be an early manifestation of cognitive decline. Furthermore, these findings highlight the value of screening patients presenting with depressive symptoms in late life for cognitive deficits.

Given the evidence pointing towards vascular disease as the link between depression and dementia (for review, see (Byers and Yaffe, 2011) and consistent with the greater vascular risk of the late onset group (Table 3, Figure 3), we had hypothesized that the propensity for cognitive decline observed in the late-onset trajectory group would be associated with shared brain pathologies and increased vascular risk factors. As predicted, the groups differed in white matter hyperintensity volume, a marker of microvascular changes (Debette and Markus, 2010), as well as mean diffusivity in white matter, suggesting impaired white matter structural integrity in the late onset group. Given that vascular factors are hypothesized to cause white matter damage (Zsoldos et al., 2018), our findings are in line with the vascular depression hypothesis linking vascular risk factors to late-life depression (Alexopoulos et al., 1997). Levels of the inflammatory markers C-reactive protein and Interleukin-6 were not specifically raised in any of the trajectory subgroups during earlier Waves 3,5 or 7 (see Supplementary Information Appendix 3) (Kivimaki et al., 2014; Virtanen et al., 2015). Contrary to our prediction, we also did not find significantly reduced fractional anisotropy in the

group with late onset of depressive symptoms. Although we had a large sample of community-dwelling adults, the majority did not present with depressive symptoms, and only a small number of participants showed a consistently high or late-onset trajectory of depressive symptoms. Accordingly, the analysis may have been underpowered to estimate a relationship with these outcomes.

Kml is a non-parametric clustering method, which has proven useful in discerning clinical and behavioral trajectories in older adults (Chamberlain et al., 2016; Demmelmaier et al., 2016). Given its non-parametric nature, kml does not test the fit between the partition found and a theoretical model. This is both a strength and a weakness. On one hand, this model-free approach allows us to derive more data-driven trajectories without the restriction of set parameters. On the other hand, without a model to be tested against, it is not possible to test the fit nor to calculate the likelihood of class assignments, as is done in model-based methods (e.g. latent class growth modelling). In order to ascertain that the trajectory-groups observed were not specific to kml, we assessed whether the late-onset group identified by kml could be replicated using latent growth curve modelling, and a similar group allocation was observed (Appendix 2). As with any trajectory analysis, it's also important to recognize that assumptions are being made regarding the interval between measures of depressive symptoms, and we cannot discard the possibility that additional changes occurred between time-points. Other limitations in our study include the use of a shortened depressive symptom measure, the GHQ-depression sub-score, which may question the validity of its assessment of depressive symptoms. Nonetheless, previous studies have found that this measure is well suited for detecting depression in the general population (Koeter, 1992; Lundin et al., 2016). In addition, the percentage of cases by Wave, as defined by the standard caseness criterion of the GHQ scale (Figure 2), was in line with the patterns observed in our trajectories. The proportions, if not the absolute values, for diagnostic frequencies (Table 1) determined by questionnaire and interview support this conclusion. Finally, since the Whitehall II cohort consists of British civil servants recruited in the 1980s, a workforce that was then predominantly male, this also true for our sample. Given that there is a higher prevalence of depressive symptoms in females than males (Kuehner, 2017), our findings although adjusted for sex may not generalize to both sexes. It would be of interest for further studies to examine the interactions between sex, depressive symptoms and brain structures in the context of risk for cognitive decline.

Strengths of our study include a long follow-up period of almost 3 decades and a well-characterized sample, enabling us to comprehensively examine a list of variables as potential predictors and confounders of depressive symptom trajectories. Also of note, the use of 9 repeated measures of

depressive symptoms, a higher frequency than most comparable studies (Musliner et al., 2016), enabled us to estimate more complex trajectory patterns. It is worth noting that at the time of scan, participants in the late-onset and consistently high symptom groups did not differ in terms of severity of depressive symptoms, as evidenced by their CES-D and GHQ-depression scores. This highlights how valuable information extracted from trajectory analyses can be, as these two groups, with different prognoses for later cognitive outcome, could not have been distinguished cross-sectionally without knowledge of their history.

In conclusion, our study of trajectories of depressive symptoms over almost 3 decades indicates that a late emerging pattern of depressive symptoms is associated with increased vascular risk over 20 years, with impaired brain white matter structure and current executive dysfunction, but not poor learning or memory. It also predicted further deterioration in MMSE scores over 3·2 years. Altogether, our findings suggest that tracking depressive symptoms over time may be a useful tool to identify "pre-clinical" phenotypes of cognitive and brain disorders in the community with different etiologies and prognoses.

Acknowledgements & Funding

We are grateful to all Whitehall II study staff and participants, as well as the radiographers and physicists of the Wellcome Centre for Integrative Neuroimaging in Oxford.

The Whitehall Imaging Sub-study was funded by UK Medical Research Council (G1001354), and HDH Wills 1965 Charitable Trust (1117747). EZs & SS were funded by EU Horizon 2020 Grant 'Lifebrain' (732592). CEM & CES were supported by NIHR Oxford Health Biomedical Research Centre. Wellcome Centre for Integrative Neuroimaging by Wellcome Trust (203139/Z/16/Z). MK & ASM by the MRC (MR/S011676/1, MR/R024227), NIH (R01AG056477, RF1AG062553), and MK also by the Academy of Finland. CES is now a full-time employee of the US Alzheimer's Association. None of the above organizations had a role in any aspect pertinent to the study, including the writing of the manuscript or the decision to submit it for publication.

References

Alexopoulos, G.S., Meyers, B.S., Young, R.C., Campbell, S., Silbersweig, D., Charlson, M., 1997. 'Vascular depression' hypothesis. Arch Gen Psychiatry 54(10), 915-922. https://doi.org/10.1001/archpsyc.1997.01830220033006.

Almeida, O.P., Hankey, G.J., Yeap, B.B., Golledge, J., Flicker, L., 2017. Depression as a modifiable factor to decrease the risk of dementia. Transl Psychiatry 7(5), e1117. https://doi.org/10.1038/tp.2017.90.

Andersson, J., Jenkinson, M., Smith, S., 2007. Non-linear registration aka spatial normalisation. FMRIB Technical Report TR07JA2. https://doi.org/https://www.fmrib.ox.ac.uk/datasets/techrep/tr07ja2/tr07ja2.pdf.

Brodaty, H., Luscombe, G., Peisah, C., Anstey, K., Andrews, G., 2001. A 25-year longitudinal, comparison study of the outcome of depression. Psychol Med 31(8), 1347-1359. https://doi.org/10.1017/s0033291701004743.

Buysse, D.J., Reynolds, C.F., 3rd, Monk, T.H., Berman, S.R., Kupfer, D.J., 1989. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 28(2), 193-213. https://doi.org/10.1016/0165-1781(89)90047-4.

Byers, A.L., Yaffe, K., 2011. Depression and risk of developing dementia. Nat Rev Neurol 7(6), 323-331. https://doi.org/10.1038/nrneurol.2011.60.

Castaneda, A.E., Suvisaari, J., Marttunen, M., Perala, J., Saarni, S.I., Aalto-Setala, T., Aro, H., Koskinen, S., Lonnqvist, J., Tuulio-Henriksson, A., 2008. Cognitive functioning in a population-based sample of young adults with a history of non-psychotic unipolar depressive disorders without psychiatric comorbidity. J Affect Disord 110(1-2), 36-45. https://doi.org/10.1016/j.jad.2007.12.239.

Chamberlain, A.M., St Sauver, J.L., Jacobson, D.J., Manemann, S.M., Fan, C., Roger, V.L., Yawn, B.P., Finney Rutten, L.J., 2016. Social and behavioural factors associated with frailty trajectories in a population-based cohort of older adults. BMJ Open 6(5), e011410. https://doi.org/10.1136/bmjopen-2016-011410.

Debette, S., Markus, H.S., 2010. The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. BMJ 341, c3666. https://doi.org/10.1136/bmj.c3666.

Demmelmaier, I., Dufour, A.B., Nordgren, B., Opava, C.H., 2016. Trajectories of Physical Activity Over Two Years in Persons With Rheumatoid Arthritis. Arthritis Care Res (Hoboken) 68(8), 1069-1077. https://doi.org/10.1002/acr.22799.

Eraydin, I.E., Mueller, C., Corbett, A., Ballard, C., Brooker, H., Wesnes, K., Aarsland, D., Huntley, J., 2019. Investigating the relationship between age of onset of depressive disorder and cognitive function. Int J Geriatr Psychiatry 34(1), 38-46. https://doi.org/10.1002/gps.4979.

Filippini, N., Zsoldos, E., Haapakoski, R., Sexton, C.E., Mahmood, A., Allan, C.L., Topiwala, A., Valkanova, V., Brunner, E.J., Shipley, M.J., Auerbach, E., Moeller, S., Ugurbil, K., Xu, J., Yacoub, E., Andersson, J., Bijsterbosch, J., Clare, S., Griffanti, L., Hess, A.T., Jenkinson, M., Miller, K.L., Salimi-Khorshidi, G., Sotiropoulos, S.N., Voets, N.L., Smith, S.M., Geddes, J.R., Singh-Manoux, A., Mackay, C.E., Kivimaki, M., Ebmeier, K.P., 2014. Study protocol: The Whitehall II imaging sub-study. Bmc Psychiatry 14, 159. https://doi.org/10.1186/1471-244X-14-159.

First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., 2002. Structured Clinical Interview for DSM-IV-TR Axis I Disorders (Research Version). Biometrics Research Department New York State Psychiatric Institute, New York.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12(3), 189-198. https://doi.org/10.1016/0022-3956(75)90026-6.

Genolini, C., Alacoque, X., Sentenac, M., Arnaud C., 2015. kml and kml3d: R Packages to Cluster Longitudinal Data. Journal of Statistical Software 65(4), 1-34. https://doi.org/10.18637/jss.v065.i04.

Genolini, C., Falissard, B., 2011. KmL: a package to cluster longitudinal data. Comput Methods Programs Biomed 104(3), e112-121. https://doi.org/10.1016/j.cmpb.2011.05.008.

Griffanti, L., Jenkinson, M., Suri, S., Zsoldos, E., Mahmood, A., Filippini, N., Sexton, C.E., Topiwala, A., Allan, C., Kivimaki, M., Singh-Manoux, A., Ebmeier, K.P., Mackay, C.E., Zamboni, G., 2018. Classification and characterization of periventricular and deep white matter hyperintensities on MRI: A study in older adults. Neuroimage 170, 174-181. https://doi.org/10.1016/j.neuroimage.2017.03.024.

Griffanti, L., Zamboni, G., Khan, A., Li, L., Bonifacio, G., Sundaresan, V., Schulz, U.G., Kuker, W., Battaglini, M., Rothwell, P.M., Jenkinson, M., 2016. BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. Neuroimage 141, 191-205. https://doi.org/10.1016/j.neuroimage.2016.07.018.

Head, J., Stansfeld, S.A., Ebmeier, K.P., Geddes, J.R., Allan, C.L., Lewis, G., Kivimaki, M., 2013. Use of self-administered instruments to assess psychiatric disorders in older people: validity of the General Health Questionnaire, the Center for Epidemiologic Studies Depression Scale and the self-completion version of the revised Clinical Interview Schedule. Psychol Med 43(12), 2649-2656. https://doi.org/10.1017/S0033291713000342.

Herrmann, L.L., Goodwin, G.M., Ebmeier, K.P., 2007. The cognitive neuropsychology of depression in the elderly. Psychol Med 37(12), 1693-1702. https://doi.org/10.1017/S0033291707001134.

Herrmann, L.L., Le Masurier, M., Ebmeier, K.P., 2008. White matter hyperintensities in late life depression: a systematic review. J Neurol Neurosurg Psychiatry 79(6), 619-624. https://doi.org/10.1136/jnnp.2007.124651.

Kivimaki, M., Shipley, M.J., Batty, G.D., Hamer, M., Akbaraly, T.N., Kumari, M., Jokela, M., Virtanen, M., Lowe, G.D., Ebmeier, K.P., Brunner, E.J., Singh-Manoux, A., 2014. Long-term inflammation increases risk of common mental disorder: a cohort study. Mol Psychiatry 19(2), 149-150. https://doi.org/10.1038/mp.2013.35.

Koeter, M.W., 1992. Validity of the GHQ and SCL anxiety and depression scales: a comparative study. J Affect Disord 24(4), 271-279. https://doi.org/10.1016/0165-0327(92)90112-j.

Kuehner, C., 2017. Why is depression more common among women than among men? Lancet Psychiatry 4(2), 146-158. https://doi.org/10.1016/S2215-0366(16)30263-2.

Lundin, A., Hallgren, M., Theobald, H., Hellgren, C., Torgen, M., 2016. Validity of the 12-item version of the General Health Questionnaire in detecting depression in the general population. Public Health 136, 66-74. https://doi.org/10.1016/j.puhe.2016.03.005.

Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., Faravelli, L., 2010. Cognitive impairment in major depression. Eur J Pharmacol 626(1), 83-86. https://doi.org/10.1016/j.ejphar.2009.08.046.

Marmot, M., Brunner, E., 2005. Cohort Profile: the Whitehall II study. Int J Epidemiol 34(2), 251-256. https://doi.org/10.1093/ije/dyh372.

Mirza, S.S., Wolters, F.J., Swanson, S.A., Koudstaal, P.J., Hofman, A., Tiemeier, H., Ikram, M.A., 2016. 10-year trajectories of depressive symptoms and risk of dementia: a population-based study. Lancet Psychiatry 3(7), 628-635. https://doi.org/10.1016/S2215-0366(16)00097-3.

Musliner, K.L., Munk-Olsen, T., Eaton, W.W., Zandi, P.P., 2016. Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes. J Affect Disord 192, 199-211. https://doi.org/10.1016/j.jad.2015.12.030.

Ownby, R.L., Crocco, E., Acevedo, A., John, V., Loewenstein, D., 2006. Depression and risk for Alzheimer disease: systematic review, meta-analysis, and metaregression analysis. Arch Gen Psychiatry 63(5), 530-538. https://doi.org/10.1001/archpsyc.63.5.530.

Radloff, L.S., 1977. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement 1, 385–401.

Reitan, R.M., 1955. The relation of the trail making test to organic brain damage. J Consult Psychol 19(5), 393-394. https://doi.org/10.1037/h0044509.

Revelle, W., 2018. psych: Procedures for Personality and Psychological Research, 1.8.12 ed. The Comprehensive R Archive Network, Illinois, USA.

RStudio Team, 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.

Sapolsky, R.M., Krey, L.C., McEwen, B.S., 1985. Prolonged glucocorticoid exposure reduces hippocampal neuron number: implications for aging. J Neurosci 5(5), 1222-1227.

Shapiro, A.M., Benedict, R.H., Schretlen, D., Brandt, J., 1999. Construct and concurrent validity of the Hopkins Verbal Learning Test-revised. Clin Neuropsychol 13(3), 348-358. https://doi.org/10.1076/clin.13.3.348.1749.

Singh-Manoux, A., Dugravot, A., Fournier, A., Abell, J., Ebmeier, K., Kivimaki, M., Sabia, S., 2017. Trajectories of Depressive Symptoms Before Diagnosis of Dementia: A 28-Year Follow-up Study. JAMA Psychiatry 74(7), 712-718. https://doi.org/10.1001/jamapsychiatry.2017.0660.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum Brain Mapp 17(3), 143-155. https://doi.org/10.1002/hbm.10062.

Taylor, W.D., Aizenstein, H.J., Alexopoulos, G.S., 2013. The vascular depression hypothesis: mechanisms linking vascular disease with depression. Mol Psychiatry 18(9), 963-974. https://doi.org/10.1038/mp.2013.20.

Virtanen, M., Shipley, M.J., Batty, G.D., Hamer, M., Allan, C.L., Lowe, G.D., Ebmeier, K.P., Akbaraly, T.N., Alenius, H., Haapakoski, R., Singh-Manoux, A., Kivimaki, M., 2015. Interleukin-6 as a predictor of symptom resolution in psychological distress: a cohort study. Psychol Med 45(10), 2137-2144. https://doi.org/10.1017/S0033291715000070.

Wickham, H., Sievert, C., 2016. ggplot2 : elegant graphics for data analysis, Second edition. ed. Springer, Switzerland.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20(1), 45-57. https://doi.org/10.1109/42.906424.

Zsoldos, E., Filippini, N., Mahmood, A., Mackay, C.E., Singh-Manoux, A., Kivimaki, M., Jenkinson, M., Ebmeier, K.P., 2018. Allostatic load as a predictor of grey matter volume and white matter integrity in old age: The Whitehall II MRI study. Sci Rep 8(1), 6411. https://doi.org/10.1038/s41598-018-24398-9.

Table 1. Diagnostic Validation

Five GHQ Depression Trajectories	*ICD10 History of a Depressive episode (CIS-R in 2012-13)*	*DSM4 History of Minor or Major Depression (SCID in 2012-16)*	*ICD Depression (F30-9) Diagnosis (self-report 2012-16)*	*Taking Antidepressants (self-report 2012-16)*	*Total Trajectory Group N*
	N (% of Group N)				
Low	205 (40·7%)	103 (20·4%)	26 (5·1%)	11 (2·2%)	505
Early	76 (61·8%)	46 (37·4%)	13 (10·6%)	6 (4·9%)	123
Middle	60 (67·4%)	36 (40·4%)	16 (18·0%)	9 (10·1%)	89
Late	18 (64·3%)	15 (51·7%)	*5 (17·2%)*	4 (13·8%)	29
High	22 (78·6%)	19 (67·9%)	12 (42·9%)	5 (17·9%)	28

Table 2. Characteristics of estimated GHQ trajectory groups

		GHQ-Dep Trajectory Groups									
		Low		**Early**		**Middle**		**Late**		**High**	
a) Description of Trajectory Groups											
Age at scan [mean SE]		70·0$_a$	0·23	69·4$_a$	0·47	69·8$_a$	0·55	70·0$_a$	0·97	69·0$_a$	0·98
Sex*	Male [N / Row-%]	423	68%	92	15%	74	12%	19	3%	**17**	3%
	Female [N / Row-%]	82$_a$	55%	31$_{a, b}$	21%	15$_{a, b}$	10%	10$_{a, b}$	7%	**11$_b$**	7%
Socio-Economic Stratum	Administrative (1) [N / Row-%]	211	66%	50	16%	43	13%	9	3%	7	2%
	Professional / Executive (2,3) "	256	64%	65	16%	43	20%	18	5%	16	4%
	Clerical / Support (4,5) "	38$_a$	68%	8$_a$	14%	3$_a$	5%	2$_a$	4%	5$_a$	9%
Full-time education (years) [mean SE][1]		14·6a	0·15	15·1a	0·30	15·1a	0·35	14·0a	0·62	15·0a	0·63
FSIQ (estimated from TOPF) [mean SE][1]		115$_a$	0·38	116$_a$	0·76	115$_a$	0·90	114$_a$	1·57	117$_a$	1·61
MMSE during Wave 11 (out of 30**) [mean SE][2]**		28·5$_a$	0·05	28·5$_a$	0·11	28·4$_a$	0·13	28·3$_a$	0·22	28·3$_a$	0·23
PSQI [mean SE] ***[1]		4·4$_a$	0·13	**5·6$_b$**	0·26	**5·8$_b$**	0·31	**6·1$_b$**	0·54	**6·4$_b$**	0·55
CES-D [mean SE] ***[1]		3·6$_a$	0·23	**5·4$_b$**	0·48	**8·1$_c$**	0·56	**15·1$_d$**	0·98	**15·1$_d$**	1·00
b) Hypothesis testing (cognitive performance)											
Hopkins Verbal Learning Test Total recall [mean SE][2]		27·5$_a$	0·19	27·3$_a$	0·39	27·7$_a$	0·46	26·2$_a$	0·80	26·6$_a$	0·82
TMTB - TMTA [sec] [mean SE] ***[2]		36·6$_a$	1·24	35·5$_a$	2·52	34·6$_a$	2·95	**61·0$_b$**	5·26	27·0$_a$	5·29

MMSE at Wave 12 (out of 30) [mean SE] **[3]	28·5$_{a}$	0·06	28·7$_{a}$	0·13	28·5$_{a,c}$	0·15	**27·6$_{b, c}$**	0·26	28·3$_{a, c}$	0·28
c) Hypothesis testing (brain measures)										
Global Mean Diffusivity (x10^{4}) [mean SE] *[4]	6·85$_{a}$	0·01	6·88$_{a, b}$	0·02	6·86$_{a, b}$	0·03	**7·00$_{b}$**	0·05	6·85$_{a, b}$	0·05
Global Fractional Anisotropy (x10^{-2}) [mean SE][4]	48·3$_{a}$	0·01	48·1$_{a}$	0·02	48·2$_{a}$	0·02	47·5$_{a}$	0·03	48·6$_{a}$	0·03
White Matter Hyperintensity Volume [% of ICV] [mean SE] *[4]	0·45$_{a, c}$	0·01	0·47$_{a, c}$	0·02	0·49$_{a, c}$	0·03	***0·58$_{a, b}$***	0·05	0·37$_{c}$	0·05

*omnibus p<0·05**p<0·01***p<0·001; [1]adjusted for age and sex; [2]adjusted for age, sex and premorbid IQ; [3]adjusted for age, sex, premorbid IQ, and MMSE at Wave 11; [4]adjusted for age, sex, premorbid IQ and scanner type; MMSE scores were transformed as Log(31-MMSE) for the analyses.

Each subscript letter denotes a subset of Five GHQ Depression-trajectories categories whose column proportions/values do not differ significantly from each other at the 0·05 level after Bonferroni correction;

Trajectory groups different from 'low group' in bold;

Trajectory groups different from 'high group' in *bold italics.*

CES-D = Center for Epidemiological Studies Depression Scale; FSIQ = Full Scale IQ; GHQ = General Health Questionnaire; ICV = Intracranial Volume; MMSE = Mini Mental State Examination; PSQI = Pittsburgh Sleep Quality Inventory; TMT = Trail Making Test; TOPF = Test of Premorbid Function.

Table 3. Vascular risks in estimated GHQ trajectory groups

	GHQ-Dep Trajectory Groups									
	Low		**Early**		**Middle**		**Late**		**High**	
Vascular Risk of Trajectory Groups										
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Mean Arterial BP at scan [mean SE] *[1]	98·5$_{a, b}$	0·53	99·9$_{a, b}$	1·08	98·1$_{a, b}$	1·26	***103·2$_{a, c}$***	2·22	94·4$_{b}$	2·26
Framingham Stroke Score Wave 3 º†1*	0·474$_{a}$	0·006	·466$_{a}$	0·012	0·470$_{a}$	0·014	**0·538$_{b,c}$**	0·027	0·462$_{a,c}$	0·026
Framingham Stroke Score Wave 5 º†1	0·521	0·008	0·518	0·016	0·545	0·019	0·580	0·035	0·526	0·034
Framingham Stroke Score Wave 7 º†1	0·633	0·010	0·644	0·019	0·655	0·023	0·728	0·042	0·556	0·041
Framingham Stroke Score Wave 9 º†1***	0·723$_{a}$	0·009	0·748$_{a,c}$	0·018	0·771$_{a,d}$	0·022	***0·843$_{b,c,d}$***	0·040	0·640$_{a}$	0·039
Framingham Stroke Score Wave 11 º†1*	0·840$_{a}$	0·010	0·854$_{a}$	0·020	0·884$_{a}$	0·024	**0·926$_{b}$**	0·044	0·758$_{a,b}$	0·043
Mean Framingham Stroke Scores *†1	0·638$_{a, b}$	0·007	0·646$_{a, b}$	0·014	0·665$_{a, b}$	0·016	***0·723$_{a, c}$***	0·031	0·589$_{b}$	0·030

* group effect p<0·05; ***p<0·001; º group x wave linear interaction effect p<0·05; †Percentage scores log10-transformed; [1]adjusted for age and sex

Each subscript letter denotes a subset of Five GHQ Depression-trajectories categories whose column proportions/values do not differ significantly from each other at the 0·05 level after Bonferroni correction;

Trajectory groups different from 'low depressed group' in **bold**;

Trajectory groups also different from 'consistent depressed group' in ***bold italics.***

Figure 1. Timeline of assessments. GHQ (General Health Questionnaire), MMSE (Mini Mental State Examination), FRS (Framingham Stroke Risk score), MRI (Magnetic Resonance Imaging), COG (Cognitive Test battery described in *{Filippini, 2014 #13}. All measurements were acquired at University College London, except for those indicated in italics, which were obtained at the University of Oxford.*

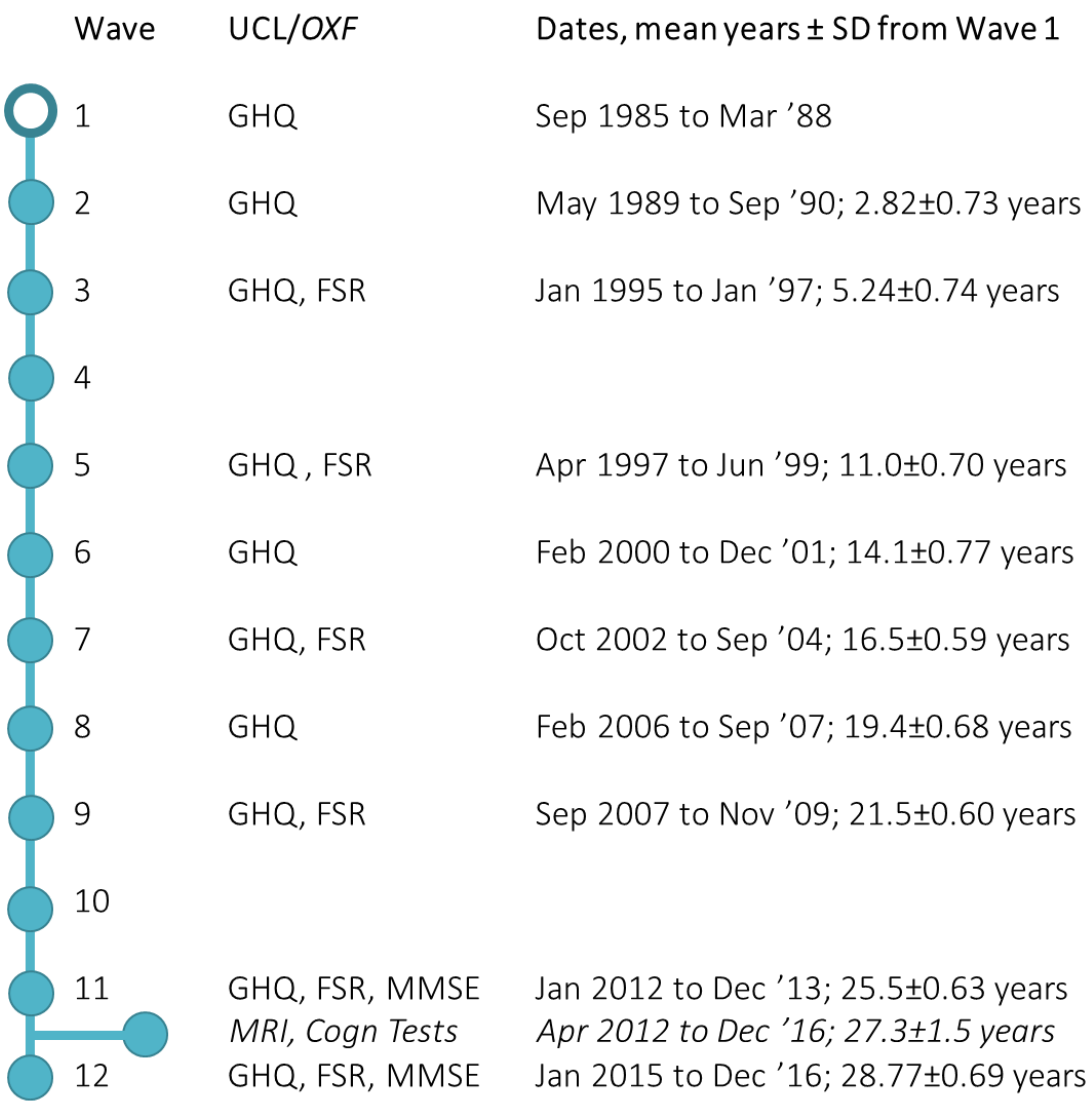


Figure 2a. Trajectories of depressive symptoms from Wave 1 (1985-1987) to Wave 11 (2012-2013), with a mean follow-up time of 25·6±0·6 years. The figure shows the five estimated trajectories of depressive symptoms measured in 774 participants using the GHQ-depression subscale. Solid lines indicate the mean trajectory, while the dotted lines connect mean individual scores for each wave.

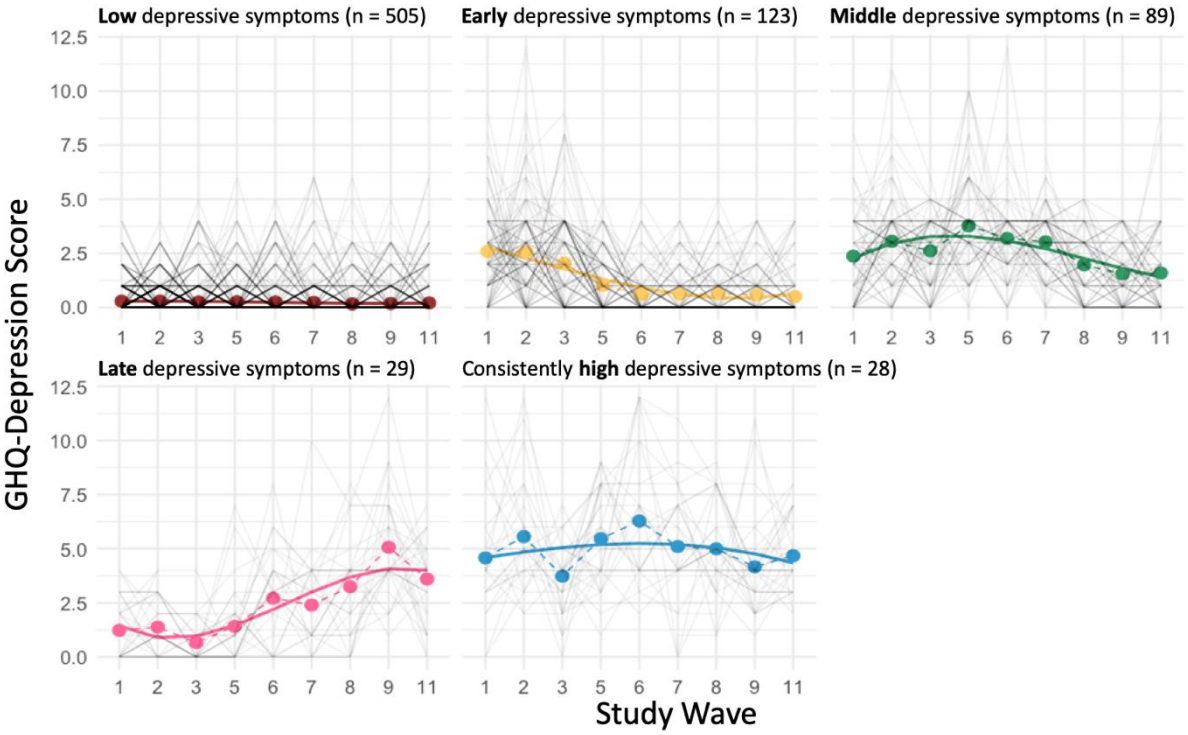


Figure 2b. Percent GHQ cases by wave in each depression trajectory group. Cases ≈ GHQ > 3.

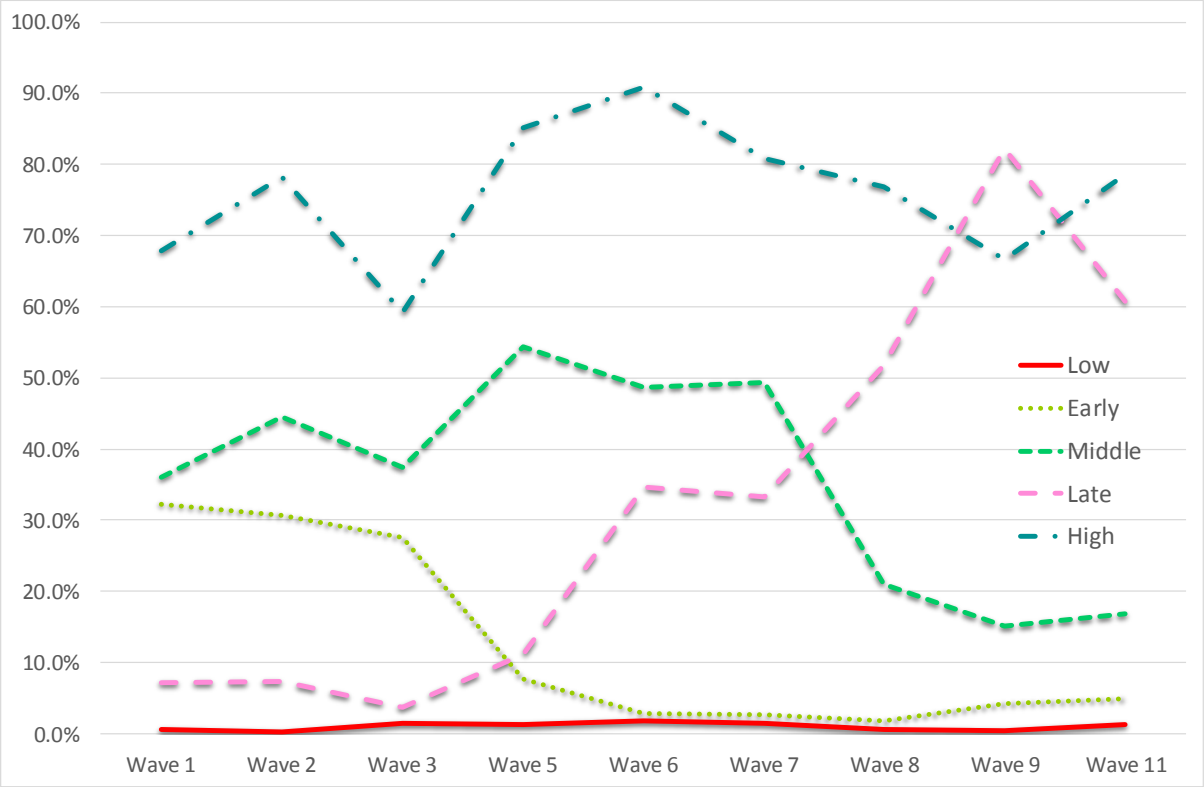


Figure 3. Framingham vascular (stroke) risks in estimated GHQ trajectory groups.

The “late” symptom onset group had higher vascular risks than the “low” symptom group in waves 3, 9, and 11. In contrast, the consistently “high” symptom group had similar vascular risk as the “low” symptom group during all waves. The “late” symptom and the consistently “high” symptom group had significantly different vascular risks in wave 9 and on average across all waves (Table 3).

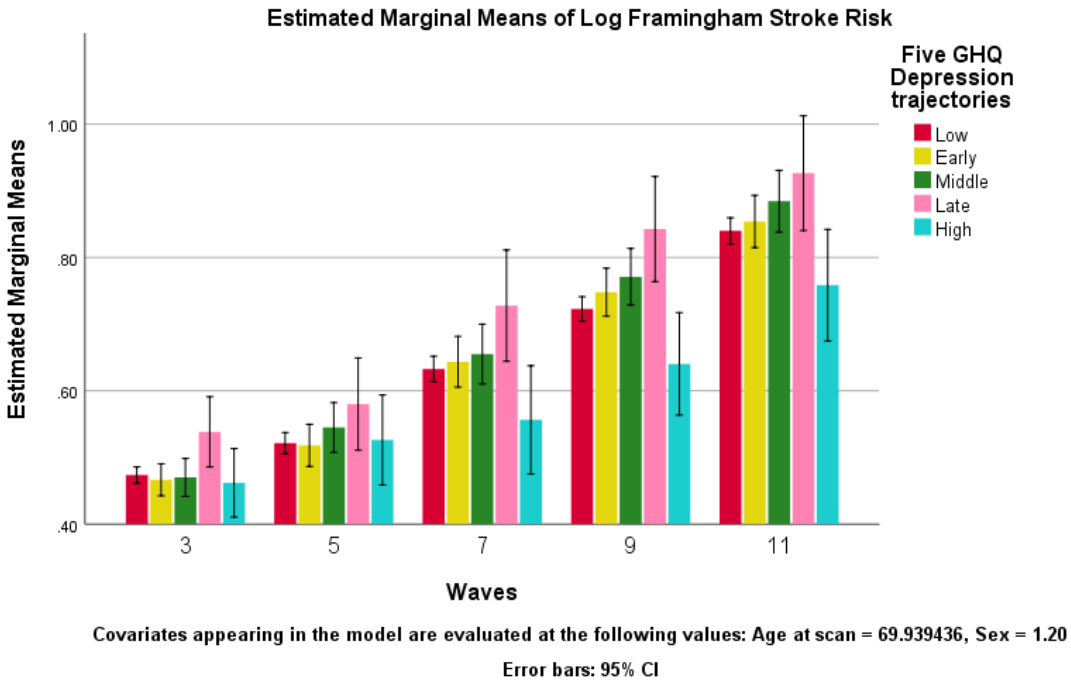
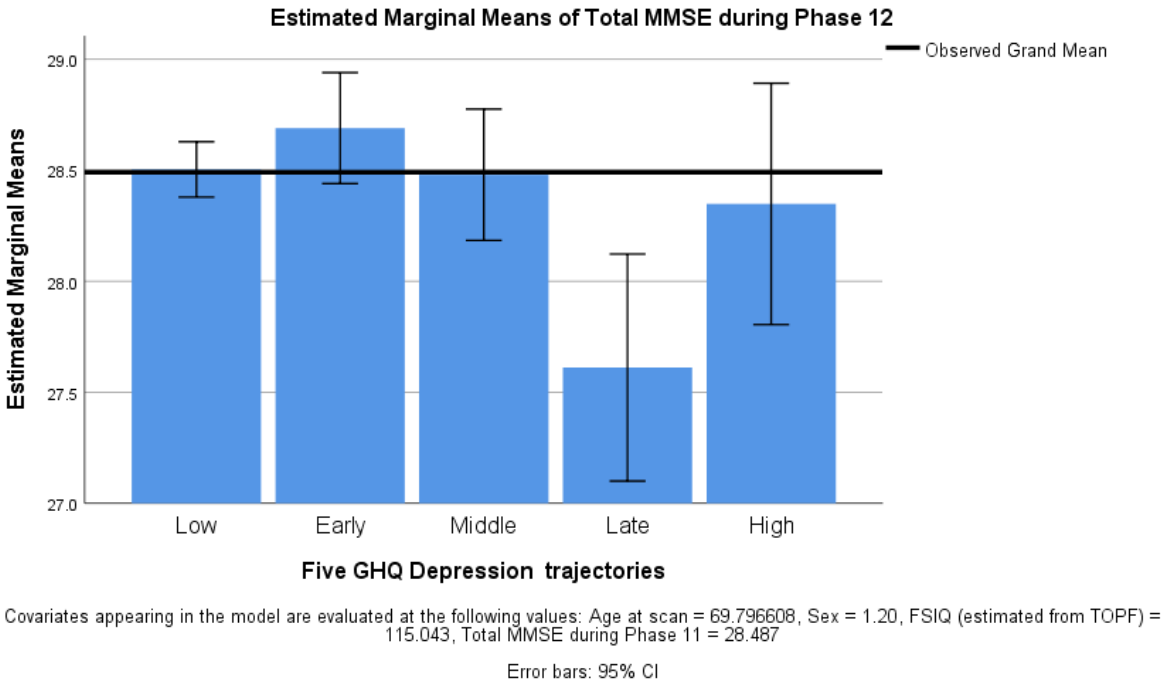


Figure 4: Comparison of MMSE score changes from Waves 11 to 12 between depression trajectory groups. After transformation of MMSE scores to Log (31-MMSE), the between-groups omnibus test was $F_{(4,714)}$=3·704, p=0·005. After Bonferroni correction, the "Late" group had significantly lower MMSE scores at Wave 12 than the "Low" and "Early" symptom groups.



Supplementary Information.

Appendix 1. Overview of missing GHQ data.

	Number of missing GHQ data-points				
	5	**4**	**3**	**2**	**1**
Number of participants (N, %)	8, 1%	9, 1·2%	22, 2·8%	52, 6·7%	129, 16·7%

The *Whitehall II MRI Subgroup* is (self-) selected, in that although 800 were randomly approached from the larger cohort (of n=6306 at the time of selection), participants who cannot travel far, do not tolerate an MRI, or have other objections against a more intense study self-excluded. As a result, there were fewer females with on average lower CES-D scores, although the groups were well age-matched.

Variable	**Study Sample**			**Wave 11 Participants**			**p-value**
	N	Mean / %	S.D.	N	Mean / %	S.D.	
Age [years]	774	69.8	5.2	6306	69.8	5.9	p=0.999
Sex Female Male	774 147 627	 19% 81%		6306 1947 4459	 29.3% 70.7%		p<0.0005
Full time education [years]	774	14.8	3.3	5101	15.1	4.2	p=0.06
CES-D	750	6.2	6.3	5855	7.3	7.6	p<0.001

Appendix 2. Allocation to a late-onset trajectory group using two methods.

To assess whether the late-onset group identified by kml could be replicated using an alternative method of classification, latent class growth models (LCGM) were performed. To facilitate comparison with the 5-group output obtained with kml, the 5-class cubic solution was selected from the LCGM analyses. For this analysis, variances of GHQ depressive symptom scores were constrained to be equal over time. LCGM analyses were conducted in MPLUS 8 (version 1.6).

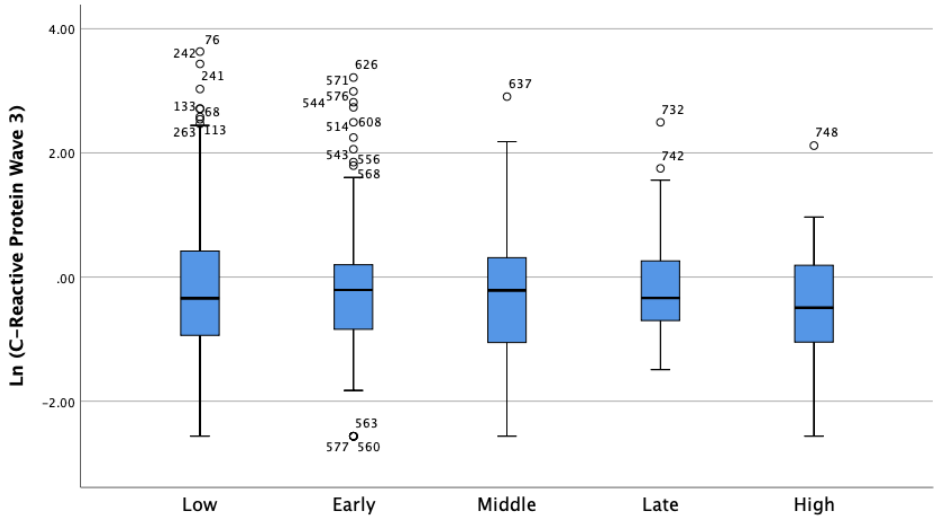
According to LCGMs, the five-class solution also indicated that there was a class of individuals with increasing depressive symptoms in late life (n = 37, 4·6%). To compare allocation into the late-onset

trajectory group between the two methods, McNemar's chi-squared test was performed. For this, a dichotomous variable was created, based on whether each participant was allocated, or not, to the late-onset group. This binary group allocation was then compared between the two methods (kml and LCGM). McNemar's test determined that there was no significant difference between the participants assigned to the late-onset group between the two methods, suggesting that we cannot reject the null hypothesis that the two classifiers made similar predictions (p = 0·36).

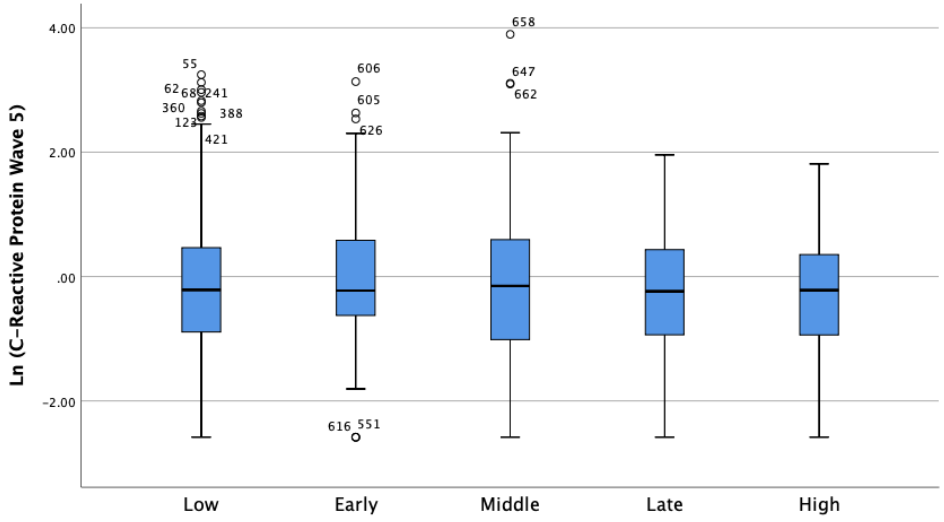
Appendix 3. Inflammatory Markers during previous waves of the Whitehall II study.

Inflammatory markers did not show a similar pattern to vascular risk, i.e. specifically higher measures on the "Late" group in Waves 3, 5 and 7. For methodological details, please see Kivimaki et al. (2014) and Virtanen et al. (2015).

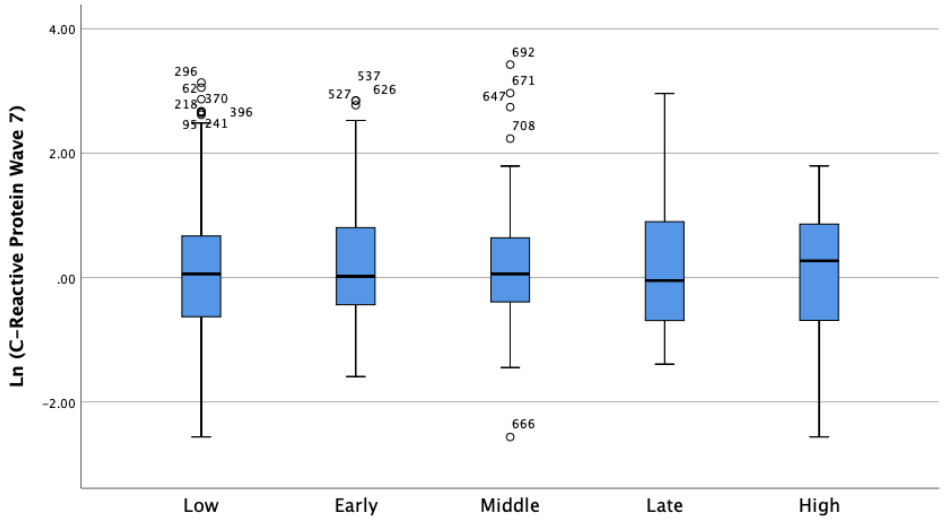
Ln (C-Reactive Protein Wave 3):



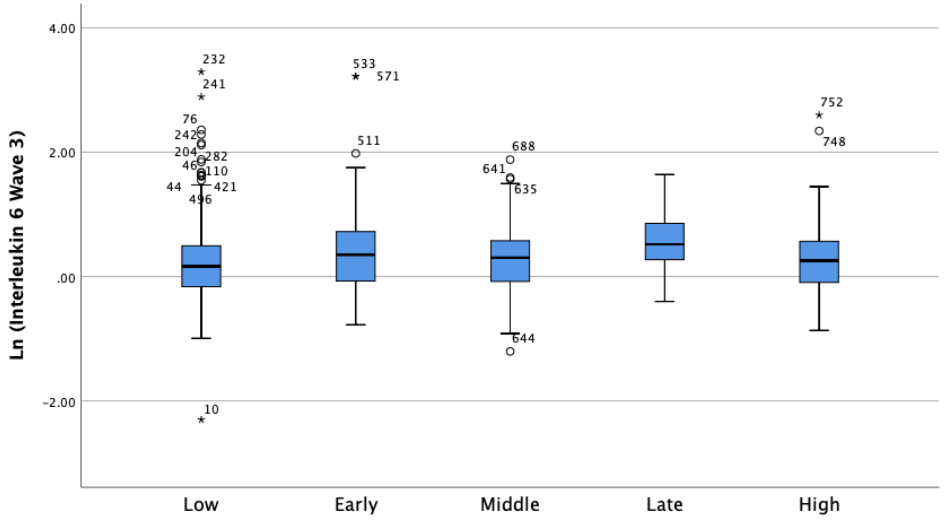
Ln (C-Reactive Protein Wave 5):


Ln (C-Reactive Protein Wave 5)
4.00
2.00
.00
-2.00
55
62
68
241
360
388
123
421
606
605
626
616
551
658
647
662
Low
Early
Middle
Late
High

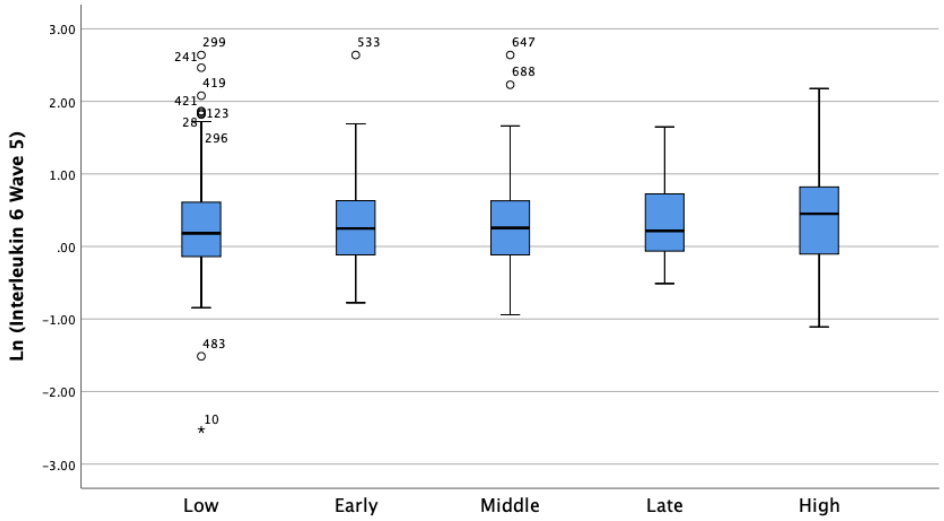
Ln (C-Reactive Protein Wave 7):



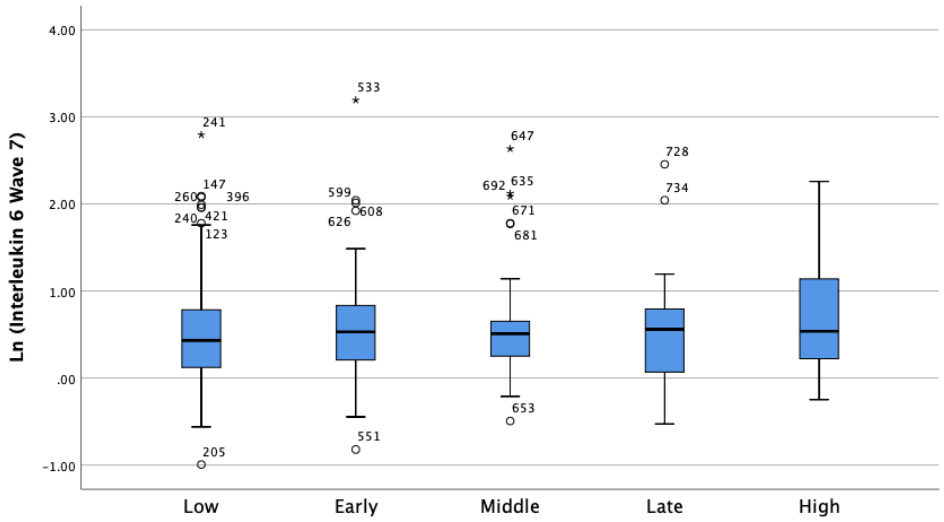
Ln (Interleukin 6 Wave 3)



Ln (Interleukin 6 Wave 5)



Ln (Interleukin 6 Wave 7)



Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

No financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work has been declared by the authors.

CRediT author statement:

Naiara Demnitz: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing
Melis Anatürk: Conceptualization, Methodology, Formal analysis, Writing - Review & Editing
Charlotte L. Allan: Investigation, Project administration, Writing - Review & Editing
Nicola Filippini: Investigation, Project administration, Formal analysis, Data Curation, Writing - Review & Editing
Ludovica Griffanti: Methodology, Formal analysis, Writing - Review & Editing
Clare E. Mackay: Funding acquisition, Supervision, Writing - Review & Editing
Abda Mahmood: Investigation, Project administration, Data Curation, Writing - Review & Editing
Claire E. Sexton: Project administration, Data Curation, Supervision, Writing - Review & Editing
Sana Suri: Project administration, Formal analysis, Data Curation, Writing - Review & Editing
Anya G. Topiwala: Investigation, Writing - Review & Editing
Enikő Zsoldos: Investigation, Project administration, Supervision, Data Curation, Writing - Review & Editing
Mika Kivimäki: Funding acquisition, Conceptualization, Data Curation, Writing - Review & Editing
Archana Singh-Manoux: Funding acquisition, Conceptualization, Data Curation, Writing - Review & Editing
Klaus P. Ebmeier: Conceptualization, Funding acquisition, Supervision, Project administration, Data Curation, Formal analysis, Writing - Review & Editing