Segmenting the human brain in population-based studies

Please check the document version of this publication:

Download date: 13 Dec. 2023

Segmenting the human brain in population-based studies:

Methodological considerations and clinical applications in diabetes, depression, and dementia

Jennifer Monereo-Sánchez

THE MAASTRICHT STUDY

Segmenting the human brain in population-based studies:

Methodological considerations and clinical applications
in diabetes, depression, and dementia

Segmenting the human brain in population-based studies:

Methodological considerations and clinical applications
in diabetes, depression, and dementia

Thesis

to obtain the degree of doctor at Maastricht University,

on the authority of the Rector Magnificus,

Prof. dr. Pamela Habibović

in accordance with the decision of the Board of Deans,

to defend in public on

Monday, 18th December 2023, at 10:00 hours

by

Jennifer Monereo-Sánchez

Born on July 26, 1988 in Barcelona

Supervisors

Prof. dr. David E.J. Linden

Prof. dr. Miranda T. Schram

Dr. Jacobus F.A. Jansen

Assessment committee

Prof. dr. Thérèse A.M.J. van Amelsvoort (chair)

Dr. Liesbeth Goossens

Dr. Thomas van Sloten (University Medical Center Utrecht, Utrecht, The Netherlands)

Prof. dr. Meike W. Vernooij (Erasmus University Rotterdam, Rotterdam, The Netherlands)

Prof. dr. Pieter J. Visser

TABLE OF CONTENTS

CHAPTER 1
GENERAL INTRODUCTION

Neuroimaging in population-based cohort studies

Neuroimaging through MRI allows the non-invasive acquisition of high spatial resolution images of the brain. This technique has been extensively used in neuroscience research to study the structure, function, and connectivity of the brain, enabling significant advances in understanding the neurobiological underpinnings of brain aging and pathophysiology.

Population-based cohorts collect data on a wide range of variables from many individuals, enabling the research of the complex interaction between genetics, environmental factors, lifestyle choices, and disease development. The use of neuroimaging from large population based cohorts further improves our understanding of the different mechanisms that occur in the brain over time or under specific circumstances, which is valuable in understanding specific metabolic conditions, such as type 2 diabetes mellitus (T2DM), neurodegenerative diseases like Alzheimer's disease, or psychiatric disorders, such as depression.

This dissertation uses data from two extensive neuroimaging cohorts, namely The Maastricht Study and the United Kingdom (UK) Biobank. These cohorts have played a pivotal role in health research in the last years. For instance, contributing to the identification of cardiovascular risk factors by shedding light on the underlying aspects that contribute to the physiopathology of the nervous system (Cox et al., 2019; van der Velde et al., 2020). Moreover, they have been instrumental in the discovery of genetic markers associated with diseases and traits, for example enhancing our understanding of the genetic underpinnings of schizophrenia (Warland, Kendall, Rees, Kirov, & Caseras, 2020). Additionally, these cohorts have been invaluable in validating biomarkers for the early detection and prognostic prediction of numerous diseases, such as retinal indices as biomarkers for brain neurodegeneration (van der Heide et al., 2023), or brain biomarkers for Alzheimer's disease (Du et al., 2021), providing crucial insights into potential diagnostic tools. Lastly, these cohorts have contributed to our comprehension of the trajectory of diseases, such as depression (Geraets et al., 2021; Geraets, Köhler, et al., 2022; Geraets, Schram, et al., 2022; Sarris et al., 2020), or cardiovascular disease (Shang et al., 2022), enabling the investigation of the natural progression and factors influencing these conditions over time.

The Maastricht Study

Chapters 2, **3**, and **4** use data from The Maastricht Study. The Maastricht Study is an ongoing observational prospective population-based cohort study (Schram et al., 2014). The study aims to investigate the etiology, pathophysiology, complications, and comorbidities of T2DM, including depression and cognitive decline. The eligible participants were individuals residing in the southern part of the Netherlands, aged between 40 and 75 years. To ensure a diverse sample, multiple recruitment strategies were employed, including mass media campaigns, outreach through municipal registries, and direct contact via mailings to the regional Diabetes Patient Registry. To enhance efficiency, recruitment was stratified based on known T2DM status, with a

deliberate oversampling of individuals with T2DM. This approach allows for a comprehensive investigation of the factors associated with T2DM and its' comorbid chronic conditions, ensuring robust findings and maximizing the research's overall impact.

The Maastricht Study is characterized by an extensive phenotyping approach, and data from various domains is used in this thesis. An oral glucose tolerance test was used to assess glucose metabolism status. This allowed to evaluate the brain associations with multiple continuous markers of hyperglycemia, and further categorize T2DM according to the World Health Organization 2006 criteria (World Health Organization, 2006). In addition, The Maastricht Study includes an assessment of depressive symptoms over time, an annual follow-up providing unique insight in the associations of brain structure with both incidence and course of depressive symptoms. Further, the extensive phenotyping approach made it possible to adjust for potential demographic, cardiovascular, and lifestyle confounders, and perform a range of sensitivity analyses. Research under The Maastricht Study has been approved by the institutional medical ethical committee (NL31329.068.10) and the Dutch Ministry of Health, Welfare, and Sports (permit 131088 105234-PG). For this dissertation, data from the first n=7,689 individuals who had completed the baseline survey between November 2010 and January 2018 was used. Among them, 5,204 had also ultra-high field 3T MRI brain images, which were further segmented and parcellated for the measurement of specific brain volumes.

UK Biobank

Chapter 5 of this thesis uses data obtained from the UK Biobank, a comprehensive biomedical database that contains medical and genetic information from half a million volunteers. The primary goal of UK Biobank is to enhance the understanding, prevention, and treatment of a wide range of serious illnesses. The recruitment for the database was conducted between 2006 and 2010, and individuals aged between 40 and 69 years from the UK were eligible to participate (Palmer, 2007). UK Biobank provides comprehensive information on both the phenotype and genotype of its participants, covering a diverse range of health-related outcomes. The data was collected using various methods, including questionnaires, physical measurements, sample assays, accelerometry, genome-wide genotyping, and multimodal imaging (Allen et al., 2012; Palmer, 2007; Sudlow et al., 2015). Notably, approximately 100,000 participants underwent an MRI scan at baseline, and 3,000 of these participants received a follow-up brain MRI scan within two years. Moreover, there are plans to conduct repeat imaging on 60,000 participants (UK Biobank, 2023). UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval. This thesis uses brain MRI images and genetic data from n=30,699 participants.

Brain MRI segmentation and segmentation quality

As previously introduced, **chapters 2**, **3**, and **4** utilize neuroimaging data from The Maastricht Study, obtained through advanced imaging techniques. Data acquisition took place at the Scannexus facilities in Maastricht, the Netherlands, utilizing a Siemens Prisma 3T MRI machine equipped with a 64-channel head coil. The imaging protocol employed a diverse range of sequences, including a T1 weighted sequence and a fluid-attenuated inversion recovery (FLAIR) sequence, among others. **Chapter 5** relies on data sourced from the UK Biobank. The imaging data for this chapter was collected at the UK Biobank imaging centers situated in Manchester, Newcastle, and Reading, using a Siemens Skyra 3T MRI machine along with a 32-channel head coil. The comprehensive acquisition protocol included various sequences, including a T1 weighted and FLAIR images.

In order to analyze structural MRI images in relation to clinical parameters, it is necessary to transform the brain images into quantitative measurements, such as volume or thickness, through a process called brain segmentation. Brain segmentation involves dividing the brain into distinct regions or structures based on their anatomical or functional characteristics. The objective is to identify and outline various structures within the brain, including the cerebral cortex, white matter, gray matter, and subcortical structures such as the thalamus, basal ganglia, and hippocampus. This intricate process entails the application of image processing procedures to the data, enabling the differentiation and identification of different regions based on their intensity, texture, shape, and location. Brain segmentation is a complex procedure that requires sophisticated image processing algorithms and expertise in neuroanatomy. There are several software tools available to automate the brain segmentation process. In this thesis, the brain segmentation and parcellation of different structures was performed using FreeSurfer v.6.0 (Fischl, 2012). FreeSurfer is a freely accessible MRI analysis software that can automatically detect the morphological boundaries between cortical and subcortical brain structures. By utilizing automation, FreeSurfer offers an efficient and reproducible segmentation solution suitable for large cohorts.

The segmentation and parcellation of brain images was performed using T1 weighted images in FreeSurfer. When available, FLAIR images were also utilized to enhance the quality of segmentation. The segmentation and parcellation process involved several essential steps. Among them, a motion correction was applied to the images to minimize any distortions caused by subject movement; image intensity normalization was conducted to address any inconsistencies arising from scanner fluctuations and coil variations. Subsequently, a skull strip technique was employed to remove non-brain tissues, followed by volumetric registration to align the images spatially. The segmentation process included the identification and delineation of both cortical and subcortical structures within the brain. Further, smooth and inflate, mapping, and registration steps were performed. Finally, cortical parcellation was completed, resulting in the partitioning of the cortical regions into distinct areas. This comprehensive process generated a segmented and parcellated brain image (see Figure 1.1) along with corresponding tabulated data, providing a detailed representation and measurement of the brain structures' characteristics.
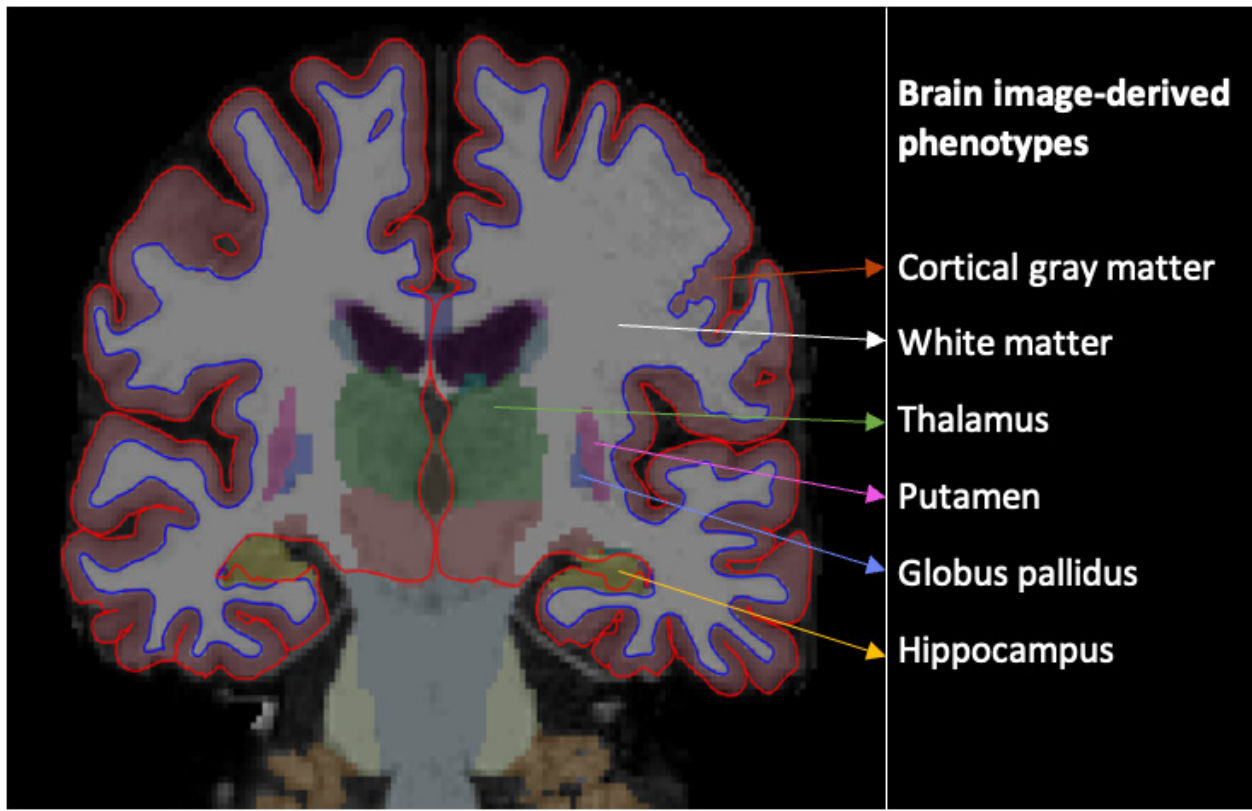
Figure 1.1 | Cortical parcellation and subcortical segmentation of the brain. The image presents a coronal section displaying the segmentation and parcellation results from a volunteer participating in The Maastricht Study. To enhance visual clarity, non-brain structures including the dura mater, skull, and other soft tissues have been removed. The outermost red line represents the pial surface, demarcating the outer boundary of the cortical gray matter. Adjacent to the gray matter, the blue line corresponds to the border between the white matter and the cortical gray matter. These delineations enable the quantification of various measures such as cortical thickness, cortical volume, and cortical surface area. In addition to cortical parcellation, the subcortical segmentation identifies and delineates specific structures within the brain. For instance, the hippocampus is depicted in yellow, the putamen and globus pallidus are displayed in pink and blue, respectively, and the thalamus is represented by the color green.

The automated segmentation and parcellation of the brain may encounter challenges and potential inaccuracies, particularly when dealing with low-quality images or images affected by artifacts. The reliability of research outcomes is directly impacted by the quality of the data employed. Hence, implementing quality control measures for brain MRI segmentations becomes crucial to ensure the trustworthiness of the obtained results. Quality control refers to the process of verifying that the data meet predetermined standards of accuracy, reliability, and validity. Several strategies can be employed to perform quality control on brain segmentation and parcellation. Currently, the gold standard method involves visual inspection and manual editing of brain surfaces. Trained researchers individually review and edit the output images generated by FreeSurfer, addressing any segmentation inaccuracies. Figure 1.2 provides an example showcasing a segmentation error and its subsequent correction through manual intervention. However, when dealing with large neuroimaging datasets, visually inspecting every acquired image for segmentation quality becomes impractical. In such cases,

automated quality control strategies emerge as attractive alternatives. These approaches aim to automate the quality control process using various computational techniques, allowing for efficient and systematic assessment of segmentation quality. The UK Biobank has a very well established quality control pipeline, which incorporates the utilization of Qoala-T (Klapwijk, Van De Kamp, Van Der Meulen, Peters, & Wierenga, 2019), a machine learning-based tool designed to detect potential issues in segmentations. Following this initial automated quality control step, a visual inspection is conducted on all flagged segmentations to ensure their accuracy. **Chapter 2** explores various automated quality control methods and compares them to visual inspection and manual correction, the current gold standard. The objective is to identify a reliable, reproducible, and time-efficient solution suitable for large-scale neuroimaging datasets. The findings were subsequently applied to the full sample in order to ensure the data quality of a large number of brain MRI segmentations within The Maastricht Study.
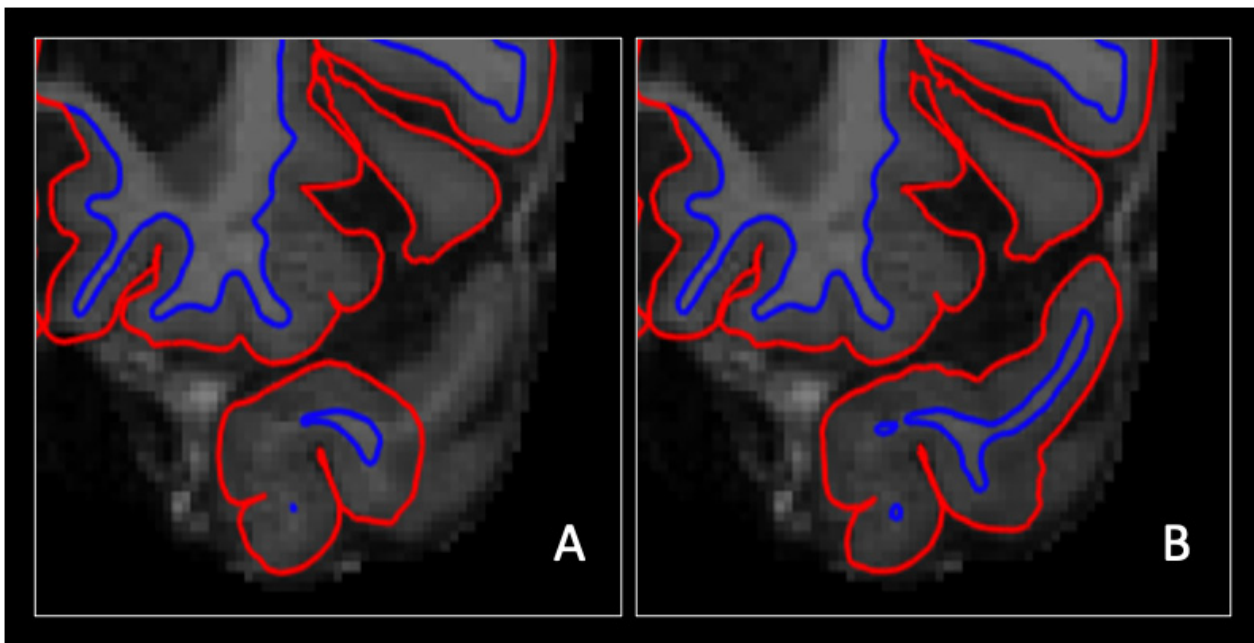
Figure 1.2 | Impact of manual edition on brain surfaces. The zoomed-in view focuses on the left temporal pole of a coronal section from a participant in The Maastricht Study. In Figure 1.2A, an inaccuracy is evident in the lateral region of the left temporal pole. The red line fails to accurately trace the surface of the gray matter, while the blue line does not accurately delineate the boundary between gray and white matter. However, in Figure 1.2B, after manual correction, the surfaces exhibit notable improvements. The lines faithfully follow the actual anatomical boundaries of the structure, providing a more accurate representation of the brain's architecture, and therefore more reliable morphological estimates.

Neuroimaging & hyperglycemia

T2DM develops when the body becomes resistant to insulin and fails to regulate blood glucose effectively, resulting in chronic hyperglycemia. Prediabetes represents an intermediate stage where blood glucose levels are higher than normal but not yet at the threshold for diagnosing T2DM (See Figure 1.3; American Diabetes Association, 2019).

Prolonged hyperglycemia is associated with an increased risk of vascular complications, which can include cardiovascular diseases such as stroke, myocardial infarctions, and microvascular disease, including nephropathy, retinopathy, or peripheral neuropathy (Grimaldi & Heurtier, 1999; Schalkwijk & Stehouwer, 2005). T2DM has also been associated with higher risk of depression (Darwish, Beroncal, Sison, & Swardfager, 2018) and dementia (Hsu, Wahlqvist, Lee, & Tsai, 2011). The prevalence of both T2DM and prediabetes is rapidly increasing on a global scale (Zhou et al., 2016), and the World Health Organization estimates that the number of individuals with diabetes will exceed 350 million by 2030 (World Health Organization, 2003). The recent 3% rise in the age-standardized mortality rate attributed to T2DM underlines the growing importance of improved prevention and treatment of T2DM for public health. Even more alarming is the staggering 13% increase in mortality rates in lower-middle income countries specifically affected by T2DM (World Health Organization, 2023). This growth of prevalence and mortality of T2DM emphasizes the urgent need for effective prevention and management strategies to mitigate the impact of T2DM and reduce the burden of associated vascular complications.

Brain damage can be considered as a (late) complication of T2DM. Research conducted as part of The Maastricht Study demonstrates that both prediabetes and T2DM are associated with structural volumetrics (van Agtmaal et al., 2018), microstructural connectivity (Vergoossen et al., 2020), and functional brain abnormalities (van Bussel et al., 2016). Moreover, a recent meta-analysis shows that T2DM is associated with reduced volumes in the entire brain and specifically in the hippocampus. (T. Zhang, Shaw, & Cherbuin, 2022). T2DM might differently affect specific hippocampal subfields because each subfield has a unique cellular composition and neurophysiology (Fanselow & Dong, 2010). Previous studies have yielded conflicting findings on hippocampal volumes (Blom et al., 2020; C. Li et al., 2020; M. Li et al., 2020; W. Zhang et al., 2021), emphasizing the need for further research adopting a comprehensive approach, incorporating deep phenotyping, and utilizing large sample sizes to gain a clearer understanding of the topic.
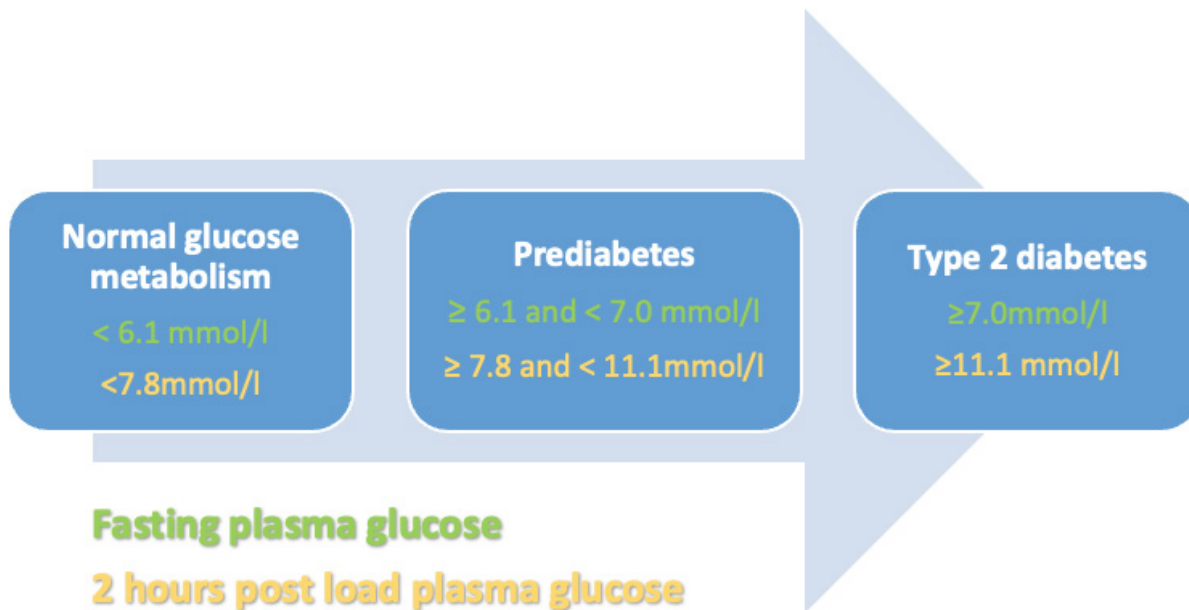
Figure 1.3| *World Health Organization diagnostic criteria for diabetes and intermediate hyperglycemia. The values are referenced according to the WHO 2006 criteria (World Health Organization, 2006) as fasting plasma glucose (in green) and plasma glucose 2 hours after the ingestion of 75g of glucose (yellow).*

Neuroimaging & Depression

A depressive episode is a mood disorder characterized by persistent feelings of sadness, hopelessness, and lack of interest in daily activities. With more than 300 million people affected by depression worldwide, it is the fourth leading cause of disability (Üstün, Ayuso-Mateos, Chatterji, Mathers, & Murray, 2004), and its prevalence has been increasing over the last decades (Moreno-Agostino et al., 2021). This is of special interest not only due to the morbidity, disability, and increased mortality associated with the disease itself (Evans et al., 2005) but also because of the related comorbidities, which include anxiety, eating disorders, and substance abuse (Richards, 2011).

The etiology of depression is currently explained as a combination of genetic and environmental factors. Stress, inflammation, and alterations in monoaminergic systems are suggested as underlying mechanisms (Saveanu & Nemeroff, 2012). From a neuroscience perspective, this involves several brain regions and associated pathways. Neuroimaging studies show that both depressive symptoms and peripheral inflammatory markers are associated with lower volumes in the hippocampus (Videbech & Ravnkilde, 2004), putamen, and caudate nucleus, as well as smaller cortical volumes in frontal regions (Han & Ham, 2021; Koolschijn, van Haren, Lensvelt-Mulders, Hulshoff Pol, & Kahn, 2009; van Tol et al., 2010). Furthermore, increasing evidence suggests an association between depression and neurodegeneration, which has led to

new efforts to understand the relationship between both diseases. According to two systematic reviews, participants with depression show a higher prevalence of neurodegenerative diseases such as Alzheimer's disease when compared to those without depression (Green et al., 2003; Ownby, Crocco, Acevedo, John, & Loewenstein, 2006). Moreover, participants with neurodegenerative diseases show a higher prevalence of depression when compared to the general population (Chi et al., 2015; Fischer, Simamyi, & Danielczyk, 1990; Vilalta-Franch et al., 2006).

This association could be explained by two constructs. On one hand, depression could be an early sign of neurodegenerative diseases, in other words, a prodromal state of the disease (Brommelhoff et al., 2009; Sun et al., 2008). On the other hand, depression could be a risk factor for developing neurodegenerative diseases (Andersen, Lolk, Kragh-Sørensen, Petersen, & Green, 2005; Green et al., 2003; Modrego & Ferrández, 2004; Ownby et al., 2006). The link between chronic depression and significant changes in brain structure and function, particularly affecting the hippocampus, provides a potential explanation for the cognitive decline often observed in depression. Evidence indicates that the persistent inflammation sometimes associated with chronic depression has the potential to contribute to brain cell damage and the abnormal accumulation of proteins, which are characteristic features of neurodegenerative diseases. However, the association between depression and neurodegeneration is still an area of ongoing research. Therefore, it is crucial to recognize the importance of studying depression not only as an isolated event in time but also with a comprehensive longitudinal approach. This approach would enable differentiation between chronic and transient depression, potentially facilitating the characterization of specific depressive conditions. By examining depression over an extended period, researchers can gather valuable insights and develop a deeper understanding of its complex relationship with neurodegenerative processes.

In this thesis, two different definitions of depression were employed. First, **Chapter 4** uses a widely accepted and validated tool to assess depressive symptoms, the Dutch version of the 9-item Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001). The PHQ-9 is a self-administered questionnaire that evaluates the presence of the nine symptoms for the DSM-IV criteria for major depressive disorder. A cut-off of ≥10 was used to define "clinically relevant depressive symptoms", which has shown good sensitivity (88%) and specificity (78%)(Pettersson, Boström, Gustavsson, & Ekselius, 2015). Further, the data on depression was assessed on an annual basis over the course of 7 years, providing valuable longitudinal data on depression. Second, **Chapter 5** uses cross-sectional data, and adopts a definition of depression described by Howard et al. (2018) as 'broad depression'. Participants who were classified as having depression met one of the following criteria: They answered positively to the question "Have you seen a general practitioner or a psychiatrist for nerves, anxiety, tension, or depression?"; or they had a hospital record diagnosis of depression. A GWAS on this definition of depression has shown to be highly genetically correlated with a GWAS using a strict clinical definition of major depressive disorder (r_g=0.85) (Howard et al., 2018)

Neuroimaging & Alzheimer's disease

Alzheimer's disease is a neurodegenerative disease characterized by a gradual loss of memory and other cognitive functions, which affects the ability to perform daily activities. It is the most common form of dementia in people over 65 years old (Cummings & Cole, 2002). Due to changes in lifestyle and environmental factors, as well as an increase in life expectancy, its prevalence is expected to rise over time. Current estimates indicate that, by 2050, 132 million people will suffer from Alzheimer's disease (Prince et al., 2015).

Alzheimer's disease has become an important health problem due to its impact on the quality of life of both patients and caregivers and its associated mortality (Markowitz, Gutterman, Sadik, & Papadopoulos, 2003; Schölzel-Dorenbos, Draskovic, Vernooij-Dassen, & Rikkert, 2009; Shin, Carter, Masterman, Fairbanks, & Cummings, 2005). Despite intensive research in recent years, the exact biological underpinnings of Alzheimer's disease are not yet completely understood. It is well established that Alzheimer's disease is the result of a combination of genetic, lifestyle, and environmental factors. However, the exact contribution of each factor is not yet clear (Gatz et al., 2006). Mutations in the APP, PSEN1 and PSEN2 genes have been associated with familial Alzheimer's disease (Piaceri, Nacmias, & Sorbi, 2013). However, over 95% of Alzheimer's disease cases are sporadic, suggesting that other factors beyond genetic variants affecting the amyloid cascade play an important role in its development (Barykin, Mitkevich, Kozin, & Makarov, 2017). The presence of the apolipoprotein ℇ4 (APOE ℇ4) allele is the most potent identified risk factor for the sporadic form of Alzheimer's disease (Cummings & Cole, 2002).

Alzheimer's disease is associated with changes in brain structure and function. Neuroimaging studies can detect brain degeneration, which has been correlated with an abnormal accumulation of tau protein in the brain and neuropsychological deficits (Bronge, Bogdanovic, & Wahlund, 2002; Frisoni, Fox, Jack Jr, Scheltens, & Thompson, 2010). Patients with Alzheimer's disease suffer from neuron and synaptic connection loss, especially in brain areas important for memory and other cognitive functions, such as the hippocampus and prefrontal cortex (Frisoni et al., 2010; Raji, Lopez, Kuller, Carmichael, & Becker, 2009; Teipel et al., 2013). Alzheimer's disease is associated with a long list of comorbidities, such as depression, anxiety, and cardiovascular disease (Santiago & Potashkin, 2021). The presence of such comorbidities has been associated with a specific course of the disease and the quality of life of patients suffering from it. For example, patients with Alzheimer's disease and comorbid depression show increased Alzheimer's disease-related brain pathology, suggesting an interaction between both diseases (Rapp et al., 2008).

Identifying individuals at risk of Alzheimer's disease poses a significant challenge within population-based cohorts. The pathophysiological changes associated with Alzheimer's disease can initiate many years before the onset of recognizable Alzheimer's disease-related symptoms (Jack et al., 2010; Sperling et al., 2011). Recognizing this preclinical phase of the disease is essential for the mitigation of the impact of Alzheimer's disease. During this prodromal stage, subtle alterations occur in the brain, including the accumulation of beta-amyloid plaques and the development of neurofibrillary tangles, which are hallmark

pathological features of Alzheimer's disease. However, these changes may not immediately manifest as noticeable symptoms or impairments in cognitive function. The analysis of cerebrospinal fluid allows the identification of individuals with preclinical Alzheimer's disease based on the presence of biomarkers before they exhibit noticeable cognitive impairment. However, this is not a feasible solution in large population-based cohorts. As a result, identifying individuals who are at risk of developing Alzheimer's disease becomes a complex task. Population-based cohorts that follow participants over extended periods offer a valuable opportunity to track the progression of Alzheimer's disease-related changes and identify potential biomarkers or risk factors associated with the disease. However, when longitudinal data is not yet available, the use of familial information might suppose a feasible solution. **Chapter 5** uses a measure of Alzheimer's by proxy. This measure was described by Jansen et al. (2019) and has been shown to be highly genetically correlated to Alzheimer's disease status (r_g = 0.81) (Jansen et al., 2019). By utilizing hospital records of both the individuals and their parents, this measure enables to, not only include individuals with diagnosed Alzheimer's disease, but also those at a high risk of developing the disease. The advantage of identifying participants before the onset of any symptoms improves in our ability to study the very early stages of Alzheimer's disease, including the morphological changes associated with this disease, and the pathophysiology of disease progression.

Outline of this thesis

The objectives of this thesis are twofold: to identify and implement a time-efficient and reproducible quality control strategy for large neuroimaging cohorts, and to utilize the resulting high-quality neuroimaging data to examine brain structure under various pathological conditions.

In **Chapter 2**, a range of commonly employed quality control methods are evaluated and compared. Recognizing the challenges of manual quality control in large cohorts, the focus is on establishing an automated quality control approach that is both efficient and reliable for The Maastricht Study dataset. A subsample of The Maastricht Study dataset undergoes visual inspection and manual editing, and the impact of this intervention on brain morphological estimates is documented. Additionally, both manual quality control and multiple automated quality control strategies are applied to the same sample. A regression model is employed to assess the changes in unexplained variance resulting from each quality control strategy, enabling a comparison of their effectiveness. The chapter concludes by identifying the most suitable quality control strategies for several scenarios.

In **Chapter 3**, high-quality cross-sectional data on glucose metabolism status and hippocampal subfield volumes is utilized. The aim is to investigate the associations of prediabetes, T2DM, and continuous measures of hyperglycemia, with the volumes of hippocampal subfields.

Chapter 4 shifts the focus to longitudinal data on depression and hippocampal subfield volumes. The objective is to examine the relationships between the volumes of hippocampal subfields and the prevalence, course, and incidence of depression.

Chapter 5 draws on data from the UK Biobank to explore the genetic overlap between depression and Alzheimer's disease, and its relationship to brain structure.

Overall, this thesis aims to enhance our understanding of brain structure under various pathological conditions by implementing an efficient quality control strategy and utilizing high-quality neuroimaging data.

References

Allen, N., Sudlow, C., Downey, P., Peakman, T., Danesh, J., Elliott, P., . . . Pell, J. (2012). UK Biobank: Current status and what it means for epidemiology. *Health Policy and Technology, 1*(3), 123-126.

American Diabetes Association. (2019). 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2019. *Diabetes care, 42*(Supplement_1), S13-S28.

Andersen, K., Lolk, A., Kragh-Sørensen, P., Petersen, N. E., & Green, A. (2005). Depression and the risk of Alzheimer disease. *Epidemiology*, 233-238.

Barykin, E. P., Mitkevich, V. A., Kozin, S. A., & Makarov, A. A. (2017). Amyloid β modification: a key to the sporadic Alzheimer's disease? *Frontiers in genetics, 8*, 58.

Blom, K., Koek, H. L., Zwartbol, M. H., Ghaznawi, R., Kuijf, H. J., Witkamp, T. D., . . . Group, U.-S. S. (2020). Vascular Risk Factors of Hippocampal Subfield Volumes in Persons without Dementia: The Medea 7T Study. *Journal of Alzheimer's Disease*(Preprint), 1-17.

Brommelhoff, J. A., Gatz, M., Johansson, B., McArdle, J. J., Fratiglioni, L., & Pedersen, N. L. (2009). Depression as a risk factor or prodromal feature for dementia? Findings in a population-based sample of Swedish twins. *Psychology and aging, 24*(2), 373.

Bronge, L., Bogdanovic, N., & Wahlund, L.-O. (2002). Postmortem MRI and histopathology of white matter changes in Alzheimer brains. *Dementia and geriatric cognitive disorders, 13*(4), 205-212.

Chi, S., Wang, C., Jiang, T., Zhu, X.-C., Yu, J.-T., & Tan, L. (2015). The prevalence of depression in Alzheimer's disease: a systematic review and meta-analysis. *Current Alzheimer Research, 12*(2), 189-198.

Cox, S. R., Lyall, D. M., Ritchie, S. J., Bastin, M. E., Harris, M. A., Buchanan, C. R., . . . Reus, L. M. (2019). Associations between vascular risk factors and brain MRI indices in UK Biobank. *European heart journal, 40*(28), 2290-2300.

Cummings, J. L., & Cole, G. (2002). Alzheimer disease. *Jama, 287*(18), 2335-2338.

Darwish, L., Beroncal, E., Sison, M. V., & Swardfager, W. (2018). Depression in people with type 2 diabetes: current perspectives. *Diabetes, metabolic syndrome and obesity: targets and therapy*, 333-343.

Du, J., Liu, Z., Hanford, L. C., Anderson, K. M., Feng, J., Ge, T., & Buckner, R. L. (2021). Exploration of Alzheimer's Disease MRI Biomarkers Using APOE4 Carrier Status in the UK Biobank. *medRxiv*, 2021.2009. 2009.21263324.

Evans, D. L., Charney, D. S., Lewis, L., Golden, R. N., Gorman, J. M., Krishnan, K. R. R., . . . Coyne, J. C. (2005). Mood disorders in the medically ill: scientific review and recommendations. *Biological psychiatry, 58*(3), 175-189.

Fanselow, M. S., & Dong, H.-W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron, 65*(1), 7-19.

Fischer, P., Simamyi, M., & Danielczyk, W. (1990). Depression in dementia of the Alzheimer type and in multi-infarct dementia. *The American journal of psychiatry*.

Fischl, B. (2012). FreeSurfer. *NeuroImage, 62*(2), 774-781.

Frisoni, G. B., Fox, N. C., Jack Jr, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology, 6*(2), 67-77.

Gatz, M., Reynolds, C. A., Fratiglioni, L., Johansson, B., Mortimer, J. A., Berg, S., . . . Pedersen, N. L. (2006). Role of genes and environments for explaining Alzheimer disease. *Archives of general psychiatry, 63*(2), 168-174.

Geraets, A. F., Köhler, S., Jansen, J. F., Eussen, S. J., Stehouwer, C. D., Schaper, N. C., . . . Schram, M. T. (2021). The association of markers of cerebral small vessel disease and brain atrophy with incidence and course of depressive symptoms-the maastricht study. *Journal of affective disorders, 292*, 439-447.

Geraets, A. F., Köhler, S., Vergoossen, L. W., Backes, W. H., Stehouwer, C. D., Verhey, F. R., . . . Schram, M. T. (2022). The association of white matter connectivity with prevalence, incidence and course of depressive symptoms: The Maastricht Study. *Psychological medicine*, 1-11.

Geraets, A. F., Schram, M. T., Jansen, J. F., Backes, W. H., Schalkwijk, C. G., Stehouwer, C. D., . . . Verhey, F. R. (2022). The cardiometabolic depression subtype and its association with clinical characteristics: The Maastricht Study. *Journal of affective disorders, 313*, 110-117.

Green, R. C., Cupples, L. A., Kurz, A., Auerbach, S., Go, R., Sadovnick, D., . . . Edeki, T. (2003). Depression as a risk factor for Alzheimer disease: the MIRAGE Study. *Archives of neurology, 60*(5), 753-759.

Grimaldi, A., & Heurtier, A. (1999). Epidemiology of cardio-vascular complications of diabetes. *Diabetes & metabolism, 25*, 12-20.

Han, K.-M., & Ham, B.-J. (2021). How inflammation affects the brain in depression: a review of functional and structural MRI studies. *Journal of Clinical Neurology (Seoul, Korea), 17*(4), 503.

Howard, D. M., Adams, M. J., Shirali, M., Clarke, T.-K., Marioni, R. E., Davies, G., . . . Barbu, M. C. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature communications, 9*(1), 1470.

Hsu, C.-C., Wahlqvist, M. L., Lee, M.-S., & Tsai, H.-N. (2011). Incidence of dementia is increased in type 2 diabetes and reduced by the use of sulfonylureas and metformin. *Journal of Alzheimer's Disease, 24*(3), 485-493.

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., . . . Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology, 9*(1), 119-128.

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., . . . Athanasiu, L. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature genetics, 51*(3), 404-413.

Klapwijk, E. T., Van De Kamp, F., Van Der Meulen, M., Peters, S., & Wierenga, L. M. (2019). Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data. *Neuroimage, 189*, 116-129.

Koolschijn, P. C. M., van Haren, N. E., Lensvelt-Mulders, G. J., Hulshoff Pol, H. E., & Kahn, R. S. (2009). Brain volume abnormalities in major depressive disorder: A meta-analysis of magnetic resonance imaging studies. *Human brain mapping, 30*(11), 3719-3735.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine, 16*(9), 606-613.

Li, C., Zuo, Z., Liu, D., Jiang, R., Li, Y., Li, H., . . . Xiong, K. (2020). Type 2 Diabetes Mellitus May Exacerbate Gray Matter Atrophy in Patients With Early-Onset Mild Cognitive Impairment. *Frontiers in neuroscience, 14*, 856.

Li, M., Huang, L., Yang, D., Luo, C., Qin, R., Zhang, B., . . . Xu, Y. (2020). Atrophy patterns of hippocampal subfields in T2DM patients with cognitive impairment. *Endocrine, 68*(3), 536.

Markowitz, J. S., Gutterman, E. M., Sadik, K., & Papadopoulos, G. (2003). Health-related quality of life for caregivers of patients with Alzheimer disease. *Alzheimer Disease & Associated Disorders, 17*(4), 209-214.

Modrego, P. J., & Ferrández, J. (2004). Depression in patients with mild cognitive impairment increases the risk of developing dementia of Alzheimer type: a prospective cohort study. *Archives of neurology, 61*(8), 1290-1293.

Moreno-Agostino, D., Wu, Y.-T., Daskalopoulou, C., Hasan, M. T., Huisman, M., & Prina, M. (2021). Global trends in the prevalence and incidence of depression: a systematic review and meta-analysis. *Journal of affective disorders, 281*, 235-243.

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. *Archives of general psychiatry, 63*(5), 530-538.

Palmer, L. J. (2007). UK Biobank: bank on it. *The Lancet, 369*(9578), 1980-1982.

Pettersson, A., Boström, K. B., Gustavsson, P., & Ekselius, L. (2015). Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nordic journal of psychiatry, 69*(7), 497-508.

Piaceri, I., Nacmias, B., & Sorbi, S. (2013). Genetics of familial and sporadic Alzheimer's disease. *Frontiers in Bioscience-Elite, 5*(1), 167-177.

Prince, M. J., Wimo, A., Guerchet, M. M., Ali, G. C., Wu, Y.-T., & Prina, M. (2015). World Alzheimer Report 2015-The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends.

Raji, C. A., Lopez, O., Kuller, L., Carmichael, O., & Becker, J. (2009). Age, Alzheimer disease, and brain structure. *Neurology, 73*(22), 1899-1905.

Rapp, M. A., Schnaider-Beeri, M., Purohit, D. P., Perl, D. P., Haroutunian, V., & Sano, M. (2008). Increased neurofibrillary tangles in patients with Alzheimer disease with comorbid depression. *The American journal of geriatric psychiatry, 16*(2), 168-174.

Richards, D. (2011). Prevalence and clinical course of depression: a review. *Clinical psychology review, 31*(7), 1117-1125.

Santiago, J. A., & Potashkin, J. A. (2021). The impact of disease comorbidities in Alzheimer's disease. *Frontiers in aging neuroscience, 13*, 631770.

Sarris, J., Thomson, R., Hargraves, F., Eaton, M., de Manincor, M., Veronese, N., . . . Firth, J. (2020). Multiple lifestyle factors and depressed mood: a cross-sectional and longitudinal analysis of the UK Biobank (N= 84,860). *BMC medicine, 18*, 1-10.

Saveanu, R. V., & Nemeroff, C. B. (2012). Etiology of depression: genetic and environmental factors. *Psychiatric clinics, 35*(1), 51-71.

Schalkwijk, C. G., & Stehouwer, C. D. (2005). Vascular complications in diabetes mellitus: the role of endothelial dysfunction. *Clinical science, 109*(2), 143-159.

Schölzel-Dorenbos, C. J., Draskovic, I., Vernooij-Dassen, M. J., & Rikkert, M. G. O. (2009). Quality of life and burden of spouses of Alzheimer disease patients. *Alzheimer Disease & Associated Disorders, 23*(2), 171-177.

Schram, M. T., Sep, S. J., van der Kallen, C. J., Dagnelie, P. C., Koster, A., Schaper, N., . . . Stehouwer, C. D. (2014). The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *European journal of epidemiology, 29*(6), 439-451.

Shang, X., Zhang, X., Huang, Y., Zhu, Z., Zhang, X., Liu, S., . . . Yu, H. (2022). Temporal trajectories of important diseases in the life course and premature mortality in the UK Biobank. *BMC medicine, 20*(1), 1-13.

Shin, I.-S., Carter, M., Masterman, D., Fairbanks, L., & Cummings, J. L. (2005). Neuropsychiatric symptoms and quality of life in Alzheimer disease. *The American journal of geriatric psychiatry, 13*(6), 469-474.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., . . . Montine, T. J. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 280-292.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine, 12*(3), e1001779.

Sun, X., Steffens, D. C., Au, R., Folstein, M., Summergrad, P., Yee, J., . . . Qiu, W. Q. (2008). Amyloid-associated depression: a prodromal depression of Alzheimer disease? *Archives of general psychiatry, 65*(5), 542-550.

Teipel, S. J., Grothe, M., Lista, S., Toschi, N., Garaci, F. G., & Hampel, H. (2013). Relevance of magnetic resonance imaging for early detection and diagnosis of Alzheimer disease. *Medical Clinics, 97*(3), 399-424.

UK Biobank. (2023, March 31st 2023). Imaging data. Retrieved from https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/imaging-data

Üstün, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C., & Murray, C. J. (2004). Global burden of depressive disorders in the year 2000. *The British journal of psychiatry, 184*(5), 386-392.

van Agtmaal, M. J., Houben, A. J., de Wit, V., Henry, R. M., Schaper, N. C., Dagnelie, P. C., . . . Kroon, A. A. (2018). Prediabetes is associated with structural brain abnormalities: the Maastricht Study. *Diabetes care, 41*(12), 2535-2543.

van Bussel, F. C., Backes, W. H., van Veenendaal, T. M., Hofman, P. A., van Boxtel, M. P., Schram, M. T., . . . Stehouwer, C. D. (2016). Functional brain networks are altered in type 2 diabetes and prediabetes: signs for compensation of cognitive decrements? The Maastricht Study. *Diabetes, 65*(8), 2404-2413.

van der Heide, F. C., Mokhtar, S., Khanna, A., Said, M., Henry, R., Kroon, A. A., . . . Schouten, J. S. (2023). Retinal Functional and Structural Neural Indices: Potential Biomarkers for the Monitoring of Cerebral Neurodegeneration: The Maastricht Study. *Journal of Alzheimer's Disease*(Preprint), 1-13.

van der Velde, J. H., Koster, A., Strotmeyer, E. S., Mess, W. H., Hilkman, D., Reulen, J. P., . . . van der Kallen, C. J. (2020). Cardiometabolic risk factors as determinants of peripheral nerve function: the Maastricht Study. *Diabetologia, 63*, 1648-1658.

van Tol, M.-J., van der Wee, N. J., van den Heuvel, O. A., Nielen, M. M., Demenescu, L. R., Aleman, A., . . . Veltman, D. J. (2010). Regional brain volume in depression and anxiety disorders. *Archives of general psychiatry, 67*(10), 1002-1011.

Vergoossen, L. W., Schram, M. T., de Jong, J. J., Stehouwer, C. D., Schaper, N. C., Henry, R. M., . . . Eussen, S. J. (2020). White matter connectivity abnormalities in prediabetes and type 2 diabetes: The Maastricht Study. *Diabetes care, 43*(1), 201-208.

Videbech, P., & Ravnkilde, B. (2004). Hippocampal volume and depression: a meta-analysis of MRI studies. *American Journal of Psychiatry, 161*(11), 1957-1966.

Vilalta-Franch, J., Garre-Olmo, J., López-Pousa, S., Turon-Estrada, A., Lozano-Gallego, M., Hernàndez-Ferràndiz, M., . . . Feijóo-Lorza, R. (2006). Comparison of different clinical diagnostic criteria for depression in Alzheimer disease. *The American journal of geriatric psychiatry, 14*(7), 589-597.

Warland, A., Kendall, K. M., Rees, E., Kirov, G., & Caseras, X. (2020). Schizophrenia-associated genomic copy number variants and subcortical brain volumes in the UK Biobank. *Molecular psychiatry, 25*(4), 854-862.

World Health Organization. (2003). *Screening for type 2 diabetes: report of a World Health Organization and International Diabetes Federation meeting*. Retrieved from https://apps.who.int/iris/handle/10665/68614

World Health Organization. (2006). *Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation* (9241594934). Retrieved from https://apps.who.int/iris/handle/10665/43588

World Health Organization. (2023, 5 April 2023). Diabetes. Retrieved from https://www.who.int/news-room/fact-sheets/detail/diabetes

Zhang, T., Shaw, M., & Cherbuin, N. (2022). Association between type 2 diabetes mellitus and brain atrophy: a meta-analysis. *Diabetes & metabolism journal, 46*(5), 781-802.

Zhang, W., Gao, C., Qing, Z., Zhang, Z., Bi, Y., Zeng, W., & Zhang, B. (2021). Hippocampal subfields atrophy contribute more to cognitive impairment in middle-aged patients with type 2 diabetes rather than microvascular lesions. *Acta Diabetologica*, 1-11.

Zhou, B., Lu, Y., Hajifathalian, K., Bentham, J., Di Cesare, M., Danaei, G., . . . Taddei, C. (2016). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4· 4 million participants. *The Lancet, 387*(10027), 1513-1530.

CHAPTER 2

QUALITY CONTROL STRATEGIES FOR BRAIN MRI SEGMENTATION AND PARCELLATION: PRACTICAL APPROACHES AND RECOMMENDATIONS, INSIGHTS FROM THE MAASTRICHT STUDY

Monereo-Sánchez, J., de Jong, J.J., Drenthen, G.S., Beran, M., Backes, W.H.,
Stehouwer, C.D., Schram, M.T., Linden, D.E. and Jansen, J.F.

Abstract

Quality control (QC) of brain segmentation is a fundamental step to ensure data quality. Manual quality control strategies are the current gold standard, although these may be unfeasible for large neuroimaging samples. Several options for automated quality control have been proposed, providing potential time efficient and reproducible alternatives. However, those have never been compared side to side, which prevents consensus in the appropriate QC strategy to use. This study aimed to elucidate the changes manual editing of brain segmentations produce in morphological estimates, and to analyze and compare the effects of different quality control strategies on the reduction of the measurement error.

Structural MR images from 259 participants of The Maastricht Study were used. Morphological estimates were automatically extracted using FreeSurfer 6.0. Segmentations with inaccuracies were manually edited, and morphological estimates were compared before and after editing. In parallel, 12 quality control strategies were applied to the full sample. Those included: two manual strategies, in which images were visually inspected and either excluded or manually edited; five automated strategies, where outliers were excluded based on the tools "MRIQC" and "Qoala-T", and the metrics "morphological global measures", "Euler numbers" and "Contrast-to-Noise ratio"; and five semi-automated strategies, where the outliers detected through the mentioned tools and metrics were not excluded, but visually inspected and manually edited. In order to quantify the effects of each QC strategy, the proportion of unexplained variance relative to the total variance was extracted after the application of each QC strategy, and the resulting differences compared.

Manually editing brain surfaces produced particularly large changes in subcortical brain volumes and moderate changes in cortical surface area, thickness and hippocampal volumes. The performance of the quality control strategies depended on the morphological measure of interest. Manual quality control strategies yielded the largest reduction in relative unexplained variance. The best performing automated alternatives were those based on Euler numbers and MRIQC scores. The exclusion of outliers based on global morphological measures produced an increase of relative unexplained variance.

Manual quality control strategies are the most reliable solution for quality control of brain segmentation and parcellation. However, measures must be taken to prevent the subjectivity associated with these strategies. The detection of inaccurate segmentations based on Euler numbers and MRIQC provide a time efficient and reproducible alternative. The exclusion of outliers based on global morphological estimates must be avoided.

Introduction

Quality control (QC) of brain MRI segmentation and parcellation, i.e. the detection and correction or exclusion of inaccuracies in segmented brain images, is a fundamental step to ensure measurement reliability. The concept of brain segmentation QC has recently gained interest, with many tools, metrics, and protocols being proposed (Lea L Backhausen et al., 2016; Esteban et al., 2017; Keshavan et al., 2018; Klapwijk, Van De Kamp, Van Der Meulen, Peters, & Wierenga, 2019; Raamana et al., 2020; Rosen et al., 2018; Waters, Mace, Sawyer, & Gansler, 2019). Manual QC strategies, consisting of visual inspection (with or without surface editing), despite their component of subjectivity, are considered the gold standard. However, neuroimaging research is shifting towards big data paradigms, with studies including thousands of brain images, such as the UK Biobank (Miller et al., 2016), or The Maastricht Study (Schram et al., 2014). Manual QC strategies have therefore become unfeasible, given the time and resources required. Thus there is a need to validate and agree on a reproducible and time efficient QC solution for large cohort studies.

Poor quality segmentation is characterized by the presence of inaccuracies. Inaccuracies occur when the boundaries that define the morphological divisions or regions of interest (ROIs) do not correspond to the anatomical boundaries, which may lead to morphological measurement errors. A commonly used tool to segment structural brain MRI is FreeSurfer (Fischl, 2012), a software for MRI analysis that provides automated subcortical segmentation and cortical parcellation of the brain. Errors in FreeSurfer's output may happen (amongst others) when sufficiently abnormal brain structure or low-quality image is provided as input. Further, image artifacts have been related to worse segmentation estimates for both cortical thickness (Reuter et al., 2015) and volumes (Savalia et al., 2017). Errors in the segmentation may result in regression attenuation (Hutcheon, Chiolero, & Hanley, 2010) as well as reduction of statistical power (Phillips & Jiang, 2016) in regression analysis of phenotypic measures with MR features. Large sample sizes can compensate for these downsides. However, when the measurement errors are systematic, recurrent, and in the same direction, a bias can be introduced, making segmentation quality a potential confounder. This type of bias has previously been shown in clinical populations compared to healthy controls (Pardoe, Hiess, & Kuzniecky, 2016), children compared to adults (Blumenthal, Zijdenbos, Molloy, & Giedd, 2002) and older adults compared to younger adults (Madan, 2018; Savalia et al., 2017; Wenger et al., 2014).

Manual QC strategies are currently the most accepted approach to ensure reliable segmentation estimates and, in absence of a better solution, they are considered the gold standard for QC. The manual QC process involves the visual inspection of each segmentation, ideally by several independent operators with knowledge of neuroimaging normal anatomy, with the possibility to manual edit the segmentations identified as inaccurate. This process is time consuming (the time required to visually inspect and edit each segmentation can range between 10 and 45 minutes) and subjective, requiring a trained operator. Moderate interrater reliability has been previously reported, with Cohen's Kappa indices that range between 0.30 (Esteban et al., 2017) and 0.48 (Savalia et al., 2017).

Studies investigating alterations in morphological estimates due to manual editing have shown mixed results, with some studies reporting significant changes (Beelen, Phan, Wouters, Ghesquière, & Vandermosten, 2020; Waters et al., 2019), while others showed no differences (McCarthy et al., 2015). Despite the potential changes in morphological estimates, manual editing does not show important effects on the sensitivity to detect differences between groups using volumetric morphological estimates (Waters et al., 2019), although it may have more impact on the sensitivity to detect differences using cortical morphological estimates such as surface area and thickness (Beelen et al., 2020).

Rather than applying manual QC strategies, one can also use automatic exclusion of cases based on quantitative parameters, such as quality metrics or morphological information that are readily available from FreeSurfer or other software, making the QC process more time-efficient. Among the most commonly used quality metrics are contrast-to-noise ratio (CNR) (Welvaert & Rosseel, 2013) and Euler numbers (EN) (Dale, Fischl, & Sereno, 1999). CNR has been used as an objective measure of image quality over many years, but it correlates weakly with human quality classifications (Yao, Lin, Ong, & Lu, 2005). EN is a measure of reconstructed brain surface complexity calculated by FreeSurfer, and has been found to correlate with movement artifacts (Rosen et al., 2018). Another option for automated QC is the exclusion of cases according to outliers based on global morphological estimates such as mean cortical thickness, total surface area or estimated total intracranial volume, a technique commonly used in neuroimaging studies, e.g. (Boedhoe et al., 2018; Guadalupe et al., 2014; Shinn et al., 2017).

Additionally, several tools for a more extensive QC are currently available. Two promising tools are MRIQC (Esteban et al., 2017) and Qoala-T (Klapwijk et al., 2019). Both tools use machine learning to provide a rating for quality. While MRIQC uses the T1 or T2 images as input to provide several quality metrics and a binary classification of image quality, Qoala-T uses FreeSurfer's segmentation and parcellation output together with FreeSurfer's output quality metrics to provide a score of segmentation quality.

To our knowledge, no previous study compared the performance between several available QC strategies.

The aims of this study were: 1) to determine the effect that manual editing of FreeSurfer's output has on the resulting morphological estimates, and 2) to identify which QC approach is best in terms of reduction of noise and measurement error.

To this end, we manually edited a sample of inaccurate segmentations and analyzed its changes. In addition, we applied 12 different QC strategies to a sample and analyzed their effects. As a test case, we used a regression of morphological brain measures against age, which allowed us to quantify the effects of the different QC strategies through the proportion of unexplained relative to total variance each strategy generated.

The time investment of each QC strategy was taken into consideration, and we provide recommendations for the optimal QC approach in diverse scenarios.

Material and Methods

Study design and participants

We used data from The Maastricht Study, an observational prospective population-based cohort study (Schram et al., 2014) . In brief, the study focuses on the etiology, pathophysiology, complications, and comorbidities of type 2 diabetes and is characterized by an extensive phenotyping approach. All individuals aged between 40 and 75 years living in the southern part of the Netherlands were eligible for participation. Participants were recruited through mass media campaigns, from the municipal registries and the regional Diabetes Patient Registry via mailings. Recruitment was stratified according to known type 2 diabetes status, with an oversampling of individuals with type 2 diabetes. The study has been approved by the institutional medical ethical committee (NL31329.068.10) and the Dutch Ministry of Health, Welfare, and Sports of the Netherlands (permit 131088-105234-PG). All participants gave written informed consent.

The present report uses cross-sectional data from the first 3451 participants who completed the baseline survey (although not necessarily brain MRI) between November 2010 and September 2013. From among the 3451 participants, 200 participants with mild cognitive impairment (MCI) were randomly selected without oversampling for type 2 diabetes. We then added 200 non-MCI participants matched on age, sex, and educational level, in order to introduce some heterogeneity in the sample. MCI diagnosis was based on: Mini-Mental State Examination (Folstein, Robins, & Helzer, 1983) scores below 24 points; more than two cognitive tests not executed; delayed recall and word learning test (Walton, 1958), or Stroop-III (Stroop, 1935) 1.5 SD below the population-mean. Of the selected 400 participants, 260 had brain MRI data available. Data processing and extraction failed in one participant, specifically in the MRIQC tool processing, and was removed. Hence, the current manuscript includes 259 participants. See participant inclusion flowchart in Figure 2.1.
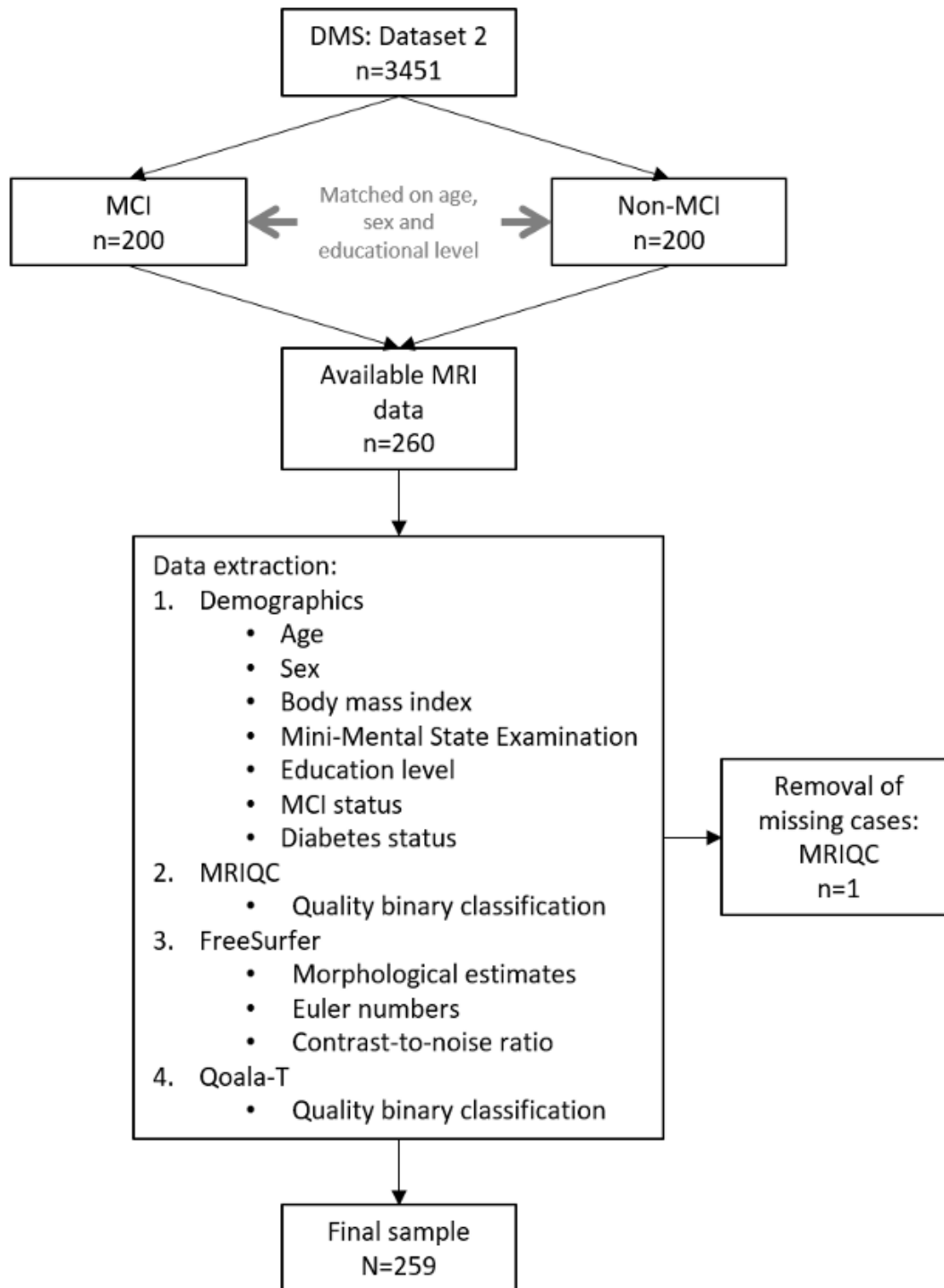
Figure 2.1 | Sample selection and data extraction. *Among the included participants in dataset 2 of The Maastricht Study (DMS), a sample of mild cognitive impairment (MCI) participants and matched non-MCI were selected. Among the selected 400, 260 participants had available brain MRI data. There was one missing case because the MRIQC tool was unable to run in one of the participants. The case with missing data was removed and the present report includes 259 subjects.*

MRI acquisition

Brain images were acquired on a 3T clinical magnetic resonance scanner (MAGNETOM Prismafit, Siemens Healthineers GmbH) located at a dedicated scanning facility (Scannexus, Maastricht, The Netherlands) using a head/neck coil with 64 elements for parallel imaging. The MRI protocol included a three-dimensional (3D) T1-weighted (T1w) magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence (repetition time/inversion time/echo time (TR/TI/TE) 2,300/900/2.98ms, 176 slices, 256 × 240 matrix size, 1.0 mm cubic reconstructed voxel size); and a fluid-attenuated inversion recovery (FLAIR) sequence (TR/TI/TE 5,000/1,800/394 ms, 176 slices, 512 × 512 matrix size, 0.49 × 0.49 × 1.0 mm reconstructed voxel size).

Brain segmentation

Brain segmentation and cortical parcellation was performed on the 259 participants with FreeSurfer v6.0 (Fischl, 2012) using T1w and FLAIR images as input. The optional arguments "-FLAIRpial" and "-3T" were used to optimize segmentation and parcellation quality. In addition, hippocampal subfields (Iglesias et al., 2015) were extracted. FreeSurfer output yielded cortical area (68 ROIs) and cortical thickness estimates (68 ROIs) in accordance with the Desikan-Killiany atlas (Desikan et al., 2006), as well as subcortical volumes (38 ROIs), and hippocampal subfields (24 ROIs). Hence, a total of 198 morphological estimates were obtained per individual brain. With no further manipulation, tabulated data was extracted. This original dataset will be referred from now on as "Non-QC dataset" (see Figure 2.2A) and used as reference for comparison with the QC datasets.

Quality control strategies

Twelve QC strategies were applied to the original sample, generating 12 new QC datasets covering all morphological estimates for cortical thickness, cortical area, subcortical volumes and hippocampal volumes. Each QC strategy resulted in different brains to edit, include and exclude. Hence, all the 12 new QC datasets (and the non-QC dataset) contain 198 morphological estimates, but differ with respect to participant inclusion and which brains underwent manual editing.

These strategies can be divided into three categories: 1) manual QC: Visual inspection of brain segmentations with either exclusion or manual editing of inaccurate cases; 2) automated QC by exclusion of outliers based on: MRIQC, Qoala-T, Morphological, EN or CNR measures; 3) semi-automated QC by visual inspection and manual editing of outliers based on: MRIQC, Qoala-T, Morphological, EN or CNR measures. In the next sections we describe the QC strategies in detail.

Manual quality control: visual inspection with exclusion or editing of segmentations

The first category, manual QC includes two QC strategies: Visual inspection with exclusion (Visual-excl) of inaccurate segmentations, and visual inspection with manual editing (Visual-edit), where the inaccurate segmentations are manually edited. Figure 2.2B shows the process of the manual strategies. The visual inspection was executed by two researchers who independently rated brain segmentations according to their quality, followed by manual editing or exclusion of segmentations identified as inaccurate.

A standard operating procedure (SOP) for visual inspection and manual editing was used. The SOP was derived from FreeSurfer's QC guidelines (AnastasiaYendiki, 2020), and tailored to the specific setting of the DMS. In short, aseg, wm, brain mask and T1 volumes, together with bilateral pial and white surfaces were loaded into FreeView. A researcher with three years of experience in hands-on QC of large MRI cohorts (J.M.) performed visual inspection of the 259 brain segmentations twice with a 6-month period gap, without and with the help of the SOP respectively. A second researcher (M.B.), without prior experience in QC, independently reviewed the same set of segmentations once, after training by rater 1, and following the same SOP. Both researchers inspected the segmentations from posterior to anterior three times: one fast screening for large inaccuracies and two detailed inspections, once for each hemisphere. The quality of the segmentations was scored from 0 to 3, where 0 referred to segmentations with perfect quality, 1 to segmentations with sufficient quality, 2 to segmentations that needed intervention, and 3 to segmentations that should be excluded due to unfixable inaccuracies. Segmentations that needed intervention included those with large and medium fixable errors, and those with frequent (more than 10 slices) small errors. Unfixable inaccuracies were defined as those where the full morphology of the brain was not visible anymore, due to large image artifacts or incidental findings.

Finally, both researchers met to review and discuss each discordant case (n-141) and reached consensus, creating a robust final agreed-upon score called from now on "Accorded rating". The segmentations with accorded ratings of 0 or 1 were accepted, and those rated with a 3 were removed from the dataset.

Subsequently, manual editing was performed on those brains with inaccurate segmentations, scored as 2 by the accorded rating. The editing was performed by changing the brain surfaces where inaccuracies were detected. This process was done through addition or removal of voxels in the white matter mask, removal of voxels in the brain mask, and addition of control points in the brain mask.

The edited subjects subsequently underwent a new segmentation pipeline. This process -visual inspection, manual editing, and production of a new segmentation- was repeated a maximum of two times when necessary, after which, the reconstructed images were visually inspected one last time. These were then accepted as accurate (scores 0 or 1) or rejected as unfixable (score of 3).

Tabulated data from only the segmentations accepted in the first visual inspection (initial scores of 0 or 1) was extracted creating the dataset "Visual-excl dataset". Tabulated data of segmentations accepted after manual editing, together with those accepted in the first visual inspection (and hence not edited) was extracted creating the dataset "Visual-edit dataset".

In order to study changes produced by manual editing in brain estimates, morphological estimates were additionally extracted before and after manual editing for only the edited segmentations (n=39). These 39 subjects were used to investigate the alterations due to manual editing on morphological FreeSurfer estimates.

Automatic quality control: exclusion of cases

Figure 2.2C shows the automatic QC process. MR images were either accepted or excluded based on the assessment by the following tools: MRIQC (Esteban et al., 2017), and Qoala-T ("alternative B") (Klapwijk et al., 2019); and the following metrics: FreeSurfer's global morphological measures, EN (Dale et al., 1999), and CNR (Welvaert & Rosseel, 2013).

Among other features, the tools MRIQC and Qoala-T use supervised machine learning to assign quality scores. We used the default setting in MRIQC, which is trained in a sample with ages ranging between 7 and 64 years old (mean age = 20.23). However, Qoala-T's default setting is trained in a pediatric sample, and in order to increase the accuracy of the outcome, we used the "alternative B" which allowed us to train the algorithm with 10% of our own sample. Both tools provided a binary quality score ("good" or "poor"). Segmentations with a quality score of "poor" were excluded. Tabulated data was then extracted, creating the datasets "Auto-MRIQC dataset" and "Auto-Qoala dataset" respectively.

Despite the fact that many studies define outliers based on standard deviation, it is a measure highly dependent on distribution, as it assumes a normal distribution, and hence not a robust method to detect outliers (Leys, Ley, Klein, Bernard, & Licata, 2013). For this reason, in this study outliers were defined as 1.5 interquartile range (IQR) below the first quartile (Q1), and 1.5 IQR above the third quartile (Q3), following the classical method proposed by Tukey (1977). Hence, the lower inner fence was defined as Q1-1.5*IQR, while the upper inner fence was Q3+1.5*IQR.

The identification of morphological outliers was specific for each type of measure, and based on the next FreeSurfer's global estimates: "left/right hemisphere mean thickness", for estimates of cortical thickness; "left/right hemisphere white surface area" for cortical area; "estimated total intracranial volume" and "mask volume" for subcortical volumes; and "left/right hemisphere whole hippocampus volume" for hippocampal subfields. Outliers were excluded below the lower and above the upper inner fences. Tabulated data was extracted for each type of morphological estimate separately (cortical thickness, cortical area, subcortical volumes and hippocampal volumes), and joined, creating the dataset "Auto-morphological dataset".

Outliers based on EN and CNR metrics were excluded only when below the lower inner fence, because high values in EN and CNR indicate a positive relation with quality. Tabulated data was extracted, creating the datasets "Auto-EN dataset" and "Auto-CNR dataset" respectively.

Semi-automated quality control: automatic detection with visual inspection and editing

Figure 2.2D shows the semi-automated QC process. Based on the same principle as for the automatic QC strategies, potentially inaccurate cases and outliers were identified with MRIQC, Qoala-T, global morphological estimates, EN, and CNR. Rather than being excluded, the potentially inaccurate cases went through visual inspection and manual editing when necessary, in an identical scheme as the one described in section "2.1.4.1 Manual quality control: visual inspection with editing or exclusion of cases".

For each approach, tabulated data was then extracted creating the QC datasets: "Semi-MRIQC dataset", "Semi-Qoala dataset", "Semi-Morphological dataset", "Semi-EN dataset", and "Semi-CNR dataset".
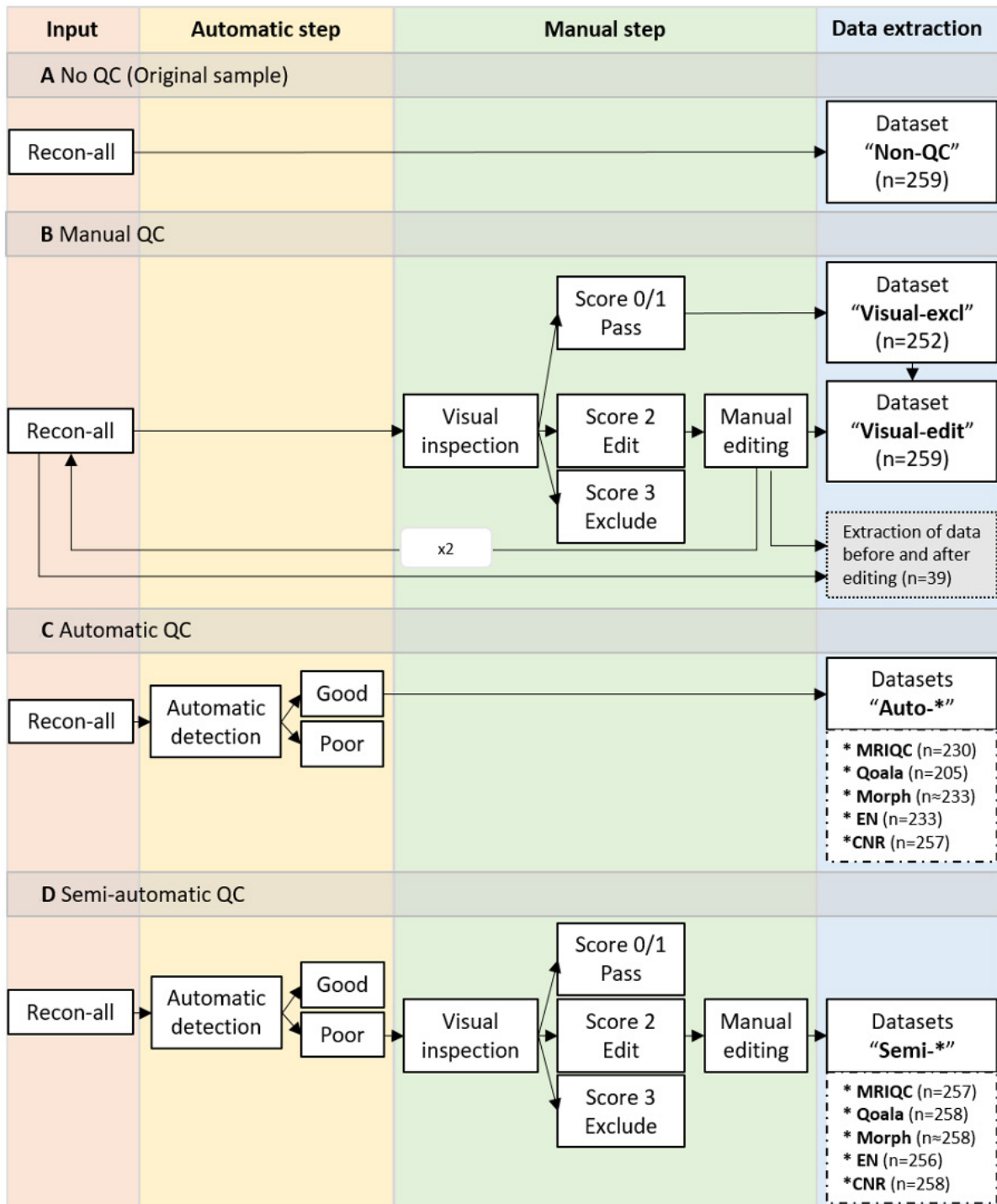
Figure 2.2 | Creation of QC datasets. **A)** *No QC: The extraction of the morphological estimates is done directly from Free-Surfer's output (recon-all) with no quality control (QC).* **B)** *Manual QC: Segmented images undergo visual inspection. Only images with a direct score of "Pass" (with no manual editing) are included in both the visual-excl and visual-edit datasets. When manual editing is necessary, it is performed a maximum of two times. All images classified as "pass" in the last round (after manual editing) are added to the visual-edit dataset. In parallel, data is independently extracted (box in gray) only for those images requiring manual editing (n=39). This is done before and after undergoing manual editing.* **C)** *Automatic QC: Segmentations are classified according to their quality by either an automated software (MRIQC or Qoala-T)*

*or outliers based on several metrics (Morphological (Morph), Euler numbers (EN), or Contrast-to-noise ratio (CNR)). Only those classified as "good" are included in the dataset. **D)** Semi-automatic QC: A "combination" between automatic and Visual-edit QC. Instead of being excluded, images classified as "poor" by the automated steps, undergo visual inspection (and manual editing when necessary). For details on the number of edited and excluded cases, sample size and time cost of each QC strategy please see Supplementary Table S2.1.*

Statistical analysis

Sample characteristics

Wilcoxon signed-rank tests and Chi-squared (χ2) tests, for continuous and categorical variables respectively, were performed to assess significant differences between the MCI and non-MCI groups.

Agreement and overlap of manual ratings

Weighted Cohen's Kappa (К) (Cohen, 1968) and percentage of agreement $\left(\frac{\text{scores with agreement}}{\text{total scores}} \times 100 \right)$ were implemented to assess inter- and intra-rater reliability among the initial visual inspection's ratings.

Manual editing effects

To assess our first aim, the percentage of change on brain morphological estimates $\left(\frac{\text{edited} - \text{unedited}}{\text{unedited}} \times 100 \right)$ after manually editing was extracted for each of the 198 brain morphological estimates separately. Wilcoxon signed-rank test, and effect size (r) (Rosenthal, Cooper, & Hedges, 1994) defined as r=Z/√N, where Z is the Z-score and N is the sample size, were used to test significance of changes before and after manual editing for each paired morphological estimate. False discovery rate (FDR) was used for multiple comparisons correction providing q-values. In addition, we extracted the average within-subject coefficient of variation (CoV) for each of the 198 morphological estimates as: $\text{CoV}_{\text{within-subject}} = \sqrt{\frac{1}{n}\sum_n \frac{s}{m^2}} \cdot \sqrt{\frac{1}{n}\sum_n \frac{s}{m^2}}$ Where n is the sample size, s the within-subject variance as $s = \frac{(x_{\text{edited}} - x_{\text{unedited}})^2}{2}$, m the within-subject mean as $m = \frac{x_{\text{edited}} + x_{\text{unedited}}}{2}$, and x is a specific brain morphological estimate.

Comparison of QC strategies

To assess our second aim, we focused on how QC strategies change the proportion of unexplained variance relative to its total variance. The background concept is based on the measurement error contained in a linear regression $\left(Y_i = \beta_0 + \sum_{j-1}^{p} \beta_j X_{i,j} + \varepsilon_i \right)$ model's stochastic component, the error term. Purely measuring the changes in unexplained variance is insufficient, as any QC strategy that would reduce the total variance of a sample will collaterally reduce the unexplained variance, but potentially also the explained one. Therefore, we here used the inverse proportion of unexplained-to-total variance. The coefficient of determination (R^2) captures this proportion $\left(R^2 = 1 - \left(\frac{\sum_i \varepsilon e_i^2}{\sum_i (y_i - \bar{y})^2} \right) = 1 - \left(\frac{\text{Unexplained variance}}{\text{Total variance}} \right) \right)$. It can be robustly

extracted from a linear regression model, and in practical terms it is not affected by sample size (Hayes, 2017). A small part of the unexplained variance in this proportion is due to noise or measurement error. The QC strategies will target this small part and potentially change it. Therefore, the measuring the small changes in the unexplained variance proportion before and after a QC strategy is applied, allows for the quantification of changes in the quality of a sample.

In order to extract R^2 for each morphological measure in each dataset, we created a test case regression model with one morphological measure as dependent variable, and age and sex as independent variables. We ran this model separately for each of the 198 morphological measures in the non-QC dataset and in each of the 12 newly created datasets, obtaining 198 R^2 for each of QC datasets (i.e. a total of 2574 R^2 values). The reason to use a model with age and sex is to ensure a wide range of R^2 values in each dataset. Age related brain atrophy has been widely studied (Gur et al., 1991; Kakimoto et al., 2016; Murphy, DeCarli, Schapiro, Rapoport, & Horwitz, 1992; Tang, Whitman, Lopez, & Baloh, 2001; Yoshii et al., 1988), and most brain regions are to some extent affected by age. Please note that purpose of the linear regression with age was to obtain a metric related to the unexplained-to-total variance, not to assess what morphological estimate has the strongest relationship with age. Using non-QC as baseline, we then individually subtracted the R^2 values obtained in the non-QC sample from their paired R^2 values obtained in each of the 12 QC datasets, obtaining 198 delta R^2 (ΔR^2) for each of the 12 QC strategies $(\Delta R^2_{xQC_i} = R^2_{x,QC_i} - R^2_{x,NonQC'}$ where x is the specific morphological estimate, and i corresponds to 1 of the 12 QC datasets, not including the baseline: Non-QC).

An increase in ΔR^2_x will then indicate a reduction of unexplained-to-total variance ratio for a specific brain morphological estimate, and hence a beneficial increase of relative explained variance. Differences in ΔR^2_x were assessed qualitatively.

To assess the overlap between QC strategies, we report the number and Dice similarity coefficient (Dsc) (Dice, 1945), of the flagged (detected as inaccurate) cases by each QC metric or tool (i.e. Visual inspection, MRIQC, Qoala-T, Global morphological measures, EN and CNR).

In addition, and to explore the robustness of the results, the distribution of ΔR^2 obtained by the original linear regression (brain ~ age + sex) was compared to the distribution obtained when adding BMI as a covariate (brain ~ age + sex + BMI).

All statistical analysis were performed in R 4.0.2 (2020-06-22) (Team, 2013). Graphs were created through ggplot2 (Wickham, 2016), and brain maps through ggseg (Mowinckel & Vidal-Piñeiro, 2020). Scripts are available at https://github.com/JenniferMosa/QCstrategies.

Results

Sample characteristics

259 participants completed FreeSurfer's recon-all and underwent all QC strategies. Supplementary Table S2.2 summarizes the characteristics of the study sample stratified for MCI and non-MCI participants. The MCI and non-MCI groups were matched for age, sex and educational level. A significantly higher BMI in MCI (p=0.017) was found, but there were no other differences on demographic parameters between participants.

Brain segmentation & creation of datasets

A total of 12 QC strategies were applied to the original sample (non-QC dataset). Each strategy differed in the amount of segmentations that were visually inspected, edited or excluded, as well as the time investment of performing each QC strategy. The sample sizes of the newly created QC datasets ranged from n=205 to n=259. The time investment to perform QC ranged from 15 minutes for the auto-QC strategies to 126 hours (approximately 30 minutes per subject) when applying the Visual-edit QC strategy. Supplementary Table S2.1 summarizes the number of segmentations visually inspected, edited and excluded, the total sample sizes of the new datasets, and the time investment of applying each QC strategy.

Visual inspection

The intra-rater agreement was 54.7%. The inter-rater agreement was 40.7% when only one rater used SOP, and 47.0% when both raters followed the SOP. Table 2.1 summarizes inter/intra-rater reliability through weighted Cohen's Kappa values (K), which takes into consideration the scores as ordered values, and confidence intervals (CI). The inter-rater reliability increases with the use of a SOP, reaching a K=0.25, similar to the intra-rater reliability, with K=0.24.

The accorded rating, agreed-upon by both raters, classified 17.8% of the segmentations as inaccurate, either by requiring manual editing or exclusion. Figure 2.3 shows the distribution of the scores given by each rater, as well as the final accorded rating.

Kappa (Confidence Interval)	Rater 1 TP1, no SOP	Rater 1 TP2, SOP	Rater 2, SOP
Rater 1 TP1, no SOP	1	0.24 (-0.45, 0.94)	0.16 (-0.41, 0.73)
Rater 1 TP2, SOP	-	1	0.25 (-0.09, 0.60)
Rater 2, SOP	-	-	1

Table 2.1 | Inter and intra-rater reliability. Cohen's kappa scores (K) and confidence intervals (CI) are reported for each rating. TP: time point; SOP: Standard operating procedure.
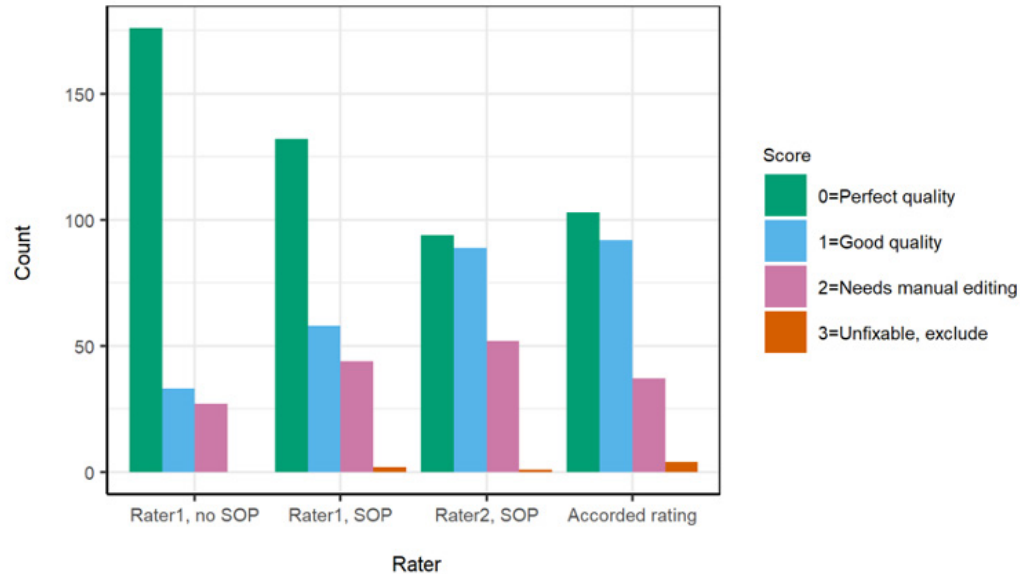
Figure 2.3 | Distribution of given scores in each visual inspection, and final accorded rating. *Abbreviation: SOP: Standard operating procedure.*

Manual editing of brain surfaces: changes in brain estimates

Segmentations of 39 out of 259 brains had an accorded rating of "2" and thus were manually edited. Manual editing resulted in changes in all morphological measures out of the 39 edited brains. The largest difference after editing, with a mean increase of 25% of its volume, was found in bilateral fimbria, followed by differences that ranged from +9 to +12% in bilateral vessel and cerebellar white matter. The largest reduction was found in the fourth ventricle, with a mean volume reduction of 7%. Figure 2.4A shows the average difference in percentage across subjects after editing.

Wilcoxon signed-rank test showed significant differences (q-value < 0.05) in 53 out of the 198 analyzed morphological measures. See Supplementary Figure S2.1 for q-values' brain maps. The effect sizes (r) ranged from 0 to 0.87, with the largest effect sizes (r>0.8) found in bilateral fimbria, brainstem, 4th ventricle, left cerebellum white matter, left lateral ventricle, and bilateral ventral DC. Figure 2.4B shows the effect size distribution for several cortical and subcortical morphological measures. A brain map legend is provided in Supplementary Figure S2.2.

The median CoV for all brain regions was 2.4%, with a standard deviation (SD) of 3.0%. The largest CoV was found within subcortical structures (CoV=4.5%, SD=4.6%), followed by hippocampal subfields (CoV=2.9%, SD=4.2%), cortical thickness (CoV=2.2%, SD=1.0%), and the smaller in cortical areas (CoV=1.7%, SD=1.6%).
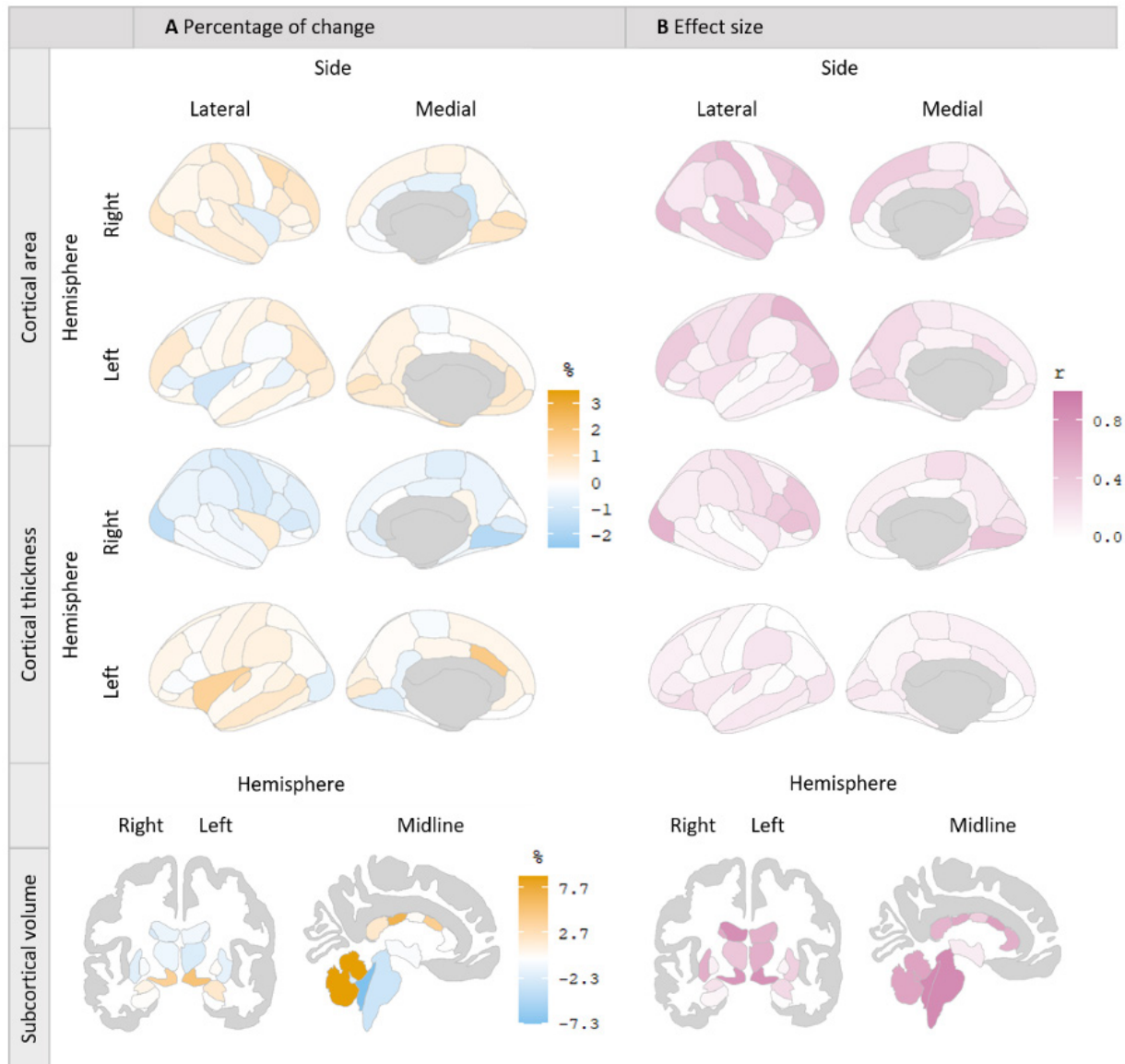
***Figure 2.4 | Brain maps. A)** Average change in percentage of each morphological measure after manual editing. Yellow indicates an increase in volume, area or thickness, while blue indicates a decrease. Color intensity shows the magnitude of the changes. Notice the percentage of change found in subcortical volumes is larger than those found in cortical areas and thickness, as reflected in its legend. **B)** Effect size (r) of these changes. Larger effect sizes are closer to 1 and represented in pink, while lower effect sizes are closer to 0 and represented in white. A brain map legend for cortical and subcortical structures can be found in Supplementary Figure S2.2. Note: The brain maps do not represent all the analyzed morphological measures, for a list of percentage of change, effect sizes, and q-values of all brain regions see Supplementary Table S2.3.*

Quality control Strategies: consequences in a regression analysis

R^2 was extracted for 198 morphological measure over 13 datasets (1 non-QC and 12 QC datasets). The distribution of the R^2 values obtained by each dataset in each brain region can be found in Supplementary Table S2.4.

The overlap of cases detected as inaccurate by the different QC strategies was relatively low, and no cases were simultaneously identified by all the strategies. The use of global morphological metrics for the identification of inaccurate cases (used in Auto-Morpho and Semi-Morpho QC strategies) detected 52 segmentations (30 of them were unique, i.e. only flagged by this specific metric); CNR detected 2 (none of them unique); EN detected 26 (3 unique); MRIQC detected 29 (10 unique); Qoala-T detected 54 (27 unique); and visual inspection detected 46 (21 unique). Supplementary Table S2.5 shows the overlap between the flagged segmentations by each tool and metric. The highest overlap was found between EN and MRIQC (n=11, Dsc=0.40,), followed by EN and Visual inspection (n=10, Dsc=0.36).

Across all brain morphological estimates, the exclusion of cases based on visual inspection (Visual-excl) yields the largest decrease of unexplained variance relative to total variance, as shown by the largest increase in R^2 (ΔR^2 =.011), followed by the exclusion of cases based on EN (Auto-EN) (ΔR^2 =.005). Auto-Morphological yields the largest decrease in R^2 (ΔR^2 =-.021). Figure 2.5.A shows the mean ΔR^2 and confidence intervals (CI) across all morphological estimates for each QC strategy.

Figure 2.5.B shows the specific mean ΔR^2 and CI per type of measure obtained by each QC strategy. The QC strategies that resulted in the highest increase in explained-to-total variance ratio as measured by larger positive ΔR^2 for cortical thickness are: Visual-edit (mean ΔR^2=.009) and visual-excl (ΔR^2 =.009); for cortical area Auto-Qoala (mean ΔR^2 =.012), Visual-excl (mean ΔR^2 =.009) and Auto-EN (mean ΔR^2 =.008); for subcortical volumes Visual-excl (mean ΔR^2 =.018), auto-MRIQC (mean ΔR^2 =.012), Auto-EN (mean ΔR^2 =.010), and Auto-Qoala (mean ΔR^2 =.006); and for hippocampal subfields Auto-EN (mean ΔR^2 =.013), Visual-excl (mean ΔR^2 =.013), Auto-MRIQC (mean $\Delta R2$=.007) and Auto-Qoala (mean ΔR^2 =.006). The exclusion of cases based on morphological estimates (Auto-Morphological) produced a relatively large reduction in R^2 for any type of measure (mean ΔR^2 = -.034 for area, mean ΔR^2 = -.022 for hippocampal subfields, mean ΔR^2 = -.017 for subcortical volumes and mean ΔR^2 = -.011 for cortical thickness), as well as Auto-Qoala, which reduced R^2 in cortical thickness by -.021 relative to the non-QC strategy.

The distribution pattern of ΔR^2 obtained in each QC dataset is preserved when adding an additional covariate (BMI) to the regression model (See Supplementary Figure S2.3).
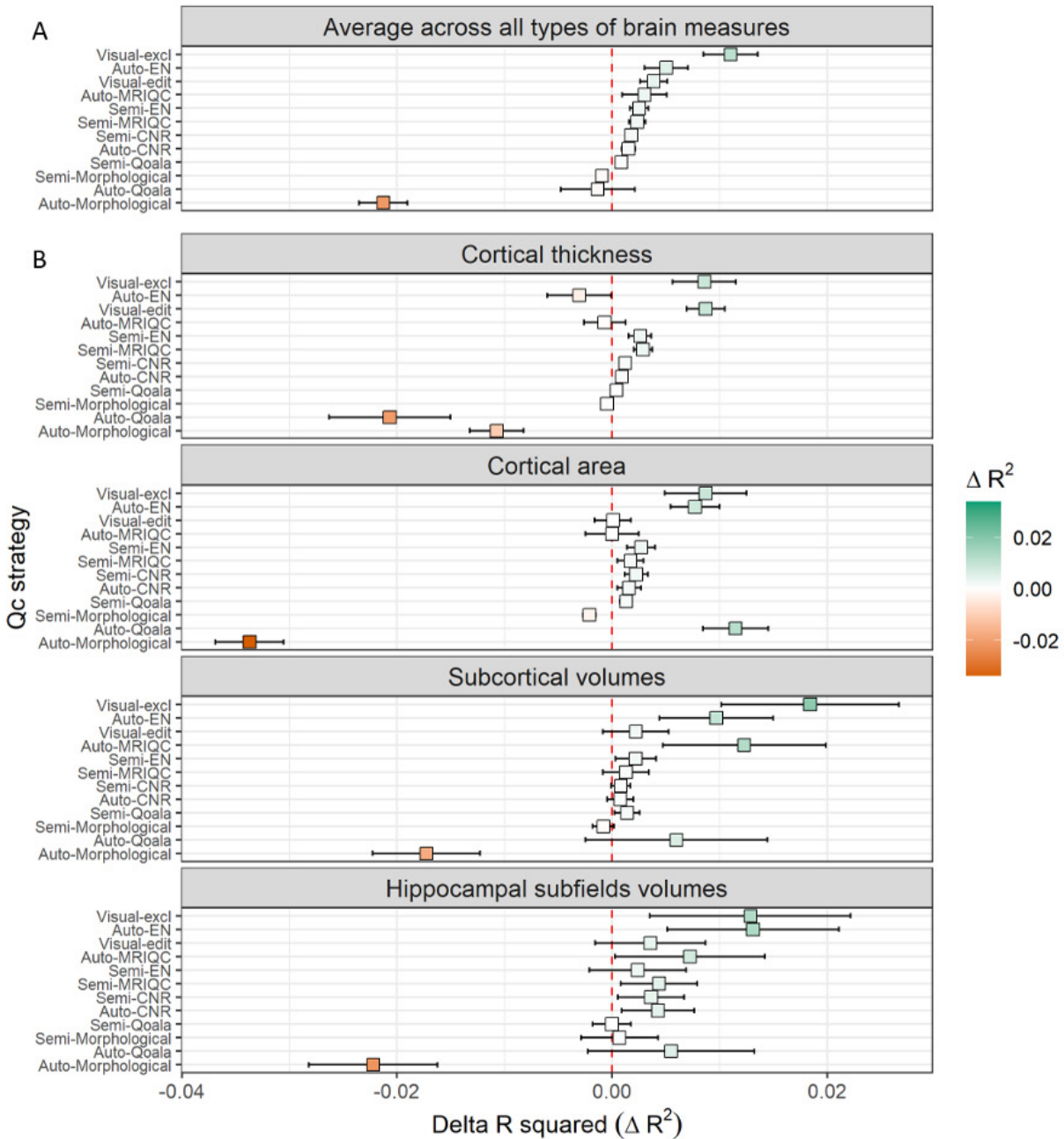
Figure 2.5 | QC strategies performance. *Mean ΔR² obtained by each QC strategy is indicated with a square. Whiskers show the 95% confidence interval for the mean. Green and orange indicate an increase and decrease in R² respectively. Red dotted line shows the mean ΔR² of the non-QC strategy, i.e. zero. A) Average ΔR² across all brain regions for each QC strategy. B) Average ΔR² across all brain regions of a specific type (Cortical thickness, area, subcortical and hippocampal subfields volumes). Abbreviations: QC: Quality control; EN: Euler numbers; CNR: Contrast-to-noise ratio; Qoala: Qoala-T.*

Discussion

In this study we investigated the influence of manual editing of brain surfaces on resulting morphological FreeSurfer estimates. We also compared the performance of different QC strategies by assessing how these alter the relative proportion of unexplained variance in a neuroimaging sample.

Manual editing brain surfaces

Manual editing of brain segmentations produces significant changes in several brain morphological estimates across thickness, surface area and volumetric measures, with large effect sizes especially found in subcortical structures. Similar studies showed mixed results: Waters et al. (2019) showed significant differences in cortical but not subcortical volume following editing. Beelen et al. (2020) found a significant increase in cortical area, and a decrease in cortical thickness after manual editing. Conversely, McCarthy et al. (2015) found no significant differences in cortical area, thickness, and subcortical volumes. The discrepancy in results could be explained by the use of different editing techniques: While Waters et al. (2019) only edited voxels in the brain mask, Beelen et al. (2020) also edited the white matter mask, and used control points; McCarthy et al. (2015) exclusively used control points and white matter modification; and we edited the brain mask, white matter mask and used control points.

The within-subject CoV that we obtained after manual editing was similar to what was found in repeatability studies for subcortical structures (Maclaren, Han, Vos, Fischbein, & Bammer, 2014; Velasco-Annis, Akhondi-Asl, Stamm, & Warfield, 2018). Further, similar CoV were found in FreeSurfer's reproducibility studies between operating systems in measures of volume and thickness (Gronenschild et al., 2012).

Our results indicate that manual editing of brain surfaces through removal and addition of voxels and control points, produces changes in all types of morphological estimates. It is important to realize that, in practice, finding a change in the morphological measures after manual intervention (or the usage of a new operating system to run FreeSurfer) is not sufficient to conclude the segmentation quality has decreased or increased, i.e. a change in a morphological estimate does not indicate a change in the quality of the data. Visual inspection of the newly generated segmentations is essential to ensure the manual intervention produced the desired effect.

Comparison between QC strategies

We created a regression model and extracted the ΔR^2 for each morphological estimate when comparing each QC strategy to non-QC. To our knowledge, this is the first study using the proportion of unexplained-to-total variance as a measure of quality.

Two previous studies have assessed the importance of manual QC by testing whether it increases the sensitivity to detect differences between clinical groups. Both studies used paired t-tests to compare the effect sizes obtained with and without manual editing, with no significant results (Beelen et al., 2020). With a different methodology, Waters et al. (2019) assessed if the correlation coefficients for brain-behavior relationships differed between edited and unedited segmentations, finding non-significant differences. Relating the quality of the data to the capacity to find significant results can be misleading, as non-randomized noise, for example caused by biases in the data (Blumenthal et al., 2002; Madan, 2018; Pardoe et al., 2016; Savalia et al., 2017; Wenger et al., 2014) can lead to more significant differences between groups. Similarly, some studies have used the variance of brain volumes before and after manual editing (McCarthy et al., 2015) or when applying a specific QC strategy (L. L. Backhausen et al., 2016) to test its adequacy. While the reduction of the total variance of a sample implies a potential reduction on unexplained variance, it can also entail a potential reduction of explained variance driven by actual morpho-physiological information. Finally, other studies assess the viability of a QC strategy by comparing it to manual ratings (Klapwijk et al., 2019; Rosen et al., 2018; Yao et al., 2005), which assumes that manual QC improves the quality of a sample. Using the proportion of unexplained-to-total variance, allowed us to capture the measurement error caused by both noise and biases for each QC strategy, without assuming a specific gold standard.

Manual quality control strategies remain the gold standard

Visual inspection with exclusion (Visual-excl) and editing (Visual-edit) of inaccurate cases, produced the largest increases in R^2 overall, and were beneficial for all types of brain measures.

Despite the good results as indicated by the increase in R^2, the low agreement between raters obtained during the visual inspection highlights the subjectivity component of the manual strategies. Enough training and experience, and the use of formal guidelines, are important to lessen the subjectivity of this process. Ultimately, the reassessment and discussion of the discordant cases between raters, may help overcome an initial low inter-rater reliability score.

Well-established formal guidelines for visual inspection of both brain MRI and FreeSurfer's output are publicly available. Notable examples are the workflow for QC of structural brain images proposed by (Lea L Backhausen et al., 2016) and the ENIGMA QC protocols (Toro, 2010) with focus on cortical parcellation. Further, recently developed tools for visual inspection of brain segmentations (Keshavan et al., 2018) and cortical parcellations (Raamana et al., 2020) may provide a solid support to increase reproducibility and reduce time cost, optimizing the visual inspection process.

Therefore, where the sample size, time requirements, and resources allow it, and if efforts are being made to ensure a reliable and reproducible score and manual intervention, manual strategies are highly recommended.

Auto-EN provides, on average, the greatest reduction in noise among automated strategies

Auto-EN is a pragmatic strategy that, on average, produces the second best reduction of unexplained variance, consistently good for three out of four types of brain measures, namely subcortical volumes, hippocampal subfields and cortical area. EN is a measure of reconstructed surface complexity and has been found to highly correlate with visual inspection scores and image artifacts in several samples (Rosen et al., 2018). It is perhaps counterintuitive that this QC strategy performs poorly for measures of cortical thickness. However, EN relates to the frequency, and not the size, of bridges and holes in the brain segmentation. Cortical thickness estimates may be more influenced by large errors (possibly more related to skull stripping) than from frequent small ones, which may be more related to underlying image quality. Hence, auto-EN is an effective and time-efficient QC strategy appropriate for large sample sizes.

The optimal QC strategy depends on the type of morphological estimate of interest and sample size

Our results show that cortical thickness benefits the most from the manual strategies (Visual-excl and Visual-edit); cortical area benefits most from Auto-Qoala and Visual-excl; and subcortical and hippocampal subfields volumes from Visual-excl, Auto-EN, and Auto-MRIQC. Interestingly, Visual-edit performs worse than Visual-excl in all types of morphological estimates except cortical thickness.

The different effects of the QC strategies on the type of morphological estimate can be explained by the particular mechanisms of each QC strategy. Manual strategies visually inspect the upper and lower boundaries of the cortical gray matter (pial and white matter surfaces respectively), which are tightly related to the cortical gray matter thickness. Manual editing changes the distance between these boundaries, directly changing the cortical thickness estimates. However, manual editing does not change the boundaries between specific ROIs, and hence morphological estimates of surface area are not substantially affected. Distinct from cortical parcellation, subcortical segmentation is defined based on image intensities and probabilistic information of ROIs positions (Fischl et al., 2002), and thus, correcting errors in surface boundaries and adding control points to correct normalization errors changes the subcortical estimations only indirectly. The QC strategies based on image quality metrics -such as CNR and especially MRIQC- may provide a time efficient indication of segmentation quality based on the fact that good image quality -for example high contrast, with no movement or artifacts- might facilitate a good segmentation performance.

The better performance of excluding (Visual-excl), rather than editing (Visual-edit) segmentations, may indicate that while visual inspection accurately targets segmentations with inaccuracies affecting all types of measures (thickness, area and volumes), manual editing these segmentations has a limited effect in the areas and volumes, preserving part of the unexplained variance of these cases. Nonetheless, small and medium samples

may benefit from the use of manual intervention. Editing, instead of excluding inaccurate segmentations, would lead to less missing cases, which is very valuable when the sample size is small. In addition, worse image quality is more often found in certain population or clinical groups (Blumenthal et al., 2002; Madan, 2018; Pardoe et al., 2016; Savalia et al., 2017; Wenger et al., 2014), which can lead to poorer segmentations. Manual editing may lessen the otherwise excluded cases among this population groups, preventing an imbalance in the dataset, and the potentially associated bias.

QC strategies based on global morphological estimates are not a suitable QC solution

Notably, for all types of measures, the ΔR^2 decreased by the commonly used strategy of excluding or visually inspecting and editing subjects based on global morphological estimates (Auto- and Semi-Morphological QC strategies). Excluding outliers based on morphological estimates naturally reduces the total variance of a sample, but our results indicate that a large part of this reduced variance is not unexplained variance but potentially relevant morpho-physiological information. In addition, a previous study found that approximately 40% of the segmentation errors are not identified by the use of morphological outliers (Waters et al., 2019). Taken together, this strongly indicates that the exclusion of subjects based on global morphological estimates is not a suitable QC strategy.

Limitations

The design of this study does not allow drawing firm conclusions beyond models using age as an independent variable. However, results are robust when adding a new covariate, BMI. And given a large proportion of published studies include age as a covariate or independent variable, information about the optimal QC procedure for such scenario is relevant for the neuroimaging community.

In addition, many QC strategies seem to overlap in performance with each other, and due to the violation of the assumption of independence, no regression or correlation tests can be performed in the resulting R^2. Therefore, the changes produced by these strategies can only be compared qualitatively, and the significance of this differences cannot be assessed.

Further, the data acquisition and segmentation pipeline from DMS was carefully designed to provide the best data quality. This led to a sample with high quality of segmentations (82% rated as accurate by visual inspection), which might not be representative of other cohorts. Different (more severely affected) clinical samples, studies with different population types, hardware or reconstruction parameters, could lead to a different pattern of segmentation inaccuracies. Further research needs to be performed to investigate these scenarios.

Finally, this study does not cover all the possible QC strategies that are available for neuroimaging. A selection was made from commonly applied, and user friendly, options. There is room for improvement in finding a time efficient and reliable option, and future studies should explore different strategies.

Conclusion

Manual editing of brain surfaces significantly alters FreeSurfer's brain morphological estimates.

The selection of a QC strategy should be determined by the type of morphological measures of interest in a study, while taking in consideration the available resources: Manual strategies (Visual-edit and Visual-excl) provide the largest decrease in unexplained variance across brain measures. Therefore, the visual inspection of FreeSurfer's output, with or without manual editing, is highly encouraged whenever possible. However, when manual options are not feasible, the exclusion of outliers based on EN (Auto-EN) is a time efficient alternative, especially for studies focused on subcortical measures, hippocampal subfields or cortical area. Auto-EN produces on average an increase in explained-to-total variance similar to the one achieved by manual strategies. Yet, when cortical thickness measures are of particular relevance for a study, and manual strategies (Visual-excl or Visual-edit) are not feasible, the visual inspection and editing of outliers based on EN (Semi-EN) provides an alternative that is, despite lower in average, consistently beneficial for all types of brain measures. It is recommended, particularly for those studies using fully automatic strategies, to report the results both, with, and without the excluded cases.

Finally, we discourage the exclusion of outliers based on global morphological estimates (Auto-morpho) as it reduces a larger proportion of explained variance.

References

AnastasiaYendiki. (2020, 27-05-2020). FreeSurferWiki. Retrieved from https://surfer.nmr.mgh.harvard.edu/fswiki

Backhausen, L. L., Herting, M. M., Buse, J., Roessner, V., Smolka, M. N., & Vetter, N. C. (2016). Quality Control of Structural MRI Images Applied Using FreeSurfer-A Hands-On Workflow to Rate Motion Artifacts. *Front Neurosci, 10*, 558. doi:10.3389/fnins.2016.00558

Backhausen, L. L., Herting, M. M., Buse, J., Roessner, V., Smolka, M. N., & Vetter, N. C. (2016). Quality control of structural MRI images applied using FreeSurfer—a hands-on workflow to rate motion artifacts. *Frontiers in Neuroscience, 10*, 558.

Beelen, C., Phan, T. V., Wouters, J., Ghesquière, P., & Vandermosten, M. (2020). Investigating the Added Value of FreeSurfer's Manual Editing Procedure for the Study of the Reading Network in a Pediatric Population. *Frontiers in Human Neuroscience, 14*, 143.

Blumenthal, J. D., Zijdenbos, A., Molloy, E., & Giedd, J. N. (2002). Motion artifact in magnetic resonance imaging: implications for automated analysis. *NeuroImage, 16*(1), 89-92.

Boedhoe, P. S., Schmaal, L., Abe, Y., Alonso, P., Ameis, S. H., Anticevic, A., . . . Beucke, J. C. (2018). Cortical abnormalities associated with pediatric and adult obsessive-compulsive disorder: findings from the ENIGMA Obsessive-Compulsive Disorder Working Group. *American journal of Psychiatry, 175*(5), 453-462.

Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychology. *Bulletin, 70, 213â, 220.*

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage, 9*(2), 179-194.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Hyman, B. T. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage, 31*(3), 968-980.

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PloS one, 12*(9).

Fischl, B. (2012). FreeSurfer. *NeuroImage, 62*(2), 774-781.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Klaveness, S. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron, 33*(3), 341-355.

Folstein, M. F., Robins, L. N., & Helzer, J. E. (1983). The mini-mental state examination. *Archives of general psychiatry, 40*(7), 812-812.

Gronenschild, E. H., Habets, P., Jacobs, H. I., Mengelers, R., Rozendaal, N., Van Os, J., & Marcelis, M. (2012). The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. *PloS one, 7*(6), e38234.

Guadalupe, T., Willems, R. M., Zwiers, M. P., Arias Vasquez, A., Hoogman, M., Hagoort, P., . . . Fisher, S. E. (2014). Differences in cerebral cortical anatomy of left-and right-handers. *Frontiers in psychology, 5*, 261.

Gur, R. C., Mozley, P. D., Resnick, S. M., Gottlieb, G. L., Kohn, M., Zimmerman, R., . . . Berretta, D. (1991). Gender differences in age effect on brain atrophy measured by magnetic resonance imaging. *Proceedings of the National Academy of Sciences, 88*(7), 2845-2849.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*: Guilford publications.

Hutcheon, J. A., Chiolero, A., & Hanley, J. A. (2010). Random measurement error and regression dilution bias. *bmj, 340*, c2289.

Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., . . . Wald, L. L. (2015). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. *NeuroImage, 115*, 117-137.

Kakimoto, A., Ito, S., Okada, H., Nishizawa, S., Minoshima, S., & Ouchi, Y. (2016). Age-related sex-specific changes in brain metabolism and morphology. *Journal of Nuclear Medicine, 57*(2), 221-225.

Keshavan, A., Datta, E., I, M. M., Madan, C. R., Jordan, K., & Henry, R. G. (2018). Mindcontrol: A web application for brain segmentation quality control. *NeuroImage, 170*, 365-372. doi:10.1016/j.neuroimage.2017.03.055

Klapwijk, E. T., Van De Kamp, F., Van Der Meulen, M., Peters, S., & Wierenga, L. M. (2019). Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data. *Neuroimage, 189*, 116-129.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764-766.

Maclaren, J., Han, Z., Vos, S. B., Fischbein, N., & Bammer, R. (2014). Reliability of brain volume measurements: a test-retest dataset. *Scientific data, 1*(1), 1-9.

Madan, C. R. (2018). Age differences in head motion and estimates of cortical morphology. *PeerJ, 6*, e5176.

McCarthy, C. S., Ramprashad, A., Thompson, C., Botti, J.-A., Coman, I. L., & Kates, W. R. (2015). A comparison of FreeSurfer-generated data with and without manual intervention. *Frontiers in neuroscience, 9*, 379.

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., . . . Andersson, J. L. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience, 19*(11), 1523-1536.

Mowinckel, A. M., & Vidal-Piñeiro, D. (2020). Visualization of brain statistics with R packages ggseg and ggseg3d. *Advances in Methods and Practices in Psychological Science, 3*(4), 466-483.

Murphy, D. G., DeCarli, C., Schapiro, M. B., Rapoport, S. I., & Horwitz, B. (1992). Age-related differences in volumes of subcortical nuclei, brain matter, and cerebrospinal fluid in healthy men as measured with magnetic resonance imaging. *Archives of Neurology, 49*(8), 839-845.

Pardoe, H. R., Hiess, R. K., & Kuzniecky, R. (2016). Motion and morphometry in clinical and nonclinical populations. *NeuroImage, 135*, 177-185.

Phillips, G. W., & Jiang, T. (2016). Measurement error and equating error in power analysis. *Practical Assessment, Research, and Evaluation, 21*(1), 9.

Raamana, P. R., Theyers, A., Selliah, T., Bhati, P., Arnott, S. R., Hassel, S., . . . Lam, R. W. (2020). Visual QC protocol for freesurfer cortical parcellations from anatomical MRI. *bioRxiv*.

Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J., & Fischl, B. (2015). Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *NeuroImage, 107*, 107-115.

Rosen, A. F., Roalf, D. R., Ruparel, K., Blake, J., Seelaus, K., Villa, L. P., . . . Elliott, M. A. (2018). Quantitative assessment of structural image quality. *Neuroimage, 169*, 407-418.

Rosenthal, R., Cooper, H., & Hedges, L. (1994). Parametric measures of effect size. *The handbook of research synthesis, 621*(2), 231-244.

Savalia, N. K., Agres, P. F., Chan, M. Y., Feczko, E. J., Kennedy, K. M., & Wig, G. S. (2017). Motion-related artifacts in structural brain images revealed with independent estimates of in-scanner head motion. *Human Brain Mapping, 38*(1), 472-492.

Schram, M. T., Sep, S. J., van der Kallen, C. J., Dagnelie, P. C., Koster, A., Schaper, N., . . . Stehouwer, C. D. (2014). The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *European journal of epidemiology, 29*(6), 439-451.

Shinn, A. K., Roh, Y. S., Ravichandran, C. T., Baker, J. T., Öngür, D., & Cohen, B. M. (2017). Aberrant cerebellar connectivity in bipolar disorder with psychosis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2*(5), 438-448.

Stroop, J. (1935). Stroop color word test. *J Exp Physiol, 18*, 643-662.

Tang, Y., Whitman, G. T., Lopez, I., & Baloh, R. W. (2001). Brain volume changes on longitudinal magnetic resonance imaging in normal older people. *Journal of Neuroimaging, 11*(4), 393-400.

Team, R. C. (2013). R: A language and environment for statistical computing. In: Vienna, Austria.

Toro, R. (2010, 29-June). Visual quality control of FreeSurfer results (ENIGMA Consortium). Retrieved from http://enigma.ini.usc.edu/protocols/imaging-protocols/protocol-for-quality-control-and-summary-statistics/#FS

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2): Reading, Mass.

Velasco-Annis, C., Akhondi-Asl, A., Stamm, A., & Warfield, S. K. (2018). Reproducibility of brain MRI segmentation algorithms: empirical comparison of local MAP PSTAPLE, FreeSurfer, and FSL-FIRST. *Journal of Neuroimaging, 28*(2), 162-172.

Walton, D. (1958). The diagnostic and predictive accuracy of the modified word learning test in psychiatric patients over 65. *Journal of Mental Science, 104*(437), 1119-1122.

Waters, A. B., Mace, R. A., Sawyer, K. S., & Gansler, D. A. (2019). Identifying errors in Freesurfer automated skull stripping and the incremental utility of manual intervention. *Brain imaging and behavior, 13*(5), 1281-1291.

Welvaert, M., & Rosseel, Y. (2013). On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PloS one, 8*(11), e77089.

Wenger, E., Mårtensson, J., Noack, H., Bodammer, N. C., Kühn, S., Schaefer, S., . . . Lindenberger, U. (2014). Comparing manual and automatic segmentation of hippocampal volumes: reliability and validity issues in younger and older brains. *Human Brain Mapping, 35*(8), 4236-4248.

Wickham, H. (2016). The ggplot Package License GPL.

Yao, S., Lin, W., Ong, E., & Lu, Z. (2005). *Contrast signal-to-noise ratio for image quality assessment.* Paper presented at the IEEE International Conference on Image Processing 2005.

Yoshii, F., Barker, W. W., Chang, J. Y., Loewenstein, D., Apicella, A., Smith, D., . . . Duara, R. (1988). Sensitivity of cerebral glucose metabolism to age, gender, brain volume, brain atrophy, and cerebrovascular risk factors. *Journal of Cerebral Blood Flow & Metabolism, 8*(5), 654-661.

Chapter 2: Quality control strategies for brain MRI segmentation and parcellation: practical approaches and recommendations, insights from The Maastricht Study

Supplementary Table S2.1

Category	QC strategy	Visual inspection (n)	Manual editing (n)	Exclusion (n)	Sample size (n)	Cumulative time investment (hours)
A. No QC	Non-QC	0	0	0	259	0.0
B. Manual QC	Visual-edit	259	39	7	252	126.0
	Visual-excl	259	0	46	213	107
C. Automatic QC	Auto-MRIQC	0	0	29	230	0.1
	Auto-Qoala	0	0	54	205	0.1
	Auto-morphological:	0	0			0.1
	Cortical thickness	0	0	15	244	
	Cortical area	0	0	24	235	
	Subcortical volumes	0	0	23	236	
	Hippocampal subfields	0	0	38	221	
	Auto-EN	0	0	26	233	0.1
	Auto-CNR	0	0	2	257	0.1
D. Semi-automatic QC	Semi-MRIQC	29	8	2	257	16.2
	Semi-Qoala	54	10	1	258	27.5
	Semi-morphological:	67	7			31.3
	Cortical thickness	49	2	0	259	
	Cortical area	38	3	1	258	
	Subcortical volumes	31	3	1	258	
	Hippocampal subfields	58	3	2	257	
	Semi-EN	26	10	3	256	16.0
	Semi-CNR	2	0	1	258	1.0

*Supplementary Table S2.1 | **Number of segmentations visually inspected, manually edited or excluded for each QC strategy, sample sizes and time investment are reported.** The Auto and Semi-morphological QC strategies (see morphological estimate's types in italic) differ in sample size, as the including, excluding, and editing criteria is based in different global estimates (see section "Automatic quality control: exclusion of cases" for details). The time investment includes only the human time: time for visual inspection (with two independent raters), for discussion to reach the accorded rating, and the time investment of manual editing. The time investment does not include the preparation time (install the software, or prepare a pipeline, for example) nor the software running time. Abbreviations: QC: Quality control; EN: Euler numbers; CNR: Contrast-to-noise ratio.*

Supplementary Table S2.2

	Non-MCI	MCI	Wilcox/χ2 p-value
N (count)	127	132	
Age (mean (SD))	59.45 (8.80)	59.20 (8.61)	0.768
BMI (mean (SD))	25.50 (3.69)	26.47 (4.14)	0.013
MMSE (mean (SD))	28.86 (1.31)	28.46 (1.66)	0.074
Sex (% females)	45.5	48.5	0.428
Level of education (n low/medium/high)	46/34/47	49/34/49	0.980
T2DM (% yes)	5.5	10.6	0.113

*Supplementary Table S2.2 | **Characteristics of the study sample (n=259) stratified by MCI and non-MCI.** Wilcoxon Signed Rank Test and Chi-Square test were used to assess differences between groups. Abbreviations: MCI: Mild cognitive impairment; SD: standard deviation; BMI: Body mass index; MMSE: Mini Mental Score Examination; T2DM: Type 2 diabetes mellitus.*

Supplementary Figure S2.1
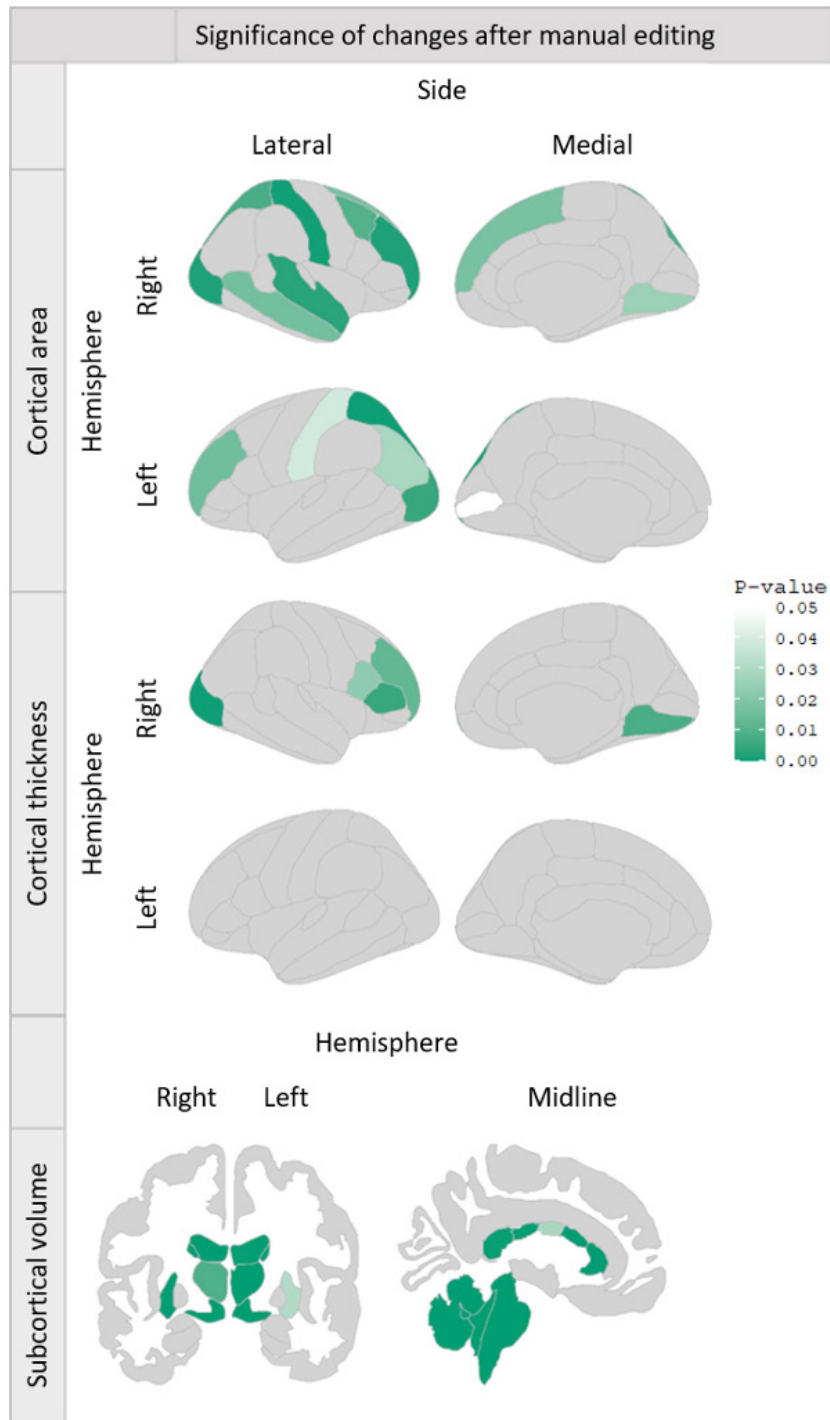
Supplementary Figure S2.1 | Brain maps show, in color, the morphological estimates that significantly changed after manual editing. False discovery rate for multiple comparisons is applied.

Supplementary Figure S2.2

bankssts	inferior parietal	middle temporal	posterior cingulate	supramarginal
caudal anterior cingulate	inferior temporal	paracentral	precentral	temporal pole
caudal middle frontal	insula	parahippocampal	precuneus	transverse temporal
corpus callosum	isthmus cingulate	pars opercularis	rostral anterior cingulate	NA
cuneus	lateral occipital	pars orbitalis	rostral middle frontal	
entorhinal	lateral orbitofrontal	pars triangularis	superior frontal	
frontal pole	lingual	pericalcarine	superior parietal	
fusiform	medial orbitofrontal	postcentral	superior temporal	

region

3rd ventricle	caudate	CC mid posterior	hippocampus	thalamus proper
4th ventricle	CC anterior	CC posterior	lateral ventricle	ventral DC
amygdala	CC central	cerebellum cortex	pallidum	NA
brain stem	CC mid anterior	cerebellum white matter	putamen	

region

Supplementary Figure S2.2 | Color coded brain maps legend with all the represented regions.

Supplementary Table S2.3

ROI	Area			Thickness		
	Percentage of change (%)	q-value	Effect size (r)	Percentage of change (%)	q-value	Effect size (r)
LH_bankssts	0,394	0,879	0,016	-0,488	0,530	0,102
LH_caudalanteriorcingulate	1,854	0,399	0,136	0,655	0,157	0,230
LH_caudalmiddlefrontal	0,097	0,796	0,042	-0,207	0,176	0,218
LH_cuneus	0,346	0,517	0,104	0,333	0,361	0,148
LH_entorhinal	0,442	0,640	0,076	1,289	0,498	0,110
LH_frontalpole	-0,152	0,298	0,168	-0,253	0,856	0,031
LH_fusiform	-0,055	0,704	0,063	0,205	0,586	0,088
LH_inferiorparietal	0,062	0,911	0,019	0,795	0,029	0,350
LH_inferiortemporal	0,394	0,430	0,127	-0,072	0,615	0,082
LH_insula	1,560	0,312	0,163	-1,075	0,157	0,228
LH_isthmuscingulate	-0,437	0,406	0,134	0,491	0,586	0,088
LH_lateraloccipital	-0,649	0,233	0,192	0,624	0,005	0,451
LH_lateralorbitofrontal	0,436	0,120	0,250	0,285	0,209	0,202
LH_lingual	-0,765	0,596	0,086	0,630	0,083	0,273
LH_medialorbitofrontal	-0,004	0,967	0,008	0,676	0,387	0,140
LH_middletemporal	0,808	0,295	0,169	0,434	0,507	0,107
LH_paracentral	-0,200	0,562	0,095	-0,162	0,325	0,159
LH_parahippocampal	-0,009	0,660	0,072	0,261	0,701	0,060
LH_parsopercularis	-0,038	0,879	0,022	0,178	0,533	0,099
LH_parsorbitalis	0,199	0,332	0,156	0,024	0,994	0,011
LH_parstriangularis	-0,121	0,477	0,115	-0,526	0,238	0,190
LH_pericalcarine	0,685	0,255	0,183	0,874	0,050	0,326
LH_postcentral	0,411	0,972	0,007	0,356	0,039	0,332
LH_posteriorcingulate	0,283	0,643	0,077	0,000	0,556	0,099
LH_precentral	0,264	0,701	0,063	0,173	0,214	0,200
LH_precuneus	0,089	0,738	0,055	0,413	0,087	0,272
LH_rostralanteriorcingulate	0,356	0,972	0,007	0,794	0,759	0,050
LH_rostralmiddlefrontal	0,321	0,498	0,110	0,727	0,016	0,388
LH_superiorfrontal	0,310	0,494	0,111	0,098	0,490	0,112
LH_superiorparietal	0,153	0,989	0,003	0,582	0,001	0,544
LH_superiortemporal	0,182	0,754	0,051	0,102	0,823	0,037
LH_supramarginal	0,458	0,230	0,193	-0,155	0,586	0,088
LH_temporalpole	2,111	0,066	0,295	-0,938	0,519	0,103
LH_transversetemporal	1,116	0,145	0,235	-0,015	0,670	0,069
RH_bankssts	-0,213	0,851	0,031	0,041	0,917	0,018
RH_caudalanteriorcingulate	-0,002	0,464	0,118	-0,361	0,507	0,107
RH_caudalmiddlefrontal	-0,460	0,459	0,120	1,133	0,010	0,411
RH_cuneus	-0,042	0,353	0,153	0,074	0,691	0,065
RH_entorhinal	0,523	0,209	0,203	-1,068	0,277	0,175

	Area			Thickness		
ROI	**Percentage of change (%)**	**q-value**	**Effect size (r)**	**Percentage of change (%)**	**q-value**	**Effect size (r)**
RH_frontalpole	-1,598	0,002	0,487	0,539	0,950	0,011
RH_fusiform	-0,343	0,928	0,016	-0,044	0,839	0,035
RH_inferiorparietal	-0,470	0,430	0,127	0,259	0,283	0,169
RH_inferiortemporal	-0,181	0,691	0,065	-0,149	0,759	0,050
RH_insula	0,680	0,222	0,197	-0,720	0,161	0,226
RH_isthmuscingulate	0,286	0,630	0,078	-1,111	0,051	0,314
RH_lateraloccipital	-1,449	0,001	0,549	0,844	0,002	0,508
RH_lateralorbitofrontal	-0,172	0,764	0,049	0,349	0,895	0,022
RH_lingual	-1,699	0,008	0,423	0,835	0,026	0,356
RH_medialorbitofrontal	-0,225	0,954	0,008	-0,068	0,917	0,018
RH_middletemporal	-0,160	0,607	0,087	0,652	0,017	0,384
RH_paracentral	-0,725	0,135	0,240	0,391	0,567	0,093
RH_parahippocampal	-0,308	0,368	0,145	0,067	0,994	0,009
RH_parsopercularis	-0,640	0,023	0,364	0,618	0,060	0,306
RH_parsorbitalis	-0,091	0,701	0,063	0,161	1,000	0,000
RH_parstriangularis	-0,940	0,005	0,455	0,360	0,630	0,078
RH_pericalcarine	-0,799	0,121	0,249	1,049	0,060	0,302
RH_postcentral	-0,851	0,267	0,194	0,690	0,001	0,515
RH_posteriorcingulate	-0,453	0,895	0,022	-0,603	0,258	0,184
RH_precentral	-0,898	0,090	0,274	-0,021	0,950	0,011
RH_precuneus	-0,479	0,325	0,159	0,163	0,582	0,088
RH_rostralanteriorcingulate	-0,724	0,422	0,130	-0,135	0,829	0,036
RH_rostralmiddlefrontal	-0,681	0,013	0,399	0,868	0,002	0,489
RH_superiorfrontal	-0,317	0,516	0,105	0,341	0,018	0,379
RH_superiorparietal	-0,625	0,315	0,162	0,383	0,008	0,424
RH_superiortemporal	-0,346	0,913	0,009	0,515	0,003	0,474
RH_supramarginal	-0,530	0,308	0,164	0,437	0,103	0,268
RH_temporalpole	-0,436	0,863	0,029	0,729	0,336	0,155
RH_transversetemporal	-0,464	0,738	0,055	0,506	0,593	0,077

Supplementary Table S2.3.A | Cortical morphological estimates. Percentage of change, q-value and effect size of the changes after manual editing.

Supplementary Table S2.3.B

ROI	Percentage of change (%)	q-value	Effect size (r)
BrainStem	-3,106	0,000	0,863
CCAnterior	0,027	0,000	0,588
CCCentral	0,337	0,029	0,351
CCMidAnterior	4,390	0,000	0,697
CCMidPosterior	7,007	0,000	0,641
CCPosterior	1,993	0,000	0,556
CSF	1,870	0,096	0,268
LH_Accumbens	3,422	0,066	0,295
LH_Amygdala	0,907	0,199	0,208
LH_Caudate	-2,168	0,000	0,632
LH_CerebellumCortex	-2,189	0,000	0,749
LH_CerebellumWhiteMatter	11,036	0,000	0,838
LH_Hippocampus	0,170	0,694	0,065
LH_InfLatVent	0,484	0,929	0,016
LH_LateralVentricle	-1,164	0,000	0,827
LH_Pallidus	-0,439	0,503	0,109
LH_Putamen	-1,766	0,000	0,589
LH_ThalamusProper	-0,985	0,009	0,416
LH_VentralDC	4,168	0,000	0,778
LH_choroidplexus	4,809	0,005	0,451
LH_vessel	12,239	0,043	0,324
OpticChiasm	0,976	0,829	0,036
RH_Accumbens	3,375	0,008	0,423
RH_Amygdala	1,938	0,085	0,277
RH_Caudate	-1,848	0,000	0,539
RH_CerebellumCortex	-1,924	0,000	0,670
RH_CerebellumWhiteMatter	8,980	0,000	0,675
RH_Hippocampus	-0,038	0,885	0,025
RH_InfLatVent	3,953	0,003	0,474
RH_LateralVentricle	-0,839	0,000	0,568
RH_Pallidus	0,226	0,625	0,079
RH_Putamen	-1,482	0,031	0,344
RH_ThalamusProper	-2,145	0,000	0,592
RH_VentralDC	5,416	0,000	0,816
RH_choroidplexus	3,298	0,075	0,286
RH_vessel	9,065	0,265	0,181
3rdVentricle	-0,303	0,410	0,133
4thVentricle	-7,212	0,000	0,863

*Supplementary Table S2.3.B | **Brain volumes.** Percentage of change, q-value and effect size of the changes after manual editing.*

Supplementary Table S2.3.C

ROI	Percentage of change (%)	q-value	Effect size (r)
LH_CA1	-0,305	0,994	0,002
LH_CA3	1,088	0,142	0,237
LH_CA4	0,261	0,372	0,145
LH_GCMLDG	0,307	0,343	0,154
LH_HATA	1,199	0,194	0,210
LH_Hippocampaltail	-2,195	0,001	0,534
LH_fimbria	24,506	0,000	0,869
LH_hippocampalfissure	-0,477	0,194	0,210
LH_molecularlayerHP	1,494	0,010	0,409
LH_parasubiculum	-0,309	0,204	0,206
LH_presubiculum	-1,074	0,051	0,313
LH_subiculum	-0,242	0,435	0,127
RH_CA1	-0,595	0,070	0,290
RH_CA3	1,974	0,001	0,505
RH_CA4	0,672	0,209	0,203
RH_GCMLDG	0,628	0,225	0,197
RH_HATA	1,774	0,704	0,063
RH_Hippocampaltail	-3,424	0,000	0,740
RH_fimbria	24,013	0,000	0,871
RH_hippocampalfissure	1,047	0,539	0,101
RH_molecularlayerHP	0,522	0,644	0,076
RH_parasubiculum	-0,145	0,983	0,004
RH_presubiculum	-2,000	0,000	0,592
RH_subiculum	-0,457	0,180	0,217

*Supplementary Table S2.3.C | **Hippocampal subfields.** Percentage of change, q-value and effect size of the changes after manual editing.*

Supplementary Figure S2.3
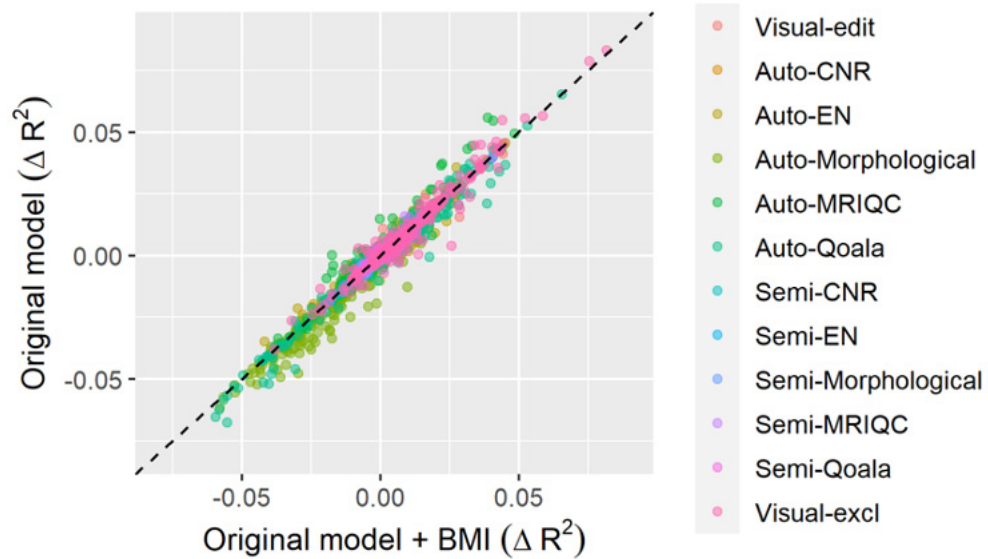
Supplementary Figure S2.3 | Distribution of ΔR^2obtained by each QC strategy using the original regression model (brain ~ age + sex) plotted vs. the distribution of ΔR^2 obtained by each QC strategy when adding BMI as a covariate (brain ~ age + sex). *Dot colors differentiate between quality control strategies. Adding BMI as a covariate, despite small variations, preserves the overall pattern.*

Supplementary Table S2.4

Region	Non-QC	Visual-edit	Auto-CNR	Auto-EN	Auto-Morpho	Auto-MRIQC	Auto-Qoala	Semi-CNR	Semi-EN	Semi-Morpho	Semi-MRIQC	Semi-Qoala	Visual-excl
Area													
LH_bankssts	0,11	0,12	0,12	0,12	0,09	0,11	0,12	0,12	0,11	0,11	0,12	0,11	0,12
LH_caudalanteriorcingulate	0,02	0,02	0,02	0,02	0,01	0,02	0,02	0,03	0,03	0,02	0,03	0,02	0,03
LH_caudalmiddlefrontal	0,13	0,14	0,14	0,14	0,10	0,14	0,14	0,14	0,14	0,14	0,14	0,14	0,18
LH_cuneus	0,14	0,15	0,15	0,17	0,11	0,15	0,14	0,15	0,16	0,14	0,16	0,14	0,19
LH_entorhinal	0,02	0,01	0,02	0,01	0,01	0,01	0,03	0,02	0,01	0,01	0,01	0,02	0,02
LH_fusiform	0,22	0,23	0,22	0,22	0,16	0,22	0,22	0,22	0,22	0,21	0,22	0,22	0,23
LH_inferiorparietal	0,10	0,11	0,11	0,11	0,06	0,10	0,10	0,11	0,11	0,10	0,11	0,10	0,11
LH_inferiortemporal	0,14	0,14	0,14	0,16	0,11	0,13	0,16	0,15	0,15	0,14	0,15	0,14	0,18
LH_isthmuscingulate	0,18	0,19	0,19	0,19	0,15	0,19	0,21	0,19	0,20	0,18	0,19	0,19	0,20
LH_lateraloccipital	0,23	0,24	0,24	0,27	0,20	0,25	0,23	0,24	0,24	0,23	0,24	0,23	0,25
LH_lateralorbitofrontal	0,19	0,19	0,20	0,22	0,16	0,19	0,22	0,20	0,20	0,19	0,19	0,20	0,20
LH_lingual	0,14	0,15	0,15	0,16	0,11	0,15	0,16	0,16	0,16	0,14	0,16	0,13	0,17
LH_medialorbitofrontal	0,16	0,16	0,16	0,16	0,12	0,15	0,18	0,16	0,16	0,16	0,16	0,16	0,15
LH_middletemporal	0,19	0,19	0,20	0,21	0,15	0,18	0,21	0,20	0,20	0,19	0,20	0,19	0,21
LH_parahippocampal	0,05	0,06	0,06	0,06	0,01	0,05	0,06	0,06	0,06	0,04	0,06	0,05	0,05
LH_paracentral	0,11	0,10	0,11	0,11	0,09	0,12	0,11	0,11	0,11	0,11	0,11	0,11	0,11
LH_parsopercularis	0,04	0,05	0,05	0,05	0,05	0,04	0,05	0,05	0,05	0,05	0,05	0,04	0,06
LH_parsorbitalis	0,15	0,15	0,14	0,15	0,11	0,14	0,18	0,15	0,15	0,14	0,15	0,15	0,13
LH_parstriangularis	0,11	0,11	0,11	0,11	0,10	0,11	0,15	0,11	0,11	0,11	0,11	0,11	0,10
LH_pericalcarine	0,08	0,09	0,09	0,10	0,05	0,09	0,08	0,09	0,09	0,08	0,09	0,08	0,10
LH_postcentral	0,26	0,25	0,26	0,25	0,22	0,26	0,27	0,26	0,26	0,26	0,26	0,26	0,28
LH_posteriorcingulate	0,14	0,14	0,14	0,14	0,10	0,14	0,15	0,14	0,14	0,14	0,14	0,15	0,16
LH_precentral	0,23	0,23	0,23	0,23	0,19	0,24	0,22	0,23	0,23	0,22	0,23	0,23	0,28
LH_precuneus	0,13	0,13	0,13	0,15	0,08	0,12	0,13	0,13	0,13	0,12	0,13	0,13	0,14
LH_rostralanteriorcingulate	0,12	0,12	0,12	0,13	0,07	0,11	0,13	0,13	0,13	0,12	0,13	0,13	0,13
LH_rostralmiddlefrontal	0,24	0,23	0,24	0,24	0,19	0,24	0,27	0,24	0,24	0,23	0,24	0,24	0,24
LH_superiorfrontal	0,25	0,25	0,25	0,27	0,20	0,25	0,26	0,25	0,25	0,25	0,25	0,25	0,26
LH_superiorparietal	0,06	0,06	0,06	0,06	0,03	0,06	0,06	0,06	0,06	0,06	0,06	0,06	0,07
LH_superiortemporal	0,20	0,21	0,20	0,21	0,16	0,21	0,21	0,20	0,20	0,20	0,20	0,20	0,20
LH_supramarginal	0,18	0,18	0,18	0,18	0,15	0,16	0,18	0,18	0,18	0,18	0,18	0,18	0,18
LH_frontalpole	0,17	0,15	0,17	0,16	0,13	0,17	0,19	0,17	0,16	0,16	0,16	0,16	0,14
LH_temporalpole	0,13	0,11	0,12	0,12	0,10	0,11	0,15	0,12	0,11	0,12	0,12	0,13	0,10
LH_transversetemporal	0,10	0,10	0,10	0,10	0,11	0,11	0,11	0,10	0,10	0,10	0,10	0,10	0,12
LH_insula	0,18	0,17	0,18	0,19	0,14	0,18	0,18	0,18	0,18	0,17	0,18	0,18	0,19

Region	Non-QC	Visual-edit	Auto-CNR	Auto-EN	Auto-Morpho	Auto-MRIQC	Auto-Qoala	Semi-CNR	Semi-EN	Semi-Morpho	Semi-MRIQC	Semi-Qoala	Visual-excl
RH_bankssts	0,06	0,06	0,06	0,06	0,03	0,05	0,08	0,06	0,06	0,05	0,05	0,06	0,06
RH_caudalanteriorcingulate	0,11	0,11	0,11	0,12	0,06	0,12	0,12	0,11	0,11	0,11	0,11	0,11	0,13
RH_caudalmiddlefrontal	0,11	0,12	0,11	0,11	0,08	0,11	0,12	0,11	0,11	0,11	0,11	0,12	0,12
RH_cuneus	0,15	0,14	0,15	0,15	0,12	0,15	0,14	0,15	0,15	0,14	0,15	0,15	0,15
RH_entorhinal	0,06	0,06	0,05	0,07	0,05	0,05	0,09	0,06	0,06	0,05	0,06	0,06	0,08
RH_fusiform	0,22	0,23	0,22	0,23	0,18	0,23	0,22	0,22	0,22	0,22	0,22	0,23	0,23
RH_inferiorparietal	0,21	0,22	0,22	0,22	0,18	0,20	0,22	0,22	0,22	0,21	0,22	0,21	0,23
RH_inferiortemporal	0,20	0,20	0,20	0,21	0,16	0,19	0,21	0,20	0,20	0,20	0,20	0,20	0,21
RH_isthmuscingulate	0,15	0,15	0,15	0,15	0,13	0,14	0,15	0,15	0,15	0,15	0,15	0,16	0,16
RH_lateraloccipital	0,21	0,21	0,21	0,22	0,18	0,21	0,22	0,21	0,21	0,21	0,21	0,22	0,22
RH_lateralorbitofrontal	0,16	0,16	0,17	0,19	0,12	0,18	0,18	0,17	0,18	0,16	0,17	0,16	0,16
RH_lingual	0,13	0,12	0,13	0,13	0,09	0,11	0,14	0,13	0,13	0,13	0,13	0,13	0,13
RH_medialorbitofrontal	0,19	0,19	0,19	0,21	0,15	0,19	0,22	0,19	0,20	0,19	0,20	0,20	0,21
RH_middletemporal	0,19	0,20	0,19	0,20	0,14	0,18	0,19	0,19	0,19	0,18	0,19	0,19	0,20
RH_parahippocampal	0,09	0,08	0,08	0,09	0,05	0,09	0,11	0,09	0,09	0,08	0,09	0,09	0,09
RH_paracentral	0,13	0,12	0,13	0,13	0,11	0,13	0,14	0,13	0,13	0,12	0,13	0,12	0,11
RH_parsopercularis	0,08	0,09	0,08	0,09	0,06	0,09	0,09	0,08	0,08	0,09	0,08	0,08	0,09
RH_parsorbitalis	0,15	0,15	0,15	0,17	0,12	0,15	0,19	0,15	0,16	0,15	0,15	0,15	0,14
RH_parstriangularis	0,15	0,15	0,15	0,15	0,13	0,16	0,17	0,15	0,15	0,15	0,15	0,15	0,14
RH_pericalcarine	0,07	0,07	0,07	0,07	0,05	0,07	0,08	0,07	0,07	0,07	0,07	0,07	0,06
RH_postcentral	0,17	0,17	0,18	0,17	0,13	0,16	0,15	0,18	0,18	0,17	0,18	0,17	0,19
RH_posteriorcingulate	0,17	0,17	0,17	0,18	0,13	0,16	0,17	0,17	0,17	0,17	0,17	0,17	0,18
RH_precentral	0,21	0,20	0,20	0,21	0,17	0,20	0,21	0,21	0,20	0,21	0,20	0,21	0,25
RH_precuneus	0,14	0,13	0,14	0,15	0,08	0,14	0,15	0,14	0,14	0,14	0,14	0,14	0,15
RH_rostralanteriorcingulate	0,09	0,09	0,09	0,09	0,05	0,09	0,09	0,09	0,09	0,09	0,08	0,09	0,08
RH_rostralmiddlefrontal	0,22	0,21	0,21	0,22	0,16	0,21	0,24	0,21	0,22	0,21	0,21	0,22	0,21
RH_superiorfrontal	0,22	0,21	0,22	0,23	0,17	0,23	0,24	0,22	0,22	0,22	0,22	0,22	0,22
RH_superiorparietal	0,10	0,10	0,11	0,13	0,07	0,10	0,12	0,11	0,11	0,10	0,11	0,10	0,13
RH_superiortemporal	0,17	0,17	0,17	0,19	0,14	0,19	0,20	0,17	0,17	0,17	0,17	0,17	0,19
RH_supramarginal	0,11	0,12	0,12	0,12	0,07	0,12	0,13	0,12	0,12	0,10	0,12	0,11	0,12
RH_frontalpole	0,15	0,12	0,15	0,15	0,10	0,14	0,16	0,15	0,14	0,14	0,14	0,14	0,14
RH_temporalpole	0,11	0,12	0,11	0,11	0,08	0,09	0,11	0,11	0,11	0,10	0,11	0,11	0,11
RH_transversetemporal	0,14	0,15	0,14	0,15	0,13	0,17	0,17	0,14	0,14	0,15	0,14	0,15	0,14
RH_insula	0,18	0,19	0,18	0,19	0,14	0,18	0,20	0,18	0,17	0,18	0,17	0,19	0,20

Region	Non-QC	Visual-edit	Auto-CNR	Auto-EN	Auto-Morpho	Auto-MRIQC	Auto-Qoala	Semi-CNR	Semi-EN	Semi-Morpho	Semi-MRIQC	Semi-Qoala	Visual-excl
Thickness													
LH_bankssts	0,09	0,10	0,09	0,07	0,07	0,09	0,04	0,09	0,09	0,09	0,09	0,09	0,09
LH_caudalanteriorcingulate	0,06	0,06	0,05	0,06	0,06	0,05	0,04	0,05	0,06	0,05	0,06	0,06	0,05
LH_caudalmiddlefrontal	0,10	0,11	0,10	0,09	0,07	0,08	0,04	0,10	0,10	0,10	0,10	0,10	0,12
LH_cuneus	0,03	0,04	0,03	0,03	0,02	0,02	0,03	0,03	0,03	0,03	0,03	0,03	0,03
LH_entorhinal	0,02	0,03	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
LH_fusiform	0,05	0,06	0,05	0,03	0,02	0,06	0,01	0,05	0,05	0,05	0,06	0,05	0,04
LH_inferiorparietal	0,12	0,13	0,12	0,11	0,10	0,12	0,06	0,12	0,12	0,12	0,12	0,12	0,14
LH_inferiortemporal	0,05	0,06	0,05	0,06	0,05	0,06	0,05	0,05	0,05	0,05	0,05	0,05	0,05
LH_isthmuscingulate	0,02	0,03	0,02	0,04	0,02	0,03	0,01	0,02	0,03	0,02	0,03	0,03	0,03
LH_lateraloccipital	0,02	0,04	0,02	0,02	0,01	0,03	0,03	0,02	0,03	0,02	0,02	0,02	0,02
LH_lateralorbitofrontal	0,03	0,04	0,03	0,04	0,02	0,03	0,02	0,03	0,03	0,03	0,03	0,03	0,03
LH_lingual	0,03	0,05	0,03	0,04	0,02	0,04	0,03	0,03	0,04	0,03	0,04	0,03	0,04
LH_medialorbitofrontal	0,02	0,03	0,02	0,03	0,01	0,02	0,03	0,02	0,02	0,01	0,02	0,02	0,02
LH_middletemporal	0,08	0,11	0,08	0,10	0,07	0,10	0,07	0,08	0,09	0,08	0,09	0,08	0,11
LH_parahippocampal	0,08	0,09	0,08	0,05	0,07	0,07	0,03	0,08	0,08	0,08	0,08	0,08	0,09
LH_paracentral	0,08	0,09	0,08	0,06	0,07	0,07	0,03	0,08	0,08	0,08	0,08	0,08	0,09
LH_parsopercularis	0,12	0,13	0,12	0,13	0,12	0,14	0,09	0,12	0,12	0,12	0,12	0,12	0,12
LH_parsorbitalis	0,03	0,03	0,03	0,03	0,02	0,03	0,01	0,03	0,03	0,03	0,03	0,03	0,04
LH_parstriangularis	0,11	0,12	0,11	0,11	0,11	0,12	0,08	0,11	0,12	0,11	0,11	0,12	0,15
LH_pericalcarine	0,02	0,01	0,01	0,01	0,02	0,02	0,03	0,01	0,01	0,02	0,01	0,01	0,02
LH_postcentral	0,09	0,11	0,09	0,08	0,07	0,09	0,04	0,09	0,09	0,09	0,09	0,09	0,10
LH_posteriorcingulate	0,03	0,03	0,03	0,02	0,02	0,02	0,01	0,03	0,03	0,03	0,03	0,03	0,03
LH_precentral	0,11	0,12	0,11	0,09	0,09	0,10	0,06	0,11	0,11	0,11	0,11	0,11	0,12
LH_precuneus	0,10	0,12	0,11	0,09	0,07	0,11	0,06	0,11	0,11	0,10	0,11	0,10	0,12
LH_rostralanteriorcingulate	0,07	0,08	0,07	0,08	0,07	0,06	0,02	0,07	0,07	0,07	0,07	0,08	0,10
LH_rostralmiddlefrontal	0,06	0,07	0,06	0,08	0,05	0,07	0,04	0,06	0,06	0,06	0,06	0,06	0,08
LH_superiorfrontal	0,12	0,13	0,12	0,11	0,11	0,12	0,05	0,12	0,12	0,12	0,12	0,12	0,13
LH_superiorparietal	0,10	0,11	0,10	0,09	0,07	0,09	0,07	0,10	0,10	0,10	0,10	0,10	0,12
LH_superiortemporal	0,14	0,16	0,14	0,13	0,11	0,14	0,09	0,14	0,14	0,14	0,14	0,13	0,16
LH_supramarginal	0,18	0,20	0,18	0,15	0,16	0,17	0,12	0,18	0,18	0,18	0,18	0,18	0,18
LH_frontalpole	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01
LH_temporalpole	0,03	0,03	0,03	0,02	0,02	0,03	0,00	0,03	0,02	0,03	0,03	0,03	0,03
LH_transversetemporal	0,02	0,04	0,02	0,01	0,01	0,02	0,00	0,02	0,03	0,02	0,03	0,02	0,03
LH_insula	0,02	0,03	0,02	0,02	0,01	0,02	0,01	0,02	0,02	0,02	0,02	0,02	0,02
RH_bankssts	0,11	0,12	0,11	0,09	0,09	0,12	0,07	0,11	0,11	0,11	0,11	0,11	0,11

Region	Non-QC	Visual-edit	Auto-CNR	Auto-EN	Auto-Morpho	Auto-MRIQC	Auto-Qoala	Semi-CNR	Semi-EN	Semi-Morpho	Semi-MRIQC	Semi-Qoala	Visual-excl
RH_caudalanteriorcingulate	0,03	0,04	0,03	0,03	0,03	0,03	0,01	0,03	0,03	0,03	0,03	0,03	0,02
RH_caudalmiddlefrontal	0,11	0,11	0,11	0,10	0,09	0,11	0,07	0,11	0,11	0,10	0,11	0,11	0,14
RH_cuneus	0,03	0,04	0,04	0,04	0,03	0,04	0,05	0,04	0,04	0,03	0,04	0,03	0,04
RH_entorhinal	0,02	0,02	0,02	0,01	0,02	0,02	0,01	0,02	0,02	0,02	0,03	0,02	0,01
RH_fusiform	0,07	0,08	0,07	0,05	0,05	0,06	0,03	0,07	0,08	0,07	0,08	0,07	0,07
RH_inferiorparietal	0,18	0,20	0,19	0,18	0,16	0,20	0,17	0,19	0,19	0,18	0,19	0,19	0,20
RH_inferiortemporal	0,05	0,06	0,05	0,05	0,04	0,04	0,05	0,05	0,06	0,05	0,05	0,05	0,04
RH_isthmuscingulate	0,03	0,03	0,03	0,03	0,04	0,03	0,03	0,03	0,03	0,04	0,03	0,03	0,03
RH_lateraloccipital	0,04	0,06	0,04	0,04	0,02	0,04	0,06	0,04	0,05	0,04	0,04	0,04	0,05
RH_lateralorbitofrontal	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,01
RH_lingual	0,02	0,03	0,02	0,03	0,01	0,01	0,03	0,02	0,02	0,02	0,02	0,02	0,02
RH_medialorbitofrontal	0,00	0,01	0,00	0,01	0,00	0,01	0,00	0,00	0,01	0,00	0,00	0,00	0,00
RH_middletemporal	0,11	0,14	0,12	0,13	0,09	0,12	0,10	0,12	0,12	0,11	0,12	0,11	0,15
RH_parahippocampal	0,07	0,08	0,08	0,07	0,07	0,07	0,04	0,07	0,08	0,07	0,08	0,07	0,08
RH_paracentral	0,07	0,08	0,07	0,06	0,06	0,07	0,03	0,07	0,07	0,07	0,08	0,07	0,09
RH_parsopercularis	0,09	0,10	0,09	0,10	0,07	0,09	0,09	0,10	0,10	0,09	0,10	0,10	0,10
RH_parsorbitalis	0,02	0,01	0,02	0,02	0,01	0,01	0,00	0,02	0,02	0,02	0,01	0,01	0,04
RH_parstriangularis	0,11	0,12	0,11	0,11	0,09	0,11	0,08	0,11	0,11	0,11	0,11	0,11	0,12
RH_pericalcarine	0,04	0,03	0,05	0,05	0,06	0,05	0,07	0,05	0,04	0,04	0,04	0,04	0,03
RH_postcentral	0,09	0,10	0,09	0,08	0,07	0,08	0,06	0,09	0,09	0,09	0,10	0,10	0,12
RH_posteriorcingulate	0,04	0,05	0,04	0,04	0,03	0,03	0,01	0,04	0,04	0,04	0,04	0,04	0,05
RH_precentral	0,10	0,11	0,11	0,09	0,09	0,10	0,07	0,11	0,11	0,10	0,11	0,10	0,13
RH_precuneus	0,15	0,16	0,15	0,14	0,13	0,14	0,14	0,15	0,16	0,15	0,15	0,15	0,14
RH_rostralanteriorcingulate	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01
RH_rostralmiddlefrontal	0,02	0,03	0,02	0,04	0,02	0,02	0,01	0,02	0,02	0,02	0,02	0,02	0,04
RH_superiorfrontal	0,08	0,08	0,08	0,07	0,06	0,06	0,04	0,08	0,08	0,08	0,08	0,07	0,09
RH_superiorparietal	0,14	0,14	0,14	0,12	0,12	0,13	0,13	0,14	0,14	0,14	0,14	0,14	0,15
RH_superiortemporal	0,20	0,21	0,20	0,19	0,16	0,19	0,13	0,20	0,20	0,20	0,20	0,20	0,22
RH_supramarginal	0,16	0,17	0,16	0,14	0,14	0,15	0,13	0,16	0,16	0,16	0,16	0,16	0,16
RH_frontalpole	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00
RH_temporalpole	0,05	0,06	0,05	0,04	0,04	0,04	0,02	0,05	0,05	0,05	0,06	0,05	0,05
RH_transversetemporal	0,03	0,03	0,02	0,02	0,01	0,02	0,01	0,03	0,03	0,02	0,03	0,03	0,04
RH_insula	0,05	0,06	0,05	0,04	0,04	0,04	0,04	0,05	0,05	0,05	0,06	0,05	0,04

Region	Non-QC	Visual-edit	Auto-CNR	Auto-EN	Auto-Morpho	Auto-MRIQC	Auto-Qoala	Semi-CNR	Semi-EN	Semi-Morpho	Semi-MRIQC	Semi-Qoala	Visual-excl
Subcortical volumes													
LH_LateralVentricle	0,22	0,23	0,22	0,22	0,22	0,21	0,17	0,22	0,23	0,23	0,22	0,23	0,23
LH_InfLatVent	0,23	0,24	0,23	0,22	0,23	0,20	0,21	0,23	0,23	0,24	0,24	0,24	0,21
LH_CerebellumWhiteMatter	0,14	0,13	0,14	0,15	0,11	0,17	0,16	0,14	0,13	0,14	0,14	0,13	0,18
LH_CerebellumCortex	0,25	0,25	0,25	0,27	0,22	0,26	0,29	0,25	0,26	0,25	0,25	0,26	0,26
LH_ThalamusProper	0,29	0,29	0,30	0,32	0,26	0,34	0,30	0,30	0,30	0,29	0,29	0,29	0,33
LH_Caudate	0,05	0,05	0,05	0,06	0,03	0,05	0,06	0,05	0,05	0,04	0,05	0,05	0,05
LH_Putamen	0,24	0,24	0,24	0,28	0,22	0,26	0,29	0,24	0,24	0,24	0,24	0,24	0,32
LH_Pallidum	0,12	0,12	0,12	0,13	0,09	0,16	0,15	0,12	0,13	0,12	0,12	0,12	0,12
3rdVentricle	0,32	0,32	0,32	0,32	0,33	0,29	0,28	0,32	0,32	0,32	0,32	0,33	0,32
4thVentricle	0,09	0,08	0,09	0,09	0,08	0,09	0,10	0,09	0,08	0,09	0,08	0,09	0,08
BrainStem	0,23	0,22	0,23	0,25	0,19	0,27	0,26	0,23	0,24	0,23	0,23	0,23	0,27
LH_Hippocampus	0,15	0,18	0,16	0,17	0,12	0,18	0,16	0,16	0,16	0,15	0,17	0,15	0,19
LH_Amygdala	0,20	0,19	0,20	0,22	0,19	0,19	0,24	0,20	0,19	0,19	0,20	0,19	0,23
CSF	0,19	0,19	0,19	0,17	0,17	0,17	0,15	0,19	0,19	0,20	0,19	0,20	0,20
LH_Accumbens	0,24	0,26	0,25	0,25	0,21	0,27	0,24	0,24	0,24	0,24	0,25	0,24	0,28
LH_VentralDC	0,26	0,27	0,27	0,30	0,21	0,30	0,30	0,27	0,27	0,26	0,27	0,27	0,31
LH_vessel	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,01	0,01	0,01	0,01	0,01	0,01
LH_choroidplexus	0,28	0,30	0,28	0,27	0,29	0,28	0,27	0,28	0,29	0,28	0,29	0,29	0,27
RH_LateralVentricle	0,19	0,21	0,19	0,20	0,19	0,19	0,15	0,19	0,20	0,20	0,20	0,20	0,20
RH_InfLatVent	0,19	0,19	0,18	0,17	0,18	0,15	0,19	0,18	0,18	0,19	0,18	0,19	0,16
RH_CerebellumWhiteMatter	0,13	0,12	0,13	0,14	0,10	0,15	0,15	0,13	0,12	0,13	0,13	0,13	0,16
RH_CerebellumCortex	0,27	0,26	0,27	0,29	0,24	0,27	0,30	0,27	0,28	0,26	0,26	0,27	0,29
RH_ThalamusProper	0,25	0,24	0,25	0,27	0,21	0,30	0,26	0,25	0,26	0,25	0,24	0,25	0,29
RH_Caudate	0,04	0,05	0,04	0,05	0,03	0,05	0,05	0,04	0,04	0,04	0,04	0,04	0,04
RH_Putamen	0,24	0,23	0,24	0,26	0,21	0,28	0,27	0,23	0,24	0,23	0,23	0,24	0,31
RH_Pallidum	0,11	0,10	0,11	0,11	0,08	0,14	0,13	0,11	0,11	0,10	0,10	0,11	0,11
RH_Hippocampus	0,17	0,18	0,18	0,21	0,17	0,18	0,17	0,17	0,18	0,17	0,18	0,17	0,19
RH_Amygdala	0,19	0,19	0,19	0,21	0,18	0,19	0,21	0,19	0,19	0,19	0,19	0,19	0,21
RH_Accumbens	0,23	0,22	0,24	0,25	0,22	0,25	0,21	0,23	0,23	0,23	0,23	0,23	0,25
RH_VentralDC	0,27	0,29	0,28	0,30	0,23	0,33	0,30	0,27	0,28	0,27	0,29	0,28	0,32
RH_vessel	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
RH_choroidplexus	0,32	0,33	0,32	0,29	0,33	0,29	0,28	0,32	0,32	0,32	0,33	0,32	0,30
OpticChiasm	0,03	0,03	0,03	0,05	0,03	0,04	0,05	0,03	0,03	0,03	0,03	0,03	0,04
CCPosterior	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,01

Region	Non-QC	Visual-edit	Auto-CNR	Auto-EN	Auto-Morpho	Auto-MRIQC	Auto-Qoala	Semi-CNR	Semi-EN	Semi-Morpho	Semi-MRIQC	Semi-Qoala	Visual-excl
CCMidPosterior	0,04	0,03	0,04	0,04	0,03	0,05	0,04	0,04	0,03	0,04	0,04	0,03	0,06
CCCentral	0,06	0,07	0,06	0,06	0,05	0,07	0,06	0,06	0,06	0,06	0,06	0,06	0,09
CCMidAnterior	0,07	0,07	0,07	0,08	0,06	0,10	0,09	0,07	0,07	0,07	0,07	0,07	0,10
CCAnterior	0,06	0,06	0,06	0,06	0,04	0,09	0,07	0,06	0,06	0,06	0,06	0,06	0,08

Hippocampal subfields													
LH_Hippocampaltail	0,12	0,16	0,15	0,15	0,13	0,17	0,11	0,15	0,16	0,15	0,16	0,12	0,17
RH_Hippocampaltail	0,14	0,15	0,14	0,15	0,11	0,16	0,15	0,14	0,15	0,14	0,15	0,15	0,16
LH_subiculum	0,10	0,11	0,11	0,12	0,07	0,12	0,11	0,11	0,11	0,10	0,11	0,11	0,11
LH_presubiculum	0,10	0,11	0,10	0,11	0,07	0,11	0,11	0,10	0,11	0,10	0,10	0,10	0,10
LH_parasubiculum	0,07	0,07	0,07	0,07	0,04	0,08	0,09	0,07	0,07	0,07	0,07	0,07	0,08
RH_subiculum	0,09	0,09	0,09	0,09	0,06	0,09	0,08	0,09	0,09	0,08	0,09	0,09	0,08
RH_presubiculum	0,11	0,11	0,11	0,11	0,10	0,11	0,13	0,11	0,11	0,11	0,11	0,11	0,10
RH_parasubiculum	0,11	0,12	0,11	0,11	0,09	0,12	0,13	0,10	0,10	0,10	0,11	0,11	0,11
LH_hippocampalfissure	0,13	0,12	0,13	0,11	0,12	0,10	0,20	0,13	0,13	0,13	0,13	0,13	0,09
RH_hippocampalfissure	0,10	0,10	0,10	0,08	0,10	0,06	0,12	0,10	0,10	0,10	0,09	0,10	0,09
LH_molecularlayerHP	0,13	0,12	0,12	0,11	0,11	0,12	0,15	0,13	0,12	0,12	0,13	0,13	0,11
RH_molecularlayerHP	0,09	0,08	0,08	0,09	0,04	0,08	0,09	0,09	0,09	0,08	0,08	0,09	0,09
LH_CA1	0,14	0,15	0,14	0,18	0,10	0,16	0,15	0,14	0,14	0,14	0,14	0,14	0,18
LH_CA3	0,14	0,15	0,14	0,13	0,13	0,14	0,12	0,14	0,14	0,14	0,14	0,14	0,15
LH_CA4	0,13	0,14	0,13	0,14	0,09	0,14	0,13	0,13	0,13	0,13	0,13	0,13	0,16
RH_CA1	0,11	0,12	0,12	0,16	0,07	0,13	0,11	0,12	0,12	0,12	0,12	0,11	0,15
RH_CA3	0,13	0,14	0,14	0,16	0,13	0,14	0,11	0,14	0,14	0,14	0,14	0,13	0,17
RH_CA4	0,12	0,12	0,12	0,15	0,08	0,13	0,11	0,12	0,12	0,11	0,12	0,12	0,14
LH_GCMLDG	0,17	0,18	0,18	0,19	0,15	0,18	0,16	0,18	0,18	0,17	0,18	0,17	0,20
RH_GCMLDG	0,14	0,15	0,14	0,18	0,13	0,16	0,13	0,14	0,15	0,14	0,15	0,14	0,16
LH_HATA	0,10	0,11	0,11	0,11	0,08	0,11	0,12	0,11	0,11	0,10	0,11	0,10	0,11
RH_HATA	0,10	0,10	0,10	0,11	0,06	0,11	0,10	0,10	0,10	0,09	0,10	0,10	0,09
LH_fimbria	0,20	0,18	0,21	0,25	0,19	0,23	0,20	0,20	0,19	0,20	0,21	0,19	0,25
RH_fimbria	0,15	0,13	0,15	0,16	0,12	0,16	0,12	0,15	0,13	0,15	0,14	0,14	0,15

Supplementary Table S2.4 | *R2 obtained by each QC strategy in each brain morphological estimate.* *Higher color intensity illustrates larger R2 values.*

Supplementary Table S2.5

Dsc / n	Morpho	CNR	EN	MRIQC	Qoala-T	Visual
Morpho	1 / 52	0,04	0,21	0,10	0,25	0,12
CNR	1	1 / 2	0,14	0,13	0,00	0,04
EN	8	2	1 / 26	0,40	0,28	0,36
MRIQC	4	2	11	1 / 29	0,22	0,27
Qoala-T	13	0	11	9	1 / 54	0,22
Visual	6	1	13	10	11	1 / 46

Supplementary Table S2.5 | Overlapping cases detected as inaccurate by the different tools and metrics composing the QC strategies. *Visual inspection (Visual), MRIQC, Qoala-T, Global morphological measures (Morpho), Euler numbers (EN) and Contrast-to-noise ratio (CNR). Lower triangle (in blue) shows the number (n) of overlapping segmentations, while the upper triangle (in green) shows the Dice similarity coefficient (Dsc) for each overlap.*

CHAPTER 3

THE ASSOCIATION OF PREDIABETES AND TYPE 2 DIABETES WITH HIPPOCAMPAL SUBFIELDS VOLUME: THE MAASTRICHT STUDY

Monereo-Sanchez, J.; Jansen, F.A.J.; Köhler, S.; van Boxtel, M. P. J.; Backes, W.H.; Stehouwer, C.D.A.; Kroon, A.A.; Kooman, J.P.; Schalkwijk, C.G.; Linden, D.E.J; Schram, M.T.

Abstract

We investigated whether prediabetes, type 2 diabetes, and continuous measures of hyperglycemia are associated with tissue volume differences in specific subfields of the hippocampus.

We used cross-sectional data from 4,724 participants (58.7±8.5 years, 51.5% women) of The Maastricht Study, a population-based prospective cohort. Glucose metabolism status was assessed with an oral glucose tolerance test, and defined as type 2 diabetes (n=869), prediabetes (n=671), or normal glucose metabolism (n=3184). We extracted 12 hippocampal subfield volumes per hemisphere with FreeSurfer v6.0 using T1w and FLAIR 3T MRI images. We used multiple linear regression and linear trend analysis, and adjusted for total intracranial volume, demographic, lifestyle, and cardiovascular risk factors.

Type 2 diabetes was significantly associated with smaller volumes in the hippocampal subfield fimbria (standardized beta coefficient ± standard error (β±SE) =-0.195±0.04, p-value<0.001), the hippocampus proper, i.e. Cornu Ammonis (CA) 1, CA2/3, CA4, dentate gyrus, subiculum and presubiculum (β±SE<-0.105±0.04, p-value<0.006); as well as the hippocampal tail (β±SE=-0.162±0.04, p-value<0.001). Prediabetes showed no significant associations. However, linear trend analysis indicated a dose-response relation from normal glucose metabolism, to prediabetes, to type 2 diabetes. Multiple continuous measures of hyperglycemia were associated with smaller volumes of the subfields fimbria (β±SE<-0.010±0.011, p-value<0.001), dentate gyrus (β±SE<-0.013±0.010, p-value<0.002), CA3 (β±SE<-0.014±0.011, p-value<0.001), and tail (β±SE<-0.006±0.012, p-value<0.003).

Type 2 diabetes and measures of hyperglycemia are associated with hippocampal subfield atrophy, independently of lifestyle and cardiovascular risk factors. We found evidence for a dose-response relationship from normal glucose metabolism, to prediabetes, to type 2 diabetes. Prediabetes stages could give a window of opportunity for the early prevention of brain disease.

Introduction

There is extensive evidence that type 2 diabetes is associated with an increased risk of both degenerative and vascular brain damage (Brundel, Kappelle, & Biessels, 2014; Zheng, Ley, & Hu, 2018), as well as with memory impairment (Callisaya et al., 2019; Sadanand, Balachandar, & Bharath, 2016). Given the involvement of the hippocampus in memory processing, the relation between type 2 diabetes and hippocampal atrophy has been widely studied. Most brain MRI studies have indeed detected an association between type 2 diabetes and smaller bilateral hippocampal volume (Cui et al., 2019; Gold et al., 2007; Hempel, Onopa, & Convit, 2012; Moran et al., 2013). Although there are exceptions (Wisse et al., 2014), a later meta-analysis based on 1,364 cases and 3,433 controls confirmed the association between type 2 diabetes and smaller hippocampal volumes (Moulton, Costafreda, Horton, Ismail, & Fu, 2015).

The hippocampus is a heterogeneous structure, composed of multiple subfields, each of which is characterized by specific cellular composition and distinctive neurophysiology (Fanselow & Dong, 2010). Therefore, type 2 diabetes pathophysiology may be differently associated with specific hippocampal subfields. Yet, there is little theoretical agreement on the hippocampal subfields that might be affected in type 2 diabetes, and previous studies found smaller volumes in different subfields (Blom et al., 2020; C. Li et al., 2020; M. Li et al., 2020; Zhang et al., 2021).

Another very relevant question to the disease course of type 2 diabetes is whether prediabetes (i.e., the intermediate hyperglycemic condition in the transition from normal glucose metabolism to type 2 diabetes) is also associated with smaller hippocampal volumes. Previous literature found no evidence of an association between prediabetes and total hippocampal volume (THV) (Marseglia et al., 2019; Schneider et al., 2017). However, whether there is an association between prediabetes and specific hippocampal subfields is still unknown.

The aim of the current study is to investigate whether prediabetes, type 2 diabetes, and continuous measures of hyperglycemia are associated with lower hippocampal subfield volumes. In addition, we aim to investigate whether potential associations are independent of demographic, lifestyle, and cardiovascular risk factors. To our knowledge, no previous study addressed the association between prediabetes, type 2 diabetes, and hippocampus subfield volumes in a population-based cohort, and taking into account potential confounders.

Material and methods

Study population and design

We used data from The Maastricht Study, an observational population-based cohort study. The rationale and methodology have been described previously (Schram et al., 2014). In brief, the study focuses on the etiology, pathophysiology, complications and comorbidities of type 2 diabetes, and is characterized by an extensive phenotyping approach. Eligible for participation were all individuals aged between 40 and 75 years and living in the southern part of the Netherlands. Participants were recruited through mass media campaigns, the municipal registries, and the regional Diabetes Patient Registry via mailings. Recruitment was stratified according to known type 2 diabetes status, with an oversampling of individuals with type 2 diabetes, for reasons of efficiency. The present report includes cross-sectional data from 7,689 participants, who completed the baseline survey between November 2010 and December 2017. The examinations of each participant were performed within a time window of three months. MRI measurements were implemented from December 2013 onwards until February 2017 and were available in 5,204 out of 7,689 participants. Additionally, 451 MRI scans had insufficient segmentation quality (Monereo-Sánchez et al., 2021). Participants with type 1 diabetes or other types of diabetes (n=29) were excluded from the analysis. In the remaining 4,724 participants, complete data on covariates was available in 4,636 participants (Figure 3.1). The study has been approved by the institutional medical ethical committee (NL31329.068.10) and the Minister of Health, Welfare and Sports of the Netherlands (Permit 131088-105234-PG). All participants gave written informed consent.
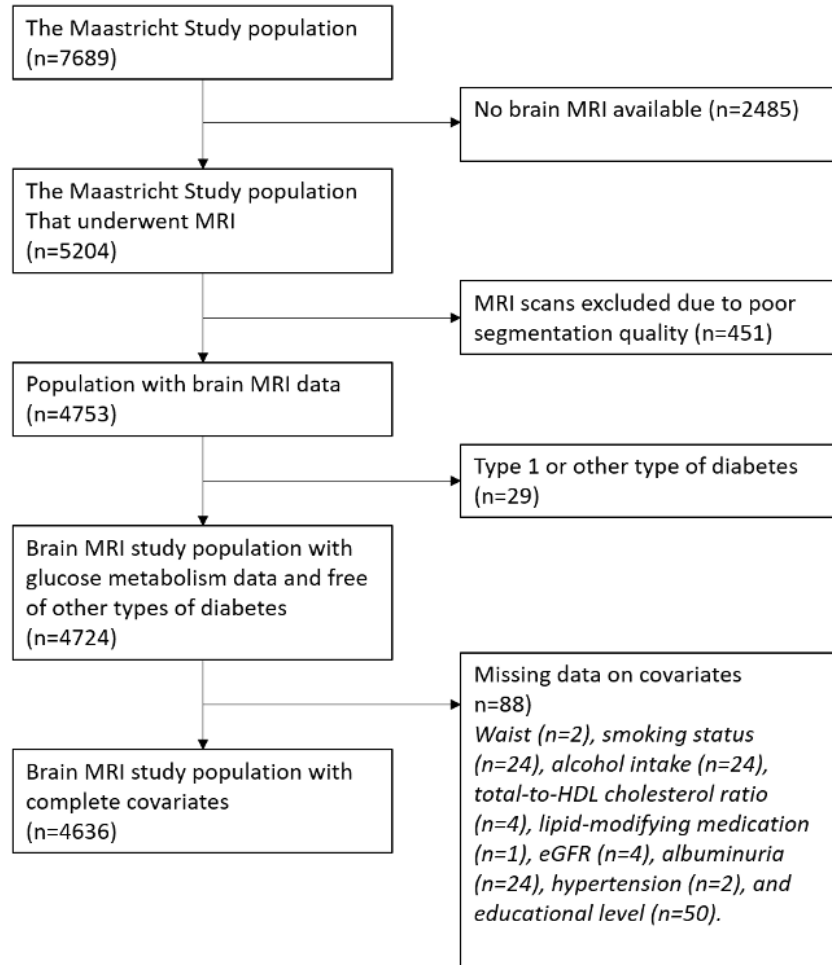
Figure 3.1 | Flowchart of the study population.

Glucose metabolism status

To determine glucose metabolism status, all participants, except those who used insulin, underwent a standardized 2h 75g oral glucose tolerance test (OGTT) after an overnight fast. For safety reasons, participants with a fasting plasma glucose level above 11.0 mmol/l, as determined by a finger prick, did not undergo the OGTT (n=50). For these individuals, fasting plasma glucose level and information about diabetes medication use were used to determine glucose metabolism status. Glucose metabolism status was defined according to the World Health Organization 2006 criteria as normal glucose metabolism (fasting plasma glucose < 6.1 mmol/l), prediabetes (fasting plasma glucose ≥ 6.1 and < 7.0 mmol/l) or type 2 diabetes (fasting plasma glucose ≥7.0mmol/l, or the use of diabetes medication) (Schram et al., 2014).

Brain imaging

Brain images were acquired on a 3T magnetic resonance scanner (MAGNETOM Prismafit, Siemens Healthineers GmbH) located at a dedicated scanning facility (Scannexus, Maastricht, The Netherlands) using a head/neck coil with 64 elements for parallel imaging. The MRI protocol included a three-dimensional (3D) T1-weighted (T1w) magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence (repetition time/inversion time/echo time (TR/TI/TE) 2,300/900/2.98ms, 176 slices, 256 × 240 matrix size, 1.0 mm cubic reconstructed voxel size); and a fluid-attenuated inversion recovery (FLAIR) sequence (TR/TI/TE 5,000/1,800/394 ms, 176 slices, 512 × 512 matrix size, 0.49 × 0.49 × 1.0 mm reconstructed voxel size).

Brain segmentation was performed with FreeSurfer v6.0 (Fischl, 2012) using T1w and FLAIR images as input. The optional arguments "FLAIRpial" and "3T" were used to optimize segmentation quality. Brain segmentations with insufficient quality, i.e. Euler numbers below 1.5 quartile (-80 for left hemisphere and -68 for right hemisphere) were excluded (Monereo-Sánchez et al., 2021). Hippocampal subfields were segmented (J. E. Iglesias et al., 2015), yielding a THV and 12 regions of interest per hemisphere. Subfields name and description can be found in Supplementary Table S3.1. Hippocampal subfields volumes were averaged between the left and right hemisphere of each participant. Supplementary Table S3.1 shows the mean volume and standard deviation of each subfield across the reference sample (n=4724), which were used for z-transformation prior to statistical analysis. Results are depicted in hippocampal subfield maps, a legend for these maps can be found in Supplementary Figure S3.1.

General characteristics and covariates

As described elsewhere (Schram et al., 2014), educational level (low, intermediate, high), alcohol intake, smoking status (never, current, former) and history of cardiovascular disease were assessed by questionnaires. Medication use was assessed in a medication interview where generic name, dose, and frequency were registered. We measured weight, height, body mass index, waist circumference, office and ambulatory 24h blood pressure, plasma glucose levels, serum creatinine, 24h urinary albumin excretion (twice), hemoglobin A1c (HbA1c), and plasma lipid profile as described elsewhere (Schram et al., 2014). Estimated glomerular filtration rate (in ml/min/1.73 m2) was calculated with the Chronic Kidney Disease Epidemiology Collaboration equation based on both serum creatinine and serum cystatin C (Inker et al., 2012).

Statistical analysis

All statistical analyses were performed by use of R 4.0.2 statistical software (2020-06-22). General characteristics of the study population were presented as mean with standard deviation, or as percentages, and were evaluated by T-tests or χ^2 tests.

We used multiple linear regression analysis to investigate the association of prediabetes, type 2 diabetes, HbA1c, fasting plasma glucose, and 2h post-load glucose levels with THV and hippocampal subfield volumes. Analyses were performed for THV and each hippocampal subfield (n=13 brain volume estimates). Analyses were adjusted for age, sex, total intracranial volume and the time between the baseline and MRI measurement, waist circumference, smoking status, alcohol intake, total-to-HDL cholesterol ratio, lipid-modifying medication, eGFR, albuminuria, hypertension, and educational level. Given THV and 12 subfields were analyzed, and to maintain a type I error rate of 5%, Matrix Spectral Decomposition (Nyholt, 2004) was used to determine the effective number of independent variables. Based on the resulting eigenvalues, the obtained effective number was n=7, therefore alpha threshold for significance was set at 0.05/7= 0.0071.

In post-hoc analysis, we tested for a linear trend using the ordinal variable of glucose metabolism status (normal glucose metabolism=1, prediabetes=2, and type 2 diabetes=3). This analysis was justified after checking the model fit with the main model (glucose metabolism status as a categorical measure). Comparison of the log likelihood ratio's showed a better fit with the ordinal variable, which indicates a dose-response relationship between glucose metabolism status and the hippocampal subfield volumes.

Several additional analyses were performed to check for robustness. To study whether the associations found in continuous measures of hyperglycemia (i.e. HbA1c, fasting plasma glucose, or 2h post-load glucose levels) were driven by the oversampling of individuals with diagnosed type 2 diabetes, we additionally excluded participants with type 2 diabetes from the analyses. To address whether the hippocampal volume differences are independent from general brain atrophy, we replaced total intracranial volume with total brain volume. We also replaced waist circumference with BMI, and total-to-HDL cholesterol ratio for LDL cholesterol level. Additionally, we report the results without the exclusion of cases with insufficient quality segmentation based on Euler numbers. Finally, an interaction term was incorporated to test for interaction among prediabetes, type 2 diabetes, and continuous measures of hyperglycemia and sex, on hippocampal subfield volumes.

Results

General characteristics of the study population

General characteristics of the study population, stratified by glucose metabolism status, are shown in Table 3.1. The study population consisted of 4724 participants; 3184 participants had normal glucose metabolism, 671 had prediabetes, and 869 had type 2 diabetes. The mean age was 58.7 ± 8.5 years, and 51.5% were female. Participants with prediabetes and type 2 diabetes were older, less often female, had a worse cardiovascular risk profile, were more often current smokers, and more often had a low educational level (Table 3.1). Mean subfields volumes can be found in Supplementary Table S3.1. Individuals who underwent MRI were younger, were less likely to have type 2 diabetes, were less often current smokers and less often had a low educational level, as compared to the study population which did not undergo MRI (Supplementary Table S3.2).

Characteristic	Normal glucose metabolism (n=3184)	Prediabetes (n=671)	Type 2 diabetes (n=869)	P-value
Demographics				
Age (years)	57.3 ± 8.5	61.3 ± 7.9	61.6 ± 7.9	<0.001
Sex (% female)	56.8	46.1	36.0	<0.001
Education level, low/medium/high (%)	26.6/28.5/44.9	36.4/27.7/35.8	41.2/28.9/29.9	<0.001
Glucose metabolism				
Fasting glucose (mmol/l)	5.1 ± 0.4	5.8 ± 0.6	7.7 ± 1.8	<0.001
2h post-load glucose (mmol/l)	5.3 ± 1.1	8.2 ± 1.7	14.2 ± 4.0	<0.001
HbA1c (mmol/mol)	35.1 ± 3.9	37.8 ± 4.4	50.0 ± 10.9	<0.001
Cardiovascular risk factors				
Waist circumference (cm)	89.7 ± 11.0	96.9 ± 12.0	103.0 ± 12.5	<0.001
Office systolic blood pressure (mmHg)	129.2 ± 16.2	135.4 ± 16.5	140.1 ± 16.9	<0.001
Office diastolic blood pressure (mmHg)	74.7 ± 9.7	76.6 ± 9.6	77.2 ± 9.4	<0.001
Hypertension (%)	36.8	59.6	77.8	<0.001
LDL (mmol/l)	3.3 ± 0.9	3.3 ± 1.0	2.5 ± 0.9	<0.001
Total-to-HDL cholesterol	3.5 ± 1.1	3.9 ± 1.3	3.7 ± 1.2	<0.001
eGFR (ml/min/1.73m2)	77.5 ± 12.9	77.1 ± 14.1	81.2 ± 18.8	<0.001
History of CVD (%)	9.5	12.7	17.5	0.654
Albuminuria, micro/macroalbuminuria (%)	3.9/0.3	5.4/0.3	15.5/1.2	<0.001
Medication use				
Antihypertensive medication (%)	19.6	39.8	64.3	<0.001
Lipid-modifying medication (%)	13.4	28.2	67.5	<0.001
Life style factors				
Smoking, never/former/current (%)	42.5/45.8/11.7	34.9/53.4/11.7	35.2/50.8/14.00	0.0106
Alcohol intake, none/low/high (%)	14.5/60.2/25.3	17.0/56.7/26.3	26.0/56.0/17.9	<0.001
Brain MRI characteristics				
Estimated total intracranial volume (mm^3)	1472257.1 ± 148293.9	1455241.6 ± 138958.9	1455750.8 ± 138233.5	<0.001
Brain volume (mm^3)	1191747.4 ± 116151.6	1171788.1 ± 113658.1	1160039.5 ± 111192.3	<0.001
MRI lag time (years)	1.2 ± 1.3	1.3 ± 1.3	1.3 ± 1.3	<0.001

Table 3.1 | General characteristics of the study population. Data are presented as mean ± standard deviation or percentage, and stratified for glucose metabolism status, i.e. normal glucose metabolism, prediabetes and type 2 diabetes. HbA1c indicates hemoglobin A1c; LDL: low-density lipoprotein; HDL, high-density lipoprotein; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease.

Associations of prediabetes and type 2 diabetes with hippocampal volume

Figure 3.2 shows the resulting associations of THV and hippocampal subfields with prediabetes and type 2 diabetes. Detailed results can be found at Supplementary Table S3.3. We found no direct significant associations between prediabetes and hippocampal subfield volumes after correction for multiple comparison.
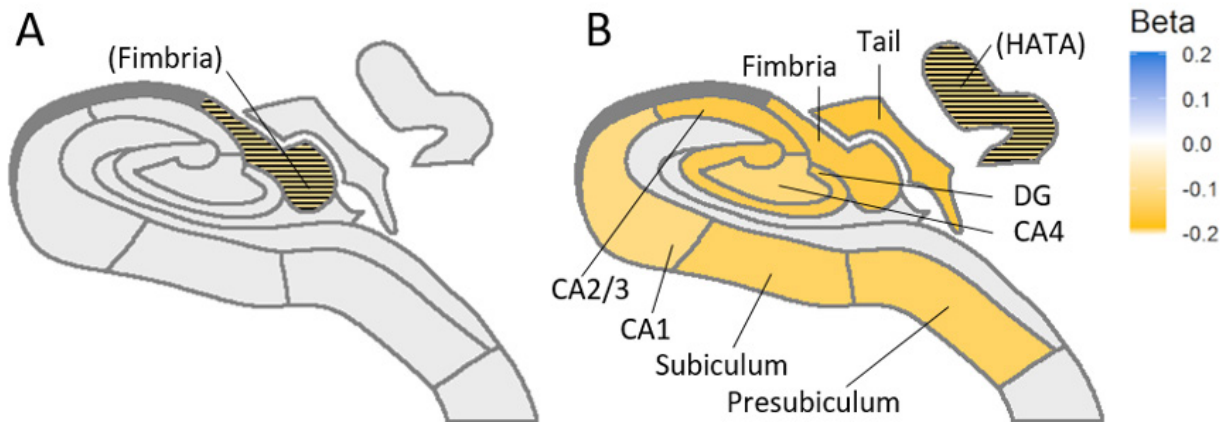
Figure 3.2 | Schematic hippocampal representation displaying the associations of A) Prediabetes and B) **Type 2 diabetes with hippocampal subfield volumes***. Yellow color represents a negative association, i.e. Type 2 diabetes is associated with smaller subfields volume. Associations that did not survive multiple comparison correction but have P<.05 are represented with a stipe pattern and are written in parenthesis. Non-significant associations are represented in gray. See Supplementary Figure S3.2 for a legend of the hippocampal map. Abbreviations: CA: Cornu Ammonis; DG: Dentate gyrus; HATA: Hippocampal-amygdalar transition area.*

After full adjustment, type 2 diabetes was significantly associated with smaller volumes in THV (β±SE = -0.151±0.04, p-value <0.001) as compared to normal glucose metabolism. In addition, type 2 diabetes was significantly associated with smaller volumes in the hippocampal fimbria (β±SE= -0.195±0.04, p-value <0.001), hippocampus proper, i.e. dentate gyrus, CA1, CA2/3, CA4, subiculum and presubiculum (β±SE< -0.105±0.04, p-value <0.006); as well as subfield tail (β±SE= -0.162±0.04, p-value <0.001). Further, type 2 diabetes was associated also with the hippocampal-amygdala transition area (HATA, β±SE= -0.098 ± 0.04, p-value= 0.015), although this was no longer significant after correction for multiple comparison.

In linear trend analysis there was a significant association between glucose metabolism status and smaller THV, fimbria, Cornu Ammonis (CA) 2/3, CA4, dentate gyrus, subiculum, presubiculum, and tail volumes (standardized beta coefficient ± standard error (β±SE) < -0.054 ± 0.02, p-for-trend < 0.003, Supplementary Table S3.4). These results indicate a dose-response relation from normal glucose metabolism, to prediabetes, to type 2 diabetes with lower hippocampal subfield volumes.

Associations of continuous measures of hyperglycemia with hippocampal volume

HbA1c, fasting plasma glucose, and 2h post-load glucose levels were associated with smaller volumes of the total hippocampus (β±SE< -0.005±0.01, p-value <0.002), fimbria (β±SE< -0.010±0.011, p-value <0.001), and tail (β±SE< -0.006±0.012, p-value <0.003) after full adjustment. Fasting plasma glucose, and 2h post-load glucose levels were additionally associated with lower volumes of dentate gyrus (β±SE< -0.013±0.010, p-value <0.002),

and CA3 (β±SE< -0.014±0.011, p-value <0.001). 2h Post-load glucose level was associated with lower volumes of the subiculum (β±SE= -0.011±0.004, p-value=0.005). Results are depicted in Figure 3.3. Detailed results can be found in Supplementary Table S3.5.
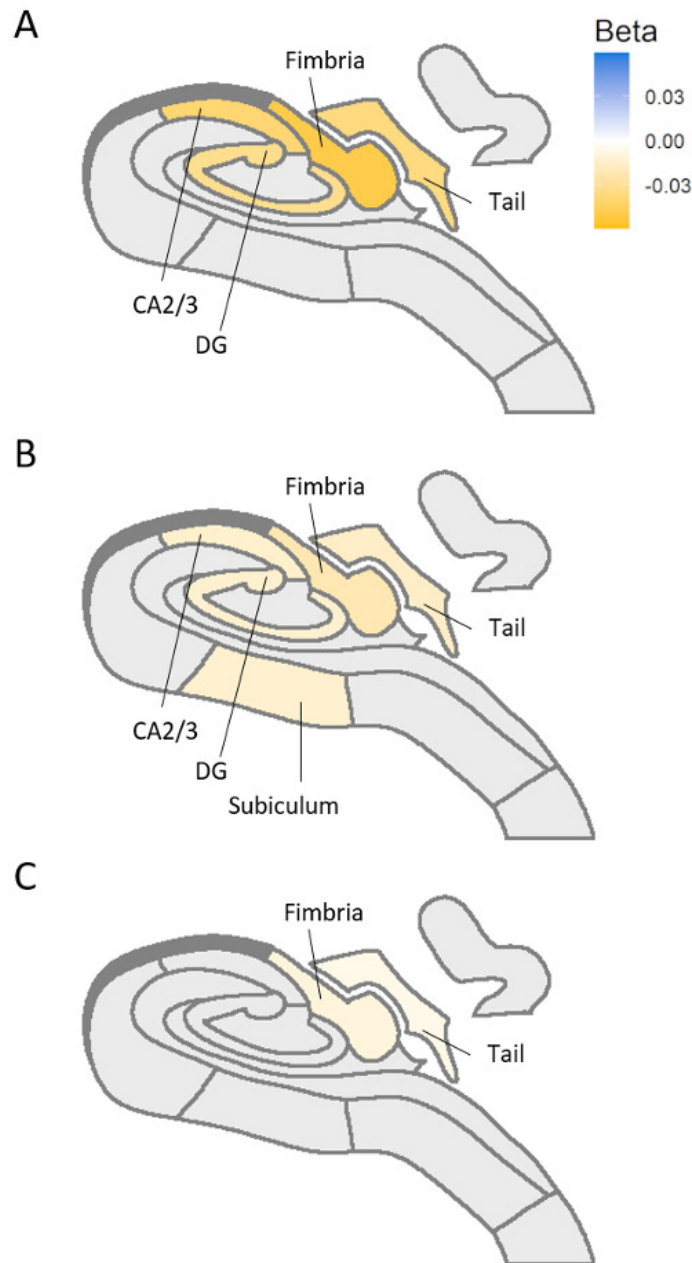
Figure 3.3 | Schematic hippocampal representation displaying the significant subfields' associations with continuous measures of hyperglycemia: A) Fasting glucose (mmol/l), B) 2h post-load glucose (mmol/l) and C) HbA1c (mmol/mol). *Yellow color represents a negative association; i.e. higher values of hyperglycemia are associated with smaller hippocampal volumes. Only those structures significant after multiple comparison correction are depicted. Hippocampal subfields with no significant associations are represented in gray. See Supplementary Figure S3.2 for a hippocampal map legend. Abbreviations: CA: Cornu Ammonis; DG: Dentate gyrus.*

Additional analyses

When we limited the analysis to participants with normal glucose metabolism and prediabetes (i.e., when we excluded participants with type 2 diabetes), the significant associations between continuous measures of hyperglycemia and hippocampal subfields previously found showed numerically similar trends for the estimates, although most associations became not significant. Yet, the association between 2h/post load glucose and smaller volumes in fimbria remained significant (Supplementary Table S3.6).

When we replaced estimated intracranial volume with total brain volume, the associations between type 2 diabetes and hippocampal subfields showed similar estimates, although they became not significant, with the exception of fimbria (Supplementary Table S3.7).

When we replaced body mass index with waist circumference the results remained consistent (Supplementary Table S3.8).

When we replaced total-to-HDL cholesterol ratio with LDL cholesterol level, the associations between type 2 diabetes and hippocampal subfields showed similar estimates. Although some estimates attenuated, the associations remained statistically significant (Supplementary Table S3.9).

The associations between prediabetes and type 2 diabetes and hippocampal subfield volumes when including the cases with low segmentation accuracy based on Euler numbers, remained consistent, and became even slightly stronger (Supplementary Table S3.10).

Discussion

In this study we evaluated the association of prediabetes and type 2 diabetes, as well as continuous measures of hyperglycemia, with hippocampal subfields volumes after adjustment for demographics, lifestyle, and cardiovascular risk factors. Through this large population-based study we confirmed that type 2 diabetes is associated with smaller THV, and smaller volumes in the subfields fimbria, dentate gyrus, CA1 to CA4, subiculum, presubiculum, and tail. We found no significant associations of prediabetes with THV or hippocampal subfield volumes. However, the results for continuous measures of hyperglycemia and the analysis for trend on glucose metabolism status suggest there is a dose-response association between glucose metabolism and lower hippocampal subfield volumes.

Consistent with previous literature, our results show that type 2 diabetes is associated with smaller volumes in dentate gyrus CA1, CA4, and subiculum (Blom et al., 2020; M. Li et al., 2020; Zhang et al., 2021). To our knowledge, we are the first to additionally find an association of type 2 diabetes with smaller volumes in fimbria, CA2/3, and tail volumes. The standardized effect sizes of this associations ranged between -0.105 in

CA1, and -0.195 in fimbria. In volume, this translates into CA1 being in average 4.22 mm^3 smaller, and fimbria being 8.98mm^3 smaller in participants with type 2 diabetes when compared to those with normal glucose metabolism status. Taking into consideration the mean volume of this structures, CA1 was 1.22% smaller, and fimbria were 5.3% smaller in participants with type 2 diabetes when compared to those with normal glucose metabolism. We also found a significant association between type 2 diabetes and smaller THV, which is in line with a meta-analysis by Moulton et.al. in 2015 (Moulton et al., 2015). Type 2 diabetes status was most strongly associated with fimbria volume, which also remained associated after correction for general atrophy of the brain. The fimbria forms a white matter bundle structure that connects the hippocampus with the rest of the brain. Smaller volumes in this specific structure could be due a loss in myelin, since white matter microstructural abnormalities (Reijmer et al., 2013) and demyelination (Li et al., 2021) has been commonly found in participants with type 2 diabetes.

In line with previous studies (Marseglia et al., 2019; Schneider et al., 2017), prediabetes showed no significant associations with hippocampal subfield volumes. However, the analysis for trend demonstrated a dose-response relationship between glucose metabolism status and lower volumes of most hippocampal subfields. This suggest there is a graded association from normal glucose metabolism to prediabetes to type 2 diabetes. Further evidence for a linear association is provided by the continuous measures of hyperglycemia, i.e., fasting plasma glucose, 2 hours post-load glucose, and HbA1c levels. Specifically, continuous measures of hyperglycemia were associated with smaller volumes in THV, fimbria, dentate gyrus, CA2/3, and tail. HbA1c shows the least strong associations among the three measures, likely because it is a treatment target in type 2 diabetes. A study by Dong et.al. (2019) showed associations of HbA1c with smaller volumes in dentate gyrus subiculum, and tail (Dong et al., 2019), while another study found associations with the CA1 to CA4 (Zhang et al., 2021), and dentate gyrus (Dong et al., 2019; Zhang et al., 2021). Yet, those studies observed an association of HbA1c with the molecular layer that was not detected in our study. It is important to notice that these studies used selected study populations and small sample sizes (Dong et al., 2019) (Zhang et al., 2021).

To assess robustness of the observed associations we performed a range of sensitivity analyses. First, to ensure that the observed associations would not be driven by participants with type 2 diabetes, we repeated the analysis after excluding participants with type 2 diabetes. Results were consistent with the main analysis, although regression coefficients were generally attenuated due to reduced sample size (from n=4724 to n=3855). However, two hours post-load glucose remained significantly associated with fimbria volume, indicating that fimbria could be one of the most sensitive subfields to hyperglycemia, and therefore bringing evidence towards subfields specificity. Second, we corrected for total brain volume instead of total intracranial volume. With this analysis we found that type 2 diabetes is associated with a smaller volume of the subfield fimbria, independent of generalized atrophy of the brain. This indicates that this subfield is affected by hyperglycemia on a higher degree than the rest of the parenchyma. Quality control of the hippocampal segmentation was provided by the exclusion of cases with outliers based on Euler numbers following guidelines(Monereo-Sánchez et al., 2021). We performed a sensitivity analysis without the exclusion of these cases. Including cases with low segmentation quality to the sample resulted in decreased p-values. This could be explained by the increase in

sample size, but it might also be due to the fact that least healthy participants tend to have worse scans and poorer segmentations, with missing parenchyma in the segmentations and therefore smaller volumes.

Our results may show some specificity for subfields, as some associations with type 2 diabetes were stronger than others. In addition, several analyses using both continuous and categorical definitions of glycaemia, as well as all the sensitivity analysis allow us to detect some subfields that might have a more severe or earlier vulnerability to hyperglycemia. Both continuous measures of glycaemia as well as type 2 diabetes are consistently associated with fimbria, dentate gyrus, CA2/3, subiculum, and tail subfield volumes. These exact subfields also show increased regional vulnerability to age (Pereira et al., 2014), which agrees with the hypothesis that type 2 diabetes can be considered accelerated aging.

Strengths of this study include the large sample size and population-based design with an oversampling of type 2 diabetes; the use of oral glucose tolerance tests to accurately characterize glucose metabolism status; and the extensive phenotyping which allowed us to adjust for major cardiovascular risk factors reducing the change of residual cofounding.

This study has some limitations. The hippocampus is a small structure, and the segmentation of the hippocampal subfields is challenging and can be subject to inaccuracies. Yet, the hippocampal volumes were extracted using FreeSurfer v6.0, which shows a good manual segmentation agreement and test-retest reliability (Tae, Kim, Lee, Nam, & Kim, 2008). We additionally improved the segmentation accuracy by adding FLAIR images for Multispectral segmentation (J. Iglesias, 2020) improving segmentation reliability (J. E. Iglesias et al., 2015; Seiger et al., 2021); finally we performed quality control by the exclusion of outliers based on Euler numbers following current recommendations (Monereo-Sánchez et al., 2021). In addition, due to the population-based nature of our cohort, our results may be subject to some selection bias because participants of cohort studies are in general more health conscious. Previous research show that this can result in over- or under-estimations of the associations under investigation (Szklo & Nieto, 2014). Further, it is important to notice that the cross-sectional design of our study does not allow to claim any causality. Yet, longitudinal studies have previously shown increased rates of brain atrophy (Kooistra et al., 2013; Samaras et al., 2014) and brain function impairment (Thambisetty et al., 2013) over time in participants with impaired glucose metabolism. This suggest that glucose metabolism might affect hippocampal volumes, but future research on longitudinal data is needed to specifically address this question.

In conclusion, type 2 diabetes was associated with generalized hippocampal atrophy, which was independent of demographics, cardiovascular, and lifestyle risk factors. The fimbria is the subfield that shows the strongest association with type 2 diabetes. Continuous measures of hyperglycemia, and analysis for trend indicate that the association between hyperglycemia and hippocampal subfields volumes is linear, and follows a dose-response curve, although we could not demonstrate significant associations of prediabetes with hippocampal subfield volumes. The latter could mean prediabetes stages represent a window of action for the early prevention of brain disease.

References

Blom, K., Koek, H. L., Zwartbol, M. H., Ghaznawi, R., Kuijf, H. J., Witkamp, T. D., . . . Group, U.-S. S. (2020). Vascular Risk Factors of Hippocampal Subfield Volumes in Persons without Dementia: The Medea 7T Study. *Journal of Alzheimer's Disease*(Preprint), 1-17.

Brundel, M., Kappelle, L. J., & Biessels, G. J. (2014). Brain imaging in type 2 diabetes. *European Neuropsychopharmacology, 24*(12), 1967-1981.

Callisaya, M. L., Beare, R., Moran, C., Phan, T., Wang, W., & Srikanth, V. K. (2019). Type 2 diabetes mellitus, brain atrophy and cognitive decline in older people: a longitudinal study. *Diabetologia, 62*(3), 448-458.

Cui, D., Liu, X., Liu, M., Cao, W., Xue, Y., Guo, Y., . . . Jiao, Q. (2019). Subcortical gray matter structural alterations in prediabetes and type 2 diabetes. *Neuroreport, 30*(6), 441-445.

Dong, S., Dongwei, L., Zhang, J., Liang, J., Sun, Z., & Fang, J. (2019). Individuals in the prediabetes stage exhibit reduced hippocampal tail volume and executive dysfunction. *Brain and behavior, 9*(8), e01351.

Fanselow, M. S., & Dong, H.-W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron, 65*(1), 7-19.

Fischl, B. (2012). FreeSurfer. *NeuroImage, 62*(2), 774-781.

Gold, S., Dziobek, I., Sweat, V., Tirsi, A., Rogers, K., Bruehl, H., . . . Convit, A. (2007). Hippocampal damage and memory impairments as possible early brain complications of type 2 diabetes. *Diabetologia, 50*(4), 711-719.

Hempel, R., Onopa, R., & Convit, A. (2012). Type 2 diabetes affects hippocampus volume differentially in men and women. *Diabetes/metabolism research and reviews, 28*(1), 76-83.

Iglesias, J. (2020, 2020-05-08). Hippocampal Subfields. Retrieved from https://surfer.nmr.mgh.harvard.edu/fswiki/HippocampalSubfields

Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., . . . Wald, L. L. (2015). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. *NeuroImage, 115*, 117-137.

Inker, L. A., Schmid, C. H., Tighiouart, H., Eckfeldt, J. H., Feldman, H. I., Greene, T., . . . Zhang, Y. L. (2012). Estimating glomerular filtration rate from serum creatinine and cystatin C. *New England journal of medicine, 367*(1), 20-29.

Kooistra, M., Geerlings, M. I., Mali, W. P., Vincken, K. L., van der Graaf, Y., Biessels, G. J., & Group, S.-M. S. (2013). Diabetes mellitus and progression of vascular brain lesions and brain atrophy in patients with symptomatic atherosclerotic disease. The SMART-MR study. *Journal of the neurological sciences, 332*(1-2), 69-74.

Li, C., Zuo, Z., Liu, D., Jiang, R., Li, Y., Li, H., . . . Xiong, K. (2020). Type 2 Diabetes Mellitus May Exacerbate Gray Matter Atrophy in Patients With Early-Onset Mild Cognitive Impairment. *Frontiers in neuroscience, 14*, 856.

Li, M., Huang, L., Yang, D., Luo, C., Qin, R., Zhang, B., . . . Xu, Y. (2020). Atrophy patterns of hippocampal subfields in T2DM patients with cognitive impairment. *Endocrine, 68*(3), 536.

Li, R., Wang, B., Wu, C., Li, D., Wu, Y., Ye, L., . . . Yuan, Y. (2021). Acidic fibroblast growth factor attenuates type 2 diabetes-induced demyelination via suppressing oxidative stress damage. *Cell death & disease, 12*(1), 1-17.

Marseglia, A., Fratiglioni, L., Kalpouzos, G., Wang, R., Bäckman, L., & Xu, W. (2019). Prediabetes and diabetes accelerate cognitive decline and predict microvascular lesions: a population-based cohort study. *Alzheimer's & Dementia, 15*(1), 25-33.

Monereo-Sánchez, J., de Jong, J. J., Drenthen, G. S., Beran, M., Backes, W. H., Stehouwer, C. D., . . . Jansen, J. F. (2021). Quality control strategies for brain MRI segmentation and parcellation: Practical approaches and recommendations-insights from the Maastricht study. *Neuroimage, 237*, 118174.

Moran, C., Phan, T. G., Chen, J., Blizzard, L., Beare, R., Venn, A., . . . Greenaway, T. M. (2013). Brain atrophy in type 2 diabetes: regional distribution and influence on cognition. *Diabetes care, 36*(12), 4036-4042.

Moulton, C. D., Costafreda, S. G., Horton, P., Ismail, K., & Fu, C. H. (2015). Meta-analyses of structural regional cerebral effects in type 1 and type 2 diabetes. *Brain imaging and behavior, 9*(4), 651-662.

Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics, 74*(4), 765-769.

Pereira, J. B., Valls-Pedret, C., Ros, E., Palacios, E., Falcón, C., Bargalló, N., . . . Junque, C. (2014). Regional vulnerability of hippocampal subfields to aging measured by structural and diffusion MRI. *Hippocampus, 24*(4), 403-414.

Reijmer, Y. D., Brundel, M., De Bresser, J., Kappelle, L. J., Leemans, A., Biessels, G. J., & Group, U. V. C. I. S. (2013). Microstructural white matter abnormalities and cognitive functioning in type 2 diabetes: a diffusion tensor imaging study. *Diabetes care, 36*(1), 137-144.

Sadanand, S., Balachandar, R., & Bharath, S. (2016). Memory and executive functions in persons with type 2 diabetes: a meta-analysis. *Diabetes/metabolism research and reviews, 32*(2), 132-142.

Samaras, K., Lutgers, H. L., Kochan, N. A., Crawford, J. D., Campbell, L. V., Wen, W., . . . Brodaty, H. (2014). The impact of glucose disorders on cognition and brain volumes in the elderly: the Sydney Memory and Ageing Study. *Age, 36*(2), 977-993.

Schneider, A. L., Selvin, E., Sharrett, A. R., Griswold, M., Coresh, J., Jack Jr, C. R., . . . Gottesman, R. F. (2017). Diabetes, prediabetes, and brain volumes and subclinical cerebrovascular disease on MRI: the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS). *Diabetes care, 40*(11), 1514-1521.

Schram, M. T., Sep, S. J., van der Kallen, C. J., Dagnelie, P. C., Koster, A., Schaper, N., . . . Stehouwer, C. D. (2014). The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *European journal of epidemiology, 29*(6), 439-451.

Seiger, R., Hammerle, F. P., Godbersen, G. M., Reed, M. B., Spurny-Dworak, B., Handschuh, P., . . . Vanicek, T. (2021). Comparison and reliability of hippocampal subfield segmentations within FreeSurfer utilizing T1-and T2-weighted multispectral MRI data. *Frontiers in neuroscience, 15*.

Szklo, M., & Nieto, F. J. (2014). *Epidemiology: beyond the basics*: Jones & Bartlett Publishers.

Tae, W. S., Kim, S. S., Lee, K. U., Nam, E.-C., & Kim, K. W. (2008). Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology, 50*(7), 569.

Thambisetty, M., Beason-Held, L. L., An, Y., Kraut, M., Metter, J., Egan, J., . . . Resnick, S. M. (2013). Impaired glucose tolerance in midlife and longitudinal changes in brain function during aging. *Neurobiology of aging, 34*(10), 2271-2276.

Wisse, L. E., de Bresser, J., Geerlings, M. I., Reijmer, Y. D., Portegies, M. L., Brundel, M., . . . Group, U. D. E. S. (2014). Global brain atrophy but not hippocampal atrophy is related to type 2 diabetes. *Journal of the neurological sciences, 344*(1-2), 32-36.

Zhang, W., Gao, C., Qing, Z., Zhang, Z., Bi, Y., Zeng, W., & Zhang, B. (2021). Hippocampal subfields atrophy contribute more to cognitive impairment in middle-aged patients with type 2 diabetes rather than microvascular lesions. *Acta Diabetologica*, 1-11.

Zheng, Y., Ley, S. H., & Hu, F. B. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nature Reviews Endocrinology, 14*(2), 88-98.

Chapter 3: The association of prediabetes and type 2 diabetes with hippocampal subfields volume: The Maastricht Study

Supplementary Table S3.1

Abbreviation	Description	Volume mm³ (mean ± SD)
THV	Total Hippocampal Volume	3422.75 ± 349.78
HATA	Hippocampus-amygdala-transition-area	61.73 ± 9.60
Fimbria	Fimbria	75.30 ± 21.64
Fissure	Hippocampal fissure	81.42 ± 14.73
Molecular layer	Molecular layer of the hippocampus	361.43 ± 45.95
Dentate gyrus	Granule cell and molecular cell layer of the dentate gyrus	333.67 ± 38.98
CA4	Cornu Ammonis 4	275.04 ± 30.15
CA2/3	Cornu Ammonis 2 and 3	240.05 ± 36.75
CA1	Cornu Ammonis 1	724.29 ± 85.54
Subiculum	Subiculum	469.41 ± 50.82
Presubiculum	Presubiculum	298.92 ± 36.79
Parasubiculum	Parasubiculum	61.40 ± 11.17
Tail	Hippocampal tail	521.53 ± 67.50

*Supplementary Table S3.1 | **List of hippocampal subfields.** Table shows the abbreviation and complete name for total hippocampal volume and each hippocampal subfield provided as output by FreeSurfer v.6.0, as well as the mean volume (mm³) ± standard deviation (SD) of the used sample (n=4724).*

Supplementary Figure S3.1
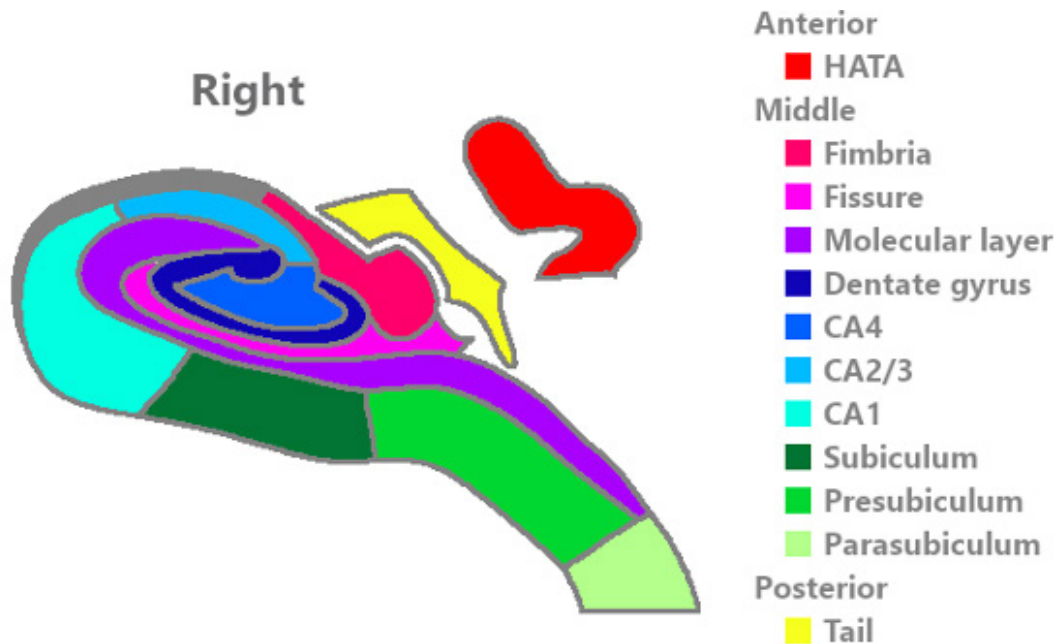
*Supplementary Figure S3.1 | **Hippocampal subfields map.** Figure shows a schematic representation of the right hippocampus in a coronal section. Hippocampal subfields are represented in different colors. Abbreviations: HATA, Hippocampus-amygdala-transition-area; Dentate gyrus, Granule cell and molecular cell layer of the dentate gyrus; CA, Cornu Ammonis. Notice that the subfield HATA is transposed from an anterior coronal section, while the tail is transposed from a posterior coronal section.*

Supplementary Table S3.2

Characteristic	Excluded (n=3053)	Included (n=4636)	P value
Demographics			
Age (years)	61.6 ± 8.6	58.7 ± 8.5	<0.001
Sex (% female)	46.7	51.5	<0.001
Education level, low/medium/high (%)	41.3/26.2/32.5	30.7/28.5/40.9	0.594
Glucose metabolism			
Status, NGM/prediabetes/T2D (%)	48.7/16.1/35.1	67.4/14.2/18.4	0.003
Fasting glucose (mmol/l)	6.3 ± 2	5.7 ± 1.3	<0.001
2h post-load glucose (mmol/l)	8.5 ± 4.4	7.1 ± 3.7	<0.001
HbA1c (mmol/mol)	42.3 ± 11.3	38.3 ± 8.2	<0.001
Cardiovascular risk factors			
Waist circumference (cm)	98.8 ± 14.7	93.2 ± 12.6	<0.001
Office systolic blood pressure (mmHg)	136.5 ± 19.1	132 ± 16.9	<0.001
Office diastolic blood pressure (mmHg)	75.8 ± 10	75.4 ± 9.7	0.128
Hypertension (%)	65.4	47.6	0.128
Total-to-HDL cholesterol	3.7 ± 1.2	3.6 ± 1.2	0.015
eGFR (ml/min/1.73m2)	77.9 ± 17.6	78.1 ± 14.4	0.594
History of CVD (%)	26.2	11.5	0.594
Albuminuria, micro/macroalbuminuria (%)	10.3/1.2	6.2/0.4	0.594
Medication use			
Antihypertensive medication (%)	50.2	30.7	0.594
Lipid-modifying medication (%)	44	25.4	0.594
Life style factors			
Smoking, never/former/current (%)	32.5/51.8/15.6	40.1/47.8/12.1	0.594
Alcohol intake, none/low/high (%)	21.1/56.8/22.2	17/59/24.1	0.594
Brain MRI characteristics			
Estimated total intracranial volume (mm3)	1484492.6 ± 175461.1	1466803.8 ± 145372.4	0.033
Brain volume (mm3)	1165349.8 ± 120357.3	1183079.6 ± 115600.2	0.002
MRI lag time (years)	1.4 ± 1.4	1.2 ± 1.3	0.003

Supplementary Table S3.2 | General characteristics of the participants stratified for study inclusion. Data are presented as means ± standard deviation or percentage, and stratified for availability of MRI data. Abbreviations: NGM, normal glucose metabolism; T2D, type 2 diabetes; HbA1c, Hemoglobin A1c; HDL, high-density lipoprotein; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease.

Supplementary Table S3.3

	Prediabetes		Type 2 diabetes	
Structure	**β±SE**	**p value**	**β±SE**	**p value**
THV	-0.036 ± 0.034	0.291	**-0.151 ± 0.036**	**<0.001**
HATA	-0.041 ± 0.038	0.279	-0.098 ± 0.040	0.015
Fimbria	-0.085 ± 0.037	0.024	**-0.195 ± 0.040**	**<0.001**
Fissure	-0.004 ± 0.040	0.923	0.026 ± 0.043	0.546
Molecular layer	0.024 ± 0.038	0.533	-0.040 ± 0.040	0.316
Dentate gyrus	-0.046 ± 0.034	0.179	**-0.145 ± 0.037**	**<0.001**
CA4	-0.039 ± 0.035	0.266	**-0.113 ± 0.037**	**0.002**
CA2/3	-0.058 ± 0.039	0.135	**-0.155 ± 0.041**	**<0.001**
CA1	0.004 ± 0.036	0.916	**-0.105 ± 0.038**	**0.006**
Subiculum	-0.021 ± 0.038	0.579	**-0.134 ± 0.041**	**0.001**
Presubiculum	-0.048 ± 0.038	0.209	**-0.118 ± 0.041**	**0.004**
Parasubiculum	0.051 ± 0.040	0.204	-0.005 ± 0.043	0.909
Tail	-0.064 ± 0.040	0.109	**-0.162 ± 0.042**	**<0.001**

*Supplementary Table S3.3 | **Associations of prediabetes and type 2 diabetes with hippocampal subfields.** Results are presented as standardized beta coefficients (β) ± standard error (SE), and p-values. Bold shows multiple comparison correction significant (p<0.0071). Normal glucose metabolism is used as reference group. Abbreviations: THV, Total hippocampal volume; HATA, Hippocampus-amygdala-transition-area; CA, Cornu Ammonis.*

Supplementary Table S3.4

Structure	**β±SE**	**p for trend**
THV	**-0.070 ± 0.017**	**<0.001**
HATA	-0.048 ± 0.019	0.014
Fimbria	**-0.096 ± 0.019**	**<0.001**
Fissure	0.011 ± 0.021	0.605
Molecular layer	-0.014 ± 0.019	0.459
Dentate gyrus	**-0.069 ± 0.018**	**<0.001**
CA4	**-0.054 ± 0.018**	**0.002**
CA2/3	**-0.075 ± 0.020**	**<0.001**
CA1	-0.045 ± 0.018	0.014
Subiculum	**-0.061 ± 0.020**	**0.002**
Presubiculum	**-0.058 ± 0.020**	**0.003**
Parasubiculum	0.005 ± 0.021	0.826
Tail	**-0.079 ± 0.020**	**<0.001**

*Supplementary Table S3.4 | **Linear associations of glucose metabolism status as a continuum with hippocampal subfield volumes.** Results are presented as standardized beta coefficients (β) ± standard error (SE), and p-values. Bold shows multiple comparison correction significant (p<0.0071). Abbreviations: THV, Total hippocampal volume; HATA, Hippocampus-amygdala-transition-area; CA, Cornu Ammonis.*

Supplementary Table S3.5

Structure	Fasting plasma glucose		2 hour post load glucose		HbA1c mmol/mol	
	β±SE	p value	β±SE	p value	β±SE	p value
THV	**-0.033 ± 0.010**	**0.001**	**-0.014 ± 0.004**	**<0.001**	**-0.005 ± 0.002**	**0.002**
HATA	-0.013 ± 0.011	0.259	-0.006 ± 0.004	0.141	-0.003 ± 0.002	0.160
Fimbria	**-0.050 ± 0.011**	**<0.001**	**-0.022 ± 0.004**	**<0.001**	**-0.010 ± 0.002**	**<0.001**
Fissure	-0.003 ± 0.012	0.820	0.003 ± 0.004	0.557	0.002 ± 0.002	0.328
Molecular layer	-0.011 ± 0.011	0.331	-0.004 ± 0.004	0.361	-0.002 ± 0.002	0.342
Dentate gyrus	**-0.032 ± 0.010**	**0.002**	**-0.013 ± 0.004**	**<0.001**	-0.004 ± 0.002	0.007
CA4	-0.027 ± 0.010	0.010	-0.010 ± 0.004	0.009	-0.003 ± 0.002	0.053
CA2/3	**-0.039 ± 0.011**	**<0.001**	**-0.014 ± 0.004**	**<0.001**	-0.004 ± 0.002	0.020
CA1	-0.025 ± 0.011	0.019	-0.009 ± 0.004	0.014	-0.003 ± 0.002	0.090
Subiculum	-0.020 ± 0.011	0.088	**-0.011 ± 0.004**	**0.005**	-0.004 ± 0.002	0.043
Presubiculum	-0.019 ± 0.011	0.097	-0.010 ± 0.004	0.017	-0.004 ± 0.002	0.027
Parasubiculum	0.001 ± 0.012	0.957	0.003 ± 0.004	0.430	-0.001 ± 0.002	0.449
Tail	**-0.036 ± 0.012**	**0.003**	**-0.016 ± 0.004**	**<0.001**	**-0.006 ± 0.002**	**<0.001**

Supplementary Table S3.5 | Associations of continuous measures of hyperglycemia with hippocampal subfield volumes. Results are presented as standardized beta coefficients (β) ± standard error (SE), and p-values. Bold shows multiple comparison correction significant (p<0.0071). Abbreviations: HbA1c, Hemoglobin A1c; THV, Total hippocampal volume; HATA, Hippocampus-amygdala-transition-area; CA, Cornu Ammonis.

Supplementary Table S3.6

Structure	Fasting glucose t0		Glucose t120		HbA1c mmol/mol	
	β±SE	p value	β±SE	p value	β±SE	p value
THV	-0.008 ± 0.026	0.755	-0.008 ± 0.008	0.320	0.001 ± 0.003	0.786
HATA	-0.033 ± 0.029	0.258	-0.011 ± 0.009	0.224	-0.002 ± 0.004	0.615
Fimbria	-0.033 ± 0.029	0.261	**-0.031 ± 0.009**	**<0.001**	-0.004 ± 0.004	0.244
Fissure	0.034 ± 0.031	0.270	0.005 ± 0.010	0.589	0.005 ± 0.004	0.185
Molecular layer	0.036 ± 0.029	0.214	0.010 ± 0.009	0.282	0.001 ± 0.004	0.765
Dentate gyrus	-0.018 ± 0.027	0.505	-0.008 ± 0.008	0.314	0.002 ± 0.003	0.622
CA4	-0.019 ± 0.027	0.491	-0.005 ± 0.008	0.559	0.003 ± 0.003	0.363
CA2/3	-0.029 ± 0.030	0.334	-0.016 ± 0.009	0.074	0.005 ± 0.004	0.220
CA1	0.011 ± 0.028	0.705	0.003 ± 0.009	0.761	0.006 ± 0.004	0.078
Subiculum	0.020 ± 0.030	0.498	0.000 ± 0.009	0.986	0.001 ± 0.004	0.736
Presubiculum	-0.023 ± 0.030	0.432	-0.008 ± 0.009	0.372	-0.008 ± 0.004	0.031
Parasubiculum	-0.007 ± 0.031	0.824	0.010 ± 0.010	0.277	-0.008 ± 0.004	0.050
Tail	-0.033 ± 0.031	0.291	-0.022 ± 0.009	0.022	-0.002 ± 0.004	0.530

Supplementary Table S3.6 | Associations of continuous measures of hyperglycemia with hippocampal subfield volumes, after the exclusion of participants with type 2 diabetes. Results are presented as standardized beta coefficients (β) ± standard error (SE), and p-values. Bold shows multiple comparison correction significant (p<0.0071). Abbreviations: HbA1c, Hemoglobin A1c; THV, Total hippocampal volume; HATA, Hippocampus-amygdala-transition-area; CA, Cornu Ammonis.

Supplementary Table S3.7

	Prediabetes		Type 2 diabetes	
Structure	β±SE	p value	β±SE	p value
THV	-0.005 ± 0.031	0.864	-0.052 ± 0.033	0.117
HATA	-0.020 ± 0.036	0.590	-0.024 ± 0.039	0.531
Fimbria	-0.064 ± 0.036	0.081	**-0.133 ± 0.039**	**<0.001**
Fissure	0.004 ± 0.040	0.912	0.062 ± 0.043	0.149
Molecular layer	0.040 ± 0.037	0.272	0.028 ± 0.039	0.482
Dentate gyrus	-0.020 ± 0.032	0.531	-0.054 ± 0.034	0.114
CA4	-0.013 ± 0.032	0.697	-0.021 ± 0.035	0.540
CA2/3	-0.032 ± 0.037	0.381	-0.078 ± 0.039	0.047
CA1	0.031 ± 0.034	0.364	-0.017 ± 0.036	0.636
Subiculum	0.005 ± 0.037	0.893	-0.049 ± 0.039	0.204
Presubiculum	-0.023 ± 0.037	0.525	-0.039 ± 0.039	0.322
Parasubiculum	0.068 ± 0.039	0.084	0.057 ± 0.042	0.176
Tail	-0.038 ± 0.038	0.322	-0.086 ± 0.041	0.034

Supplementary Table S3.7 | Associations of glucose metabolism status with hippocampal subfield volumes, after replacing intracranial volume for total brain volume. Results are presented as standardized beta coefficients (β) ± standard error (SE), and p-values. Bold shows multiple comparison correction significant (p<0.0071). Abbreviations: THV, Total hippocampal volume; HATA, Hippocampus-amygdala-transition-area; CA, Cornu Ammonis.

Supplementary Table S3.8

	Prediabetes		Type 2 diabetes	
Structure	β±SE	p value	β±SE	p value
THV	-0.035 ± 0.034	0.307	**-0.147 ± 0.036**	**<0.001**
HATA	-0.037 ± 0.038	0.321	-0.090 ± 0.040	0.024
Fimbria	-0.084 ± 0.037	0.024	**-0.192 ± 0.039**	**<0.001**
Fissure	0.001 ± 0.040	0.975	0.038 ± 0.043	0.371
Molecular layer	0.026 ± 0.038	0.489	-0.032 ± 0.040	0.424
Dentate gyrus	-0.044 ± 0.034	0.197	**-0.140 ± 0.036**	**<0.001**
CA4	-0.037 ± 0.035	0.286	**-0.108 ± 0.037**	**0.003**
CA2/3	-0.060 ± 0.039	0.120	**-0.158 ± 0.041**	**<0.001**
CA1	0.007 ± 0.036	0.855	-0.098 ± 0.038	0.009
Subiculum	-0.022 ± 0.038	0.571	**-0.134 ± 0.040**	**<0.001**
Presubiculum	-0.048 ± 0.038	0.207	**-0.117 ± 0.040**	**0.004**
Parasubiculum	0.055 ± 0.040	0.170	0.003 ± 0.042	0.939
Tail	-0.065 ± 0.040	0.104	**-0.162 ± 0.042**	**<0.001**

Supplementary Table S3.8 | Associations of glucose metabolism status with hippocampal subfield volumes, after changing waist for BMI. Results are presented as standardized beta coefficients (β) ± standard error (SE), and p-values. Bold shows multiple comparison correction significant (p<0.0071). Abbreviations: THV, Total hippocampal volume; HATA, Hippocampus-amygdala-transition-area; CA, Cornu Ammonis.

Supplementary Table S3.9

	Prediabetes		Type 2 diabetes	
Structure	β±SE	p value	β±SE	p value
THV	-0.028 ± 0.034	0.400	**-0.136 ± 0.036**	**<0.001**
HATA	-0.039 ± 0.038	0.296	**-0.092 ± 0.040**	**0.023**
Fimbria	**-0.076 ± 0.037**	**0.043**	**-0.175 ± 0.040**	**<0.001**
Fissure	-0.002 ± 0.040	0.970	0.035 ± 0.043	0.419
Molecular layer	0.028 ± 0.038	0.465	-0.039 ± 0.040	0.336
Dentate gyrus	-0.040 ± 0.034	0.240	**-0.133 ± 0.037**	**<0.001**
CA4	-0.034 ± 0.035	0.332	**-0.104 ± 0.037**	**0.005**
CA2/3	-0.051 ± 0.038	0.181	**-0.141 ± 0.041**	**<0.001**
CA1	0.011 ± 0.036	0.750	**-0.089 ± 0.038**	**0.020**
Subiculum	-0.014 ± 0.038	0.711	**-0.120 ± 0.041**	**0.003**
Presubiculum	-0.043 ± 0.038	0.260	**-0.107 ± 0.041**	**0.009**
Parasubiculum	0.053 ± 0.040	0.185	-0.010 ± 0.043	0.825
Tail	-0.059 ± 0.040	0.137	**-0.148 ± 0.043**	**<0.001**

Supplementary Table S3.9 | Associations of type 2 diabetes with hippocampal subfield volumes, after changing total-to-HDL cholesterol ratio for LDL cholesterol level. Results are presented as standardized beta coefficients (β) ± standard error (SE), and p-values. Bold shows multiple comparison correction significant (p<0.0071). Abbreviations: THV, Total hippocampal volume; HATA, Hippocampus-amygdala-transition-area; CA, Cornu Ammonis.

Supplementary Table S3.10

	Prediabetes		Type 2 diabetes	
Structure	β±SE	p value	β±SE	p value
THV	-0.030 ± 0.033	0.356	**-0.172 ± 0.034**	**<0.001**
HATA	-0.040 ± 0.036	0.267	**-0.128 ± 0.038**	**<0.001**
Fimbria	-0.056 ± 0.035	0.114	**-0.224 ± 0.037**	**<0.001**
Fissure	-0.017 ± 0.038	0.651	0.021 ± 0.040	0.595
Molecular layer	0.006 ± 0.036	0.868	-0.017 ± 0.038	0.656
Dentate gyrus	-0.035 ± 0.033	0.291	**-0.171 ± 0.035**	**<0.001**
CA4	-0.031 ± 0.033	0.361	**-0.139 ± 0.035**	**<0.001**
CA2/3	-0.056 ± 0.037	0.124	**-0.184 ± 0.038**	**<0.001**
CA1	0.010 ± 0.034	0.769	**-0.125 ± 0.036**	**<0.001**
Subiculum	-0.012 ± 0.037	0.735	**-0.158 ± 0.038**	**<0.001**
Presubiculum	-0.039 ± 0.036	0.288	**-0.141 ± 0.038**	**<0.001**
Parasubiculum	0.023 ± 0.039	0.545	-0.000 ± 0.040	0.997
Tail	-0.058 ± 0.038	0.127	**-0.171 ± 0.039**	**<0.001**

Supplementary Table S3.10 | Associations of glucose metabolism status with hippocampal subfield volumes, with no exclusion of cases based on segmentation accuracy (additional cases n=451). Results are presented as standardized beta coefficients (β) ± standard error (SE), and p-values. Bold shows multiple comparison correction significant (p<0.0071). Abbreviations: THV, Total hippocampal volume; HATA, Hippocampus-amygdala-transition-area; CA, Cornu Ammonis.

CHAPTER 4

ASSOCIATION OF HIPPOCAMPAL SUBFIELD VOLUMES WITH PREVALENCE, COURSE AND INCIDENCE OF DEPRESSIVE SYMPTOMS - THE MAASTRICHT STUDY

Monereo-Sanchez, J.; Jansen, J.F.A.; van Boxtel, M.P.J.; Backes W.H.; Köhler, S.; C.D.A. Stehouwer, C.D.A.; Linden, D.E.J.; Schram, M.T.

Abstract

Late-life depression has been associated with volume changes of the hippocampus. However, little is known about its association with specific hippocampal subfields over time. We investigated whether hippocampal subfield volumes were associated with prevalence, course and incidence of depressive symptoms.

We extracted 12 hippocampal subfield volumes per hemisphere with FreeSurfer v6.0 using T1w and FLAIR 3T MRI images. Depressive symptoms were assessed at baseline and annually over 7 years of follow-up (PHQ-9). We used negative binominal, logistic, and Cox regression analyses, corrected for multiple comparisons, and adjusted for demographic, cardiovascular, and lifestyle factors.

A total of n=4,174 participants were included (mean age 60.0±8.6 years, 51.8% female). Larger right hippocampal fissure volume was associated with prevalent depressive symptoms (OR[95%CI]; 1.26[1.08;1.48]). Larger bilateral hippocampal fissure (OR[CI]= 1.37-1.40[1.14;1.71], larger right molecular layer (OR[CI]= 1.51[1.14;2.00]), and smaller right cornu ammonis (CA)2/3 volumes (OR[CI]= 0.61[0.48;0.79]) were associated with prevalent depressive symptoms with a chronic course. No associations of hippocampal subfield volumes with incident depressive symptoms were found. Yet, lower left hippocampal amygdala transition area (HATA) volume was associated with incident depressive symptoms with chronic course (HR[CI]= 0.70[0.55;0.89]).

Differences in hippocampal fissure, molecular layer and CA volumes might co-occur or follow the onset of depressive symptoms, in particular with a chronic course. Smaller HATA was associated with an increased risk of incident (chronic) depression. Our results could capture a biological foundation for the development of chronic depressive symptoms and stress the need to discriminate subtypes of depression to unravel its biological underpinnings.

Introduction

Background

The contribution of structural brain changes to the prevalence, course, and incidence of late-life depression is a key topic of psychiatric neuroscience. Neuroanatomical substrates of depression could play a major role in diagnosis, prognosis, stratification of depression subtypes, and treatment monitoring. Although some robust associations have been identified previously, the field has not yet yielded information that is clinically applicable, and contributions to pathophysiological understanding have been limited. The most replicated finding among older adults has been an association between smaller hippocampus volume with depression (Hickie et al., 2005; Kempton et al., 2011; Sexton et al., 2012; Zhang, Peng, Sweeney, Jia, & Gong, 2018). This association may be especially noticeable in a longer depression duration or a larger number of depressive episodes (Campbell, Marriott, Nahmias, & MacQueen, 2004; Geerlings & Gerritsen, 2017; Kronmüller et al., 2008; Roddy et al., 2019; Treadway et al., 2015). Little is known about the temporality of this association. Yet one study suggested that longer duration and severity of depression lead to faster development of hippocampal atrophy (Taylor et al., 2014). Conversely, there is insufficient longitudinal data available to assess whether hippocampal atrophy may precede incident depression (den Heijer et al., 2011). Further, given that the hippocampus is a heterogeneous structure, composed of several subfields, each of which is characterized by specific cellular composition and characteristic neurophysiology (Fanselow & Dong, 2010), one may expect that different hippocampal subfields might be differentially associated with depression pathophysiology. Whereas this has been explored previously (Ballmaier et al., 2008; Lim et al., 2012; Roddy et al., 2019; Treadway et al., 2015), conflicting results have been presented, likely due to limited sample sizes and a lack of longitudinal data.

Aims

The aim of the present study was to investigate the associations of hippocampal subfield volumes with prevalence, course and incidence of depressive symptoms using a large neuroimaging sample. Specifically, we investigated the cross-sectional associations of hippocampal subfield volumes and depressive symptoms at baseline, and the longitudinal associations of hippocampal subfields volumes and depressive symptoms at follow up. In both cases we further subdivided the analysis according to the course of depression, i.e. chronic or transient, and corrected the analysis for demographic, cardiovascular, and lifestyle risk factors.

Method

Study population and design

We used data from The Maastricht Study, an observational prospective population-based cohort study. The rationale and methodology have been described previously (Schram et al., 2014). In brief, the study focuses on the etiology, pathophysiology, complications, and comorbidities of type 2 diabetes mellitus (T2DM) heart disease, and other chronic conditions, and is characterized by an extensive phenotyping approach. Eligible for participation were all individuals aged between 40 and 75 years and living in the southern part of the Netherlands. Participants were recruited through mass media campaigns, the municipal registries, and the regional Diabetes Patient Registry via mailings. Recruitment was stratified according to known T2DM status, with an oversampling of individuals with T2DM, for reasons of efficiency. Baseline data were collected between November 2010 and January 2018. Lag time between MRI and depression assessment at baseline was 102±120 days. The study has been approved by the institutional Medical Ethical Committee (NL31329.068.10) and the Minister of Health, Welfare and Sports of the Netherlands (Permit 131088-105234-PG). All participants gave written informed consent.

For the current analysis complete data was available from 4653 for cross sectional and 4154 participants for longitudinal analysis. Supplementary Figure S4.1 shows the flowchart of the study population.

Brain MRI

Brain images were acquired on a 3T clinical magnetic resonance scanner (MAGNETOM Prismafit, Siemens Healthineers GmbH) located at a dedicated scanning facility (Scannexus, Maastricht, The Netherlands) using a head/neck coil with 64 elements for parallel imaging. The MRI protocol included a three-dimensional (3D) T1-weighted (T1w) magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence (repetition time/inversion time/echo time (TR/TI/TE) 2,300/900/2.98ms, 176 slices, 256 × 240 matrix size, 1.0 mm cubic reconstructed voxel size); and a fluid-attenuated inversion recovery (FLAIR) sequence (TR/TI/TE 5,000/1,800/394 ms, 176 slices, 512 × 512 matrix size, 0.49 × 0.49 × 1.0 mm reconstructed voxel size).

Brain segmentation was performed with FreeSurfer v6.0 (Fischl, 2012) using both T1w and FLAIR images as input. The arguments "-FLAIRpial" and "-3T" were used to optimize segmentation quality. Brain segmentations with insufficient quality, i.e. Euler numbers below 1.5 quartile (-80 for left hemisphere and -68 for right hemisphere) were excluded (Monereo-Sánchez et al., 2021). Hippocampal subfields (J. E. Iglesias et al., 2015) were segmented using multispectral segmentation, yielding hippocampus total volume and 12 hippocampal subfields per hemisphere (Supplementary Table S4.1). All extracted volumes were z-transformed prior to statistical analysis with respect to the distribution in the complete sample (n=4,643). Results will be depicted in hippocampal subfields maps, a legend for these maps can be found in Supplementary Figure S4.2.

Depression

Depressive symptoms were assessed by a validated Dutch version of the 9-item Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer, & Williams) both at baseline and follow-up. Follow-up data was collected annually over a period of seven years. The PHQ-9 (Kroenke et al.) is a self-administered questionnaire that assesses the presence of the nine symptoms for the DSM-IV criteria for a major depressive disorder on a 4-point Likert-scale ranging from 0 "not at all" to 4 "nearly every day" (American Psychiatric Association, 1994). When one or two items were missing, the total score was calculated as 9 × (total points/9–number of missing items) and rounded to the nearest integer. When more items were missing, the total score was scored as missing. A cut-off score of ≥10 is most often used as a dichotomous scoring system for defining clinically relevant depressive symptoms, with a good sensitivity (88%) and specificity (78%)(Pettersson, Boström, Gustavsson, & Ekselius, 2015). The internal consistency of the PHQ-9 in The Maastricht Study was good (Cronbach's alpha = 0.82 without T2DM, and 0.87 with T2DM) (Janssen et al., 2016). There was a time lag between the baseline data collection and the date of the MRI scan. Therefore, the PHQ-9 score obtained closest to the date of the MRI scan, regardless of whether the assessment was before or after the scan, was chosen as the baseline score for each individual. Subsequent assessments were labelled as follow-up 1, follow-up 2, and so forth, based on the order in which they occurred after the baseline assessment.

Here, we use the term prevalent *depressive symptoms* to indicate the use of PHQ-9 scores as a continuum at baseline. We use the term prevalent *depression* to indicate clinically relevant depressive symptoms (PHQ-9 ≥10) at baseline. We subdivided prevalent depression according to its course as 1) Prevalent depression with a chronic course i.e., clinically relevant depressive symptoms (PHQ-9 ≥10) at baseline and clinically relevant depressive symptoms (PHQ-9 ≥10) on at least one follow-up moment; and 2) Prevalent depression with a transient course i.e. clinically relevant depressive symptoms (PHQ-9 ≥10) at baseline and no clinically relevant depressive symptoms (PHQ-9 <10) during follow-up.

We use the term *incident depression* to indicate no clinically relevant depressive symptoms (PHQ-9 <10) at baseline and presence of clinically relevant depressive symptoms (PHQ-9 ≥10) on at least one follow-up moment. We subdivided incident depression according to its course as 1) Incident depression with a chronic course i.e. no clinically relevant depressive symptoms (PHQ-9 <10) at baseline and clinically relevant depressive symptoms (PHQ-9 ≥10) on two or more follow-up moments; or (2) incident depression with a transient course, i.e. no clinically relevant depressive symptoms (PHQ-9 <10) at baseline and clinically relevant depressive symptoms (PHQ-9 ≥10) on one follow-up.

We used the term, *no depressive symptoms* as comparison group, and include those participants with no clinically relevant depressive symptoms (PHQ-9 <10) at baseline and no clinically relevant depressive symptoms (PHQ-9 <10) at follow-up.

General characteristics and covariates

General characteristics and covariates were measured at baseline. Educational level (low, intermediate, high), history of cardiovascular diseases (CVD) (Rose, 1962), smoking status (never, current, former), alcohol

consumption (none, low, high) were assessed by questionnaires. We measured, height, weight, waist circumference, office blood pressure, plasma lipid profile, and 24h urinary albumin excretion (twice) as described elsewhere (Schram et al., 2014). To determine T2DM status, all participants (except those who used insulin) underwent a standardized 7-point oral glucose tolerance test (OGTT) after an overnight fast. Glucose metabolism status was defined according to the World Health Organization 2006 criteria (World Health Organization, 2006) . Participants were considered to have T2DM if they had a fasting blood glucose ≥7.0 mmol/L or a 2-h post load blood glucose ≥11.1 mmol/L or used oral glucose-lowering medication or insulin. Medication use was assessed in a medication interview where generic name, dose, and frequency were registered.

Statistical analyses

General characteristics of the study population were evaluated using independent T-tests, or χ2 tests when appropriate.

We investigated the association between both hippocampus total volume and hippocampal subfields volumes and 1) prevalent depressive symptoms (negative binomial regression analysis); 2) prevalent depression (logistic regression analysis); 3) prevalent depression subdivided by its course, i.e. chronic or transient, using no depressive symptoms as comparison group (multinomial logistic regression analysis); 4) incident depression (Cox proportional hazards regression with time to event on the time-axis); and 5) incident depression subdivided by its course, i.e. chronic or transient (Cox proportional hazards regression with time to event on the time-axis).

We studied left and right hemispheres separately, thus analyzed 1 total hippocampal volume, and 12 hippocampal subfields for each hemisphere. Associations were adjusted for potential confounders: Model 1, adjusted for total or left/right brain volume, MRI lag time, age and sex; and Model 2, additionally adjusted for T2DM status, education level, waist circumference, history of cardiovascular disease, total-to-HDL cholesterol ratio, use of alcohol, and smoking status.

Correction for multiple comparisons was done in accordance to Matrix Spectral Decomposition (Nyholt, 2004). Based on the resulting eigenvalues, alpha was set at 0.05/13= 0.0039.

Several sensitivity analyses were performed based on Model 2: First, we excluded individuals with T2DM to assess whether they drive the observed associations. Second, we adjusted for antidepressant medication use. Third, we excluded participants who used antidepressant medication. Fourth, to restrict analyses to 'de novo' depression, we excluded participants who had a history of major depressive disorder diagnosis before baseline. Fifth, we tested whether these associations differed according to sex, and T2DM status, by use of interaction analyses. All statistical analyses were performed in R 4.0.2 (2020-06-22) (Team, 2013).

Results

General characteristics of the study population

The cross-sectional study population (N=4,643) had a mean age of 60.0±8.6 years, and 51.8% were women, 229 participants had prevalent depression (PHQ-9 ≥ 10).

Table 4.1 shows the general characteristics of the study population for longitudinal analysis (n=4,174) stratified for depressive status. A total of 190 participants showed prevalent depression, 141 of them had a chronic course during follow-up, and 49 had a transient course. Out of 3984 participants free of depression at baseline, 376 developed incident depression. Participants with no depressive symptoms were more often men, and had a better cardiovascular profile than those with prevalent or incident depression. Demographics of participants not included in this study due to missing data or bad segmentation quality can be seen in Supplementary Table S4.2.

Characteristic	No depressive symptoms n=3608	Prevalent depression n=190	Pval$^\alpha$	Incident depression n=376	Pval$^\alpha$
Age	60.4 ± 8.4	**57.3 ± 8.2**	**p<0.001**	**59.1 ± 8.7**	**0.005**
Sex (% women)	51.0	**61.1**	**0.007**	**57.2**	**0.022**
BMI (kg/m2)	26.1 ± 3.9	**27.8 ± 5.0**	**p<0.001**	**27.3 ± 4.7**	**p<0.001**
Waist (cm)	92.5 ± 12.1	**95.7 ± 14.0**	**0.002**	**95.3 ± 14.3**	**p<0.001**
Educational level (% Low, Medium, High)	28.8, 28.4, 42.9	**38.1, 33.3, 28.6**	**p<0.001**	**34.7, 29.5, 35.8**	**0.017**
Alcohol consumption (% None, Low, High)	14.9, 59.4, 25.7	**28.0 54.5 17.5**	**p<0.001**	**23.4, 54.0, 22.6**	**p<0.001**
Smoking status (% Never, Former, Current)	40.7, 49.3, 10.1	**40.2 40.7, 19.0**	**p<0.001**	**37.5, 46.1, 16.4**	**p<0.001**
Partner (% yes)	86.6	**77.4**	**p<0.001**	**78.7**	**p<0.001**
T2DM (% yes) [β]	16.0	**26.8**	**p<0.001**	**25.3**	**p<0.001**
CVD (% yes)	11.2	**16.5**	**0.026**	10.8	0.846
Hypertension (% yes)	46.9	52.1	0.157	51.3	0.098
Cholesterol ratio	3.6 ± 1.1	3.7 ± 1.3	0.066	**3.7 ± 1.2**	**0.049**
Cholesterol medication (% yes)	24.9	27.4	0.447	27.4	0.292
Antidepressants (% yes)	4.6	**27.9**	**p<0.001**	**15.2**	**p<0.001**
History of depression (% yes)	24.7	**79.1**	**p<0.001**	**51.8**	**p<0.001**

Table 4.1| General characteristics of the study population (n=4,174) stratified for depressive status. Data is presented as mean ± SD or percentage as appropriate. Abbreviations: No depressive symptoms, no clinically relevant depressive symptoms at baseline nor at follow up; Prevalent depression, clinically relevant depressive symptoms at baseline (PHQ-9 ≥ 10); Incident depression, no clinically relevant depressive symptoms at baseline and clinically relevant depressive symptoms at follow up; Pval, p-value; PHQ-9, patient health questionnaire; BMI, Body Mass Index; T2DM, Type 2 diabetes mellitus; CVD, cardiovascular disease. Bold indicates p<0.05.

[α] *Compared to no clinically relevant depressive symptoms at baseline and follow-up.*
[β] *The study is oversampled with individuals with type 2 diabetes by design.*

Hippocampal subfields and prevalent depression

We found no associations between total hippocampus volume and prevalent depressive symptoms (PHQ-9 score as a continuum; Supplementary Table S4.3). Larger volumes in bilateral molecular layer (left RR[CI]=1.11[1.05;1.17] / right RR[CI]=1.14[1.08;1.20]) and right hippocampal fissure (RR[CI]=1.06[1.02;1.11]), and smaller volumes in right dentate gyrus (RR[CI]=0.89[0.83;0.96]) and right Cornu Ammonis (CA) 2/3 (RR[CI]=0.91[0.87;0.96]) and CA4 (RR[CI]=0.89[0.83;0.96]) were significantly associated with depressive symptoms in Model 1. The association between smaller volumes in CA4 and depressive symptoms remained significant (RR[CI]=0.89[0.83;0.96]) after full adjustment. Hippocampal subfields associations in Model 2 are depicted in Supplementary Figure S4.3.

No association of total hippocampus volume with prevalent depression (PHQ-9≥10) was found (Supplementary Table S4.4). A larger right molecular layer (OR[CI]= 1.46[1.18;1.81]) and hippocampal fissure (OR[CI]= 1.32[1.13;1.53]), and smaller CA-2/3 (OR[CI]= 0.72[0.59;0.87]) were associated with prevalent depression in Model 1. The association with right hippocampal fissure remained significant (OR[CI]= 1.26[1.08;1.48]) after full adjustment (Model 2). Results are depicted in Figure 4.1.
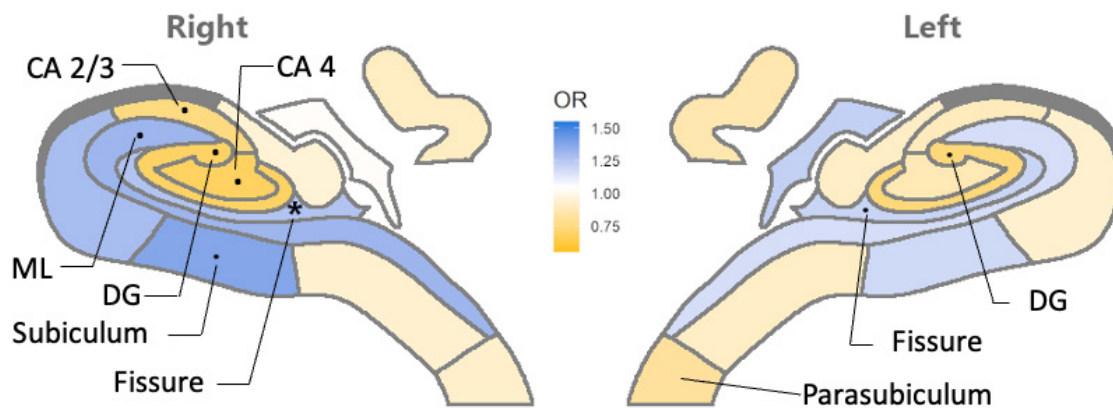
Figure 4.1| *Hippocampal map showing the associations between hippocampal subfield volumes and prevalent depression (PHQ-9≥10). Diagram displaying the subfields' associations with prevalent depression, after full adjustment (Model 2). Blue color represents a positive association: a higher volume is associated with higher odd ratio (OR) for prevalent depression, while yellow represent a negative association. Dots show the hippocampal subfields with associations of p<0.05; stars show the subfields that are significant after multiple comparison correction (p<0.0039). See Supplementary figure S4.2 for a hippocampal map legend. Abbreviations: Prevalent depression, clinically relevant depressive symptoms at baseline (PHQ-9 ≥ 10); CA, Cornu Ammonis; DG, Dentate gyrus; ML, Molecular layer.*

Hippocampal subfields volumes and course of prevalent depression

A significant association between lower volumes in the right hippocampus total volume and chronic course of prevalent depression (OR[CI]= 0.68[0.52;0.87]) was found after full adjustment (Model 2). Larger bilateral hippocampal fissure (left OR[CI]= 1.42[1.18;1.70]/ right OR[CI]= 1.46[1.21;1.77]) and molecular layer (left OR[CI]= 1.45[1.13;1.85]/right OR[CI]= 1.66[1.27;2.19]), as well as smaller left parasubiculum (OR[CI]= 0.73[0.59;0.90]) and right CA2/3 (OR[CI]= 0.60[0.47;0.77]) were associated with a higher risk ratio of chronic course of prevalent depression in Model 1. After full adjustment (Model 2), higher volumes in bilateral hippocampal fissure (left OR[CI]= 1.37[1.14;1.64]/ right OR[CI]= 1.40[1.15;1.71]), and right molecular layer (OR[CI]= 1.51[1.14;2.00]), as well as smaller volumes in right CA2/3 (OR[CI]= 0.61[0.48;0.79]) remained significantly associated with chronic course of prevalent depression. Results are depicted in Figure 4.2, and details can be found in Supplementary Table S4.5. No significant associations were found for transient course of prevalent depression.
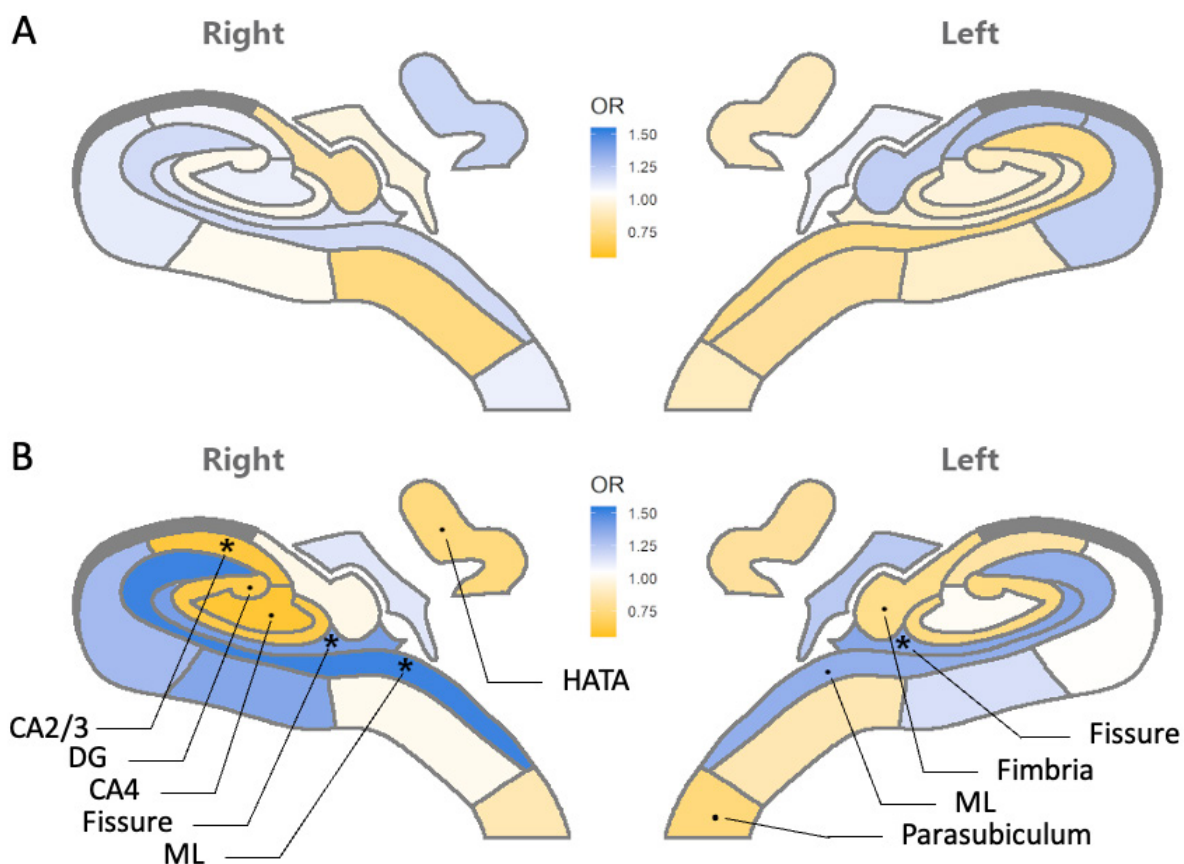
Figure 4.2| The hippocampal map shows the associations between hippocampal subfield volumes and (A) transient or (B) chronic course of prevalent depression. Diagram displaying the subfields' associations with A) Transient course of prevalent depression, and B) Chronic course of prevalent depression in Model 2. Blue color represents a positive association: a higher volume is associated with higher odds ratio (OR) for depression, while yellow represents a negative association. Dots show the subfields with associations of p<0.05, stars show the subfields that are significant after multiple comparison correction (p<0.0039). See Supplementary figure S4.2 for a hippocampal map legend. Abbreviations: Prevalent depression, clinically relevant depressive symptoms at baseline; CA, Cornu Ammonis; DG, Dentate gyrus; ML, Molecular layer.

Hippocampal subfields and incident depression

No significant associations were found between hippocampal volumes and incident depression (Supplementary Table S4.6).

Hippocampal subfields and course of incident depression

A statistically significant association between lower volume in left HATA and chronic course of incident depression was found (HR[CI]= 0.70[0.55;0.89]), while we found no associations with transient course of incident depression (Table S4.7).

Sensitivity analysis

Sensitivity analysis show results with preserved direction of effect and higher p-values when A) excluding participants with T2DM, B) adjusting for antidepressant medication, C) excluding participants using antidepressant medication, and D) excluding participants with a lifetime of major depressive disorder diagnosis. Results are detailed for prevalent depressive symptoms (Table S4.8), prevalent depression (Table S4.9), and prevalent depression with a chronic course (Table S4.10). No interactions with sex or T2DM were found in the associations of depression and hippocampal volumes (data not shown).

Discussion

In this middle-to-older-aged population, we studied the associations between hippocampal subfield volumes and prevalence, course, and incidence of depressive symptoms. We show that specific hippocampal subfields are associated with prevalent depression, especially with a chronic course. One subfield was also associated with incident depression, yet only when the course was chronic. To our knowledge, this is the first study that investigated the association of specific hippocampal subfield volumes with depressive symptoms in a population-based sample.

Larger right hippocampal fissure and bilateral molecular layer, as well as smaller right dentate gyrus, and CA2-4, were associated with prevalent depressive symptoms (PHQ-9 score as a continuum). Larger right fissure and molecular layer, as well as smaller right CA2/3 were associated with prevalent depression (PHQ-9 ≥ 10), independently of age, sex, and total hippocampal volume. The associations between hippocampal subfields and depression severity have been previously explored among major depressive disorder (MDD) patients in small clinical samples (n=41 to 163). In line with our results, Hu et.al. (2019) (Hu et al., 2019) found that lower CA2/3 and CA4/DG volumes were associated with more severe depressive symptoms,

with a similar trend also found by Brown et.al. (2019)(S. Brown et al., 2019). We could not replicate their findings on a significant association with lower volumes in subiculum. In addition, our results are in line with previous studies in clinical samples that compared patients with MDD to controls (Ballmaier et al., 2008; Han, Won, Sim, & Tae, 2016; Roddy et al., 2019), whom found smaller volumes in CA structures, subiculum and tail associated with MDD. A main difference is, though, that we find these associations more often in the right hemisphere, while the previous studies reported differences in both hemispheres. These differences may be explained by the differences in study samples (clinical vs population-based), and difference in instruments to assess depression (MDD diagnosis vs depressive symptoms). Although our definition of depression status is a reliable approach for MDD screening (Janssen et al., 2016; Muñoz-Navarro et al., 2017), our study sample likely includes less severely affected individuals. This might mean that CA1, subiculum, and tail have more subtle or later roles in depression pathophysiology, being only detectable in more severe depression, in line with results by Roddy et.al.(Roddy et al., 2019). Further, we found an association with the hippocampal fissure volume, which has not been reported before. The hippocampal fissure is not a tissue structure per se, but a CSF filled cavity, defined as a space between the dentate gyrus and the molecular layer (J. E. Iglesias et al., 2015). A larger volume could be driven by the general atrophy of the hippocampus. However, given the larger volume we find in the molecular layer (probably due to swelling), we hypothesize that the observed larger hippocampal fissure volume may be the result of reshaping of the hippocampus, which might create a larger relative distance between the dentate gyrus and the molecular layer, increasing the volume of the hippocampal fissure. We corrected for total hippocampal volume which allowed us to detect subtle subfield-specific changes of the hippocampus, and may explain why this finding has not been reported previously.

We further studied the association of hippocampal subfields with the course of prevalent depression. We found some hippocampal subfields were associated with a chronic course, while none was associated with a transient course. Specifically, larger bilateral fissure and molecular layer, and smaller volumes of left parasubiculum, right CA2/3 and right total hippocampal volume were associated with a chronic course. Previous studies found an association between depression recurrence and total hippocampal atrophy (Campbell et al., 2004; Kronmüller et al., 2008; Roddy et al., 2019), yet only one study explored this association on hippocampal subfields, finding smaller volumes in dentate gyrus (Treadway et al., 2015). The different patterns of hippocampal morphology in patients with transient or chronic depression may suggest that hippocampus atrophy is of importance in the pathophysiology of chronic depression, while is not in transient depression. Some studies also explored the utility of hippocampal subfields in the measurement of treatment response, with promising results finding an increase in hippocampal volumes after some treatments, and remission of depression (Cao et al., 2018; Hu et al., 2019; Kraus et al., 2019). Overall, our results suggest that the different subfields of the hippocampus might have a different sensitivity to depression. Cytology studies suggest that a deficiency in neurotrophic support might be the cause (Stockmeier et al., 2004), and that the compensation of neurotrophic factors through pharmacological therapy could reverse the pathological process of depression (Nestler et al., 2002).

We found no significant associations between hippocampal subfield volumes and risk of incident depression, in line with a previous smaller population-based study (den Heijer et al., 2011). Yet, when we subdivided this analysis according to the course, we found lower volumes in left HATA to be associated with incident depression with a chronic course. This might indicate that there is a different etiology in incident depression with a chronic course versus a transient course. Replication of our findings is needed, and future studies should clarify whether changes in hippocampal volumes are specific for subtypes of depression.

Strengths of this study include the large sample size and population-based design, the extensive assessment of potential confounders which reduces the chance of residual confounding and the annual assessment of depressive symptoms over a 7-year period. To assess robustness of observed associations we performed a range of sensitivity analyses. Results remained similar after additionally adjusting for antidepressant medication or limiting the sample to de-novo depression. Potential selection and/or attrition bias, which is inherent to prospective population-based studies, may have resulted in underestimation of the observed associations. In addition, depression was measured with PHQ-9 questionnaire, which is a reliable and valid tool for the measurement of depressive symptoms, but is not equal to a clinical diagnosis of MDD. (Kroenke et al.; Martin, Rief, Klaiberg, & Braehler, 2006; Negeri et al., 2021). Finally, hippocampal volumes were extracted using FreeSurfer v6.0 automated tool. FreeSurfer v6.0 has proven to be a reliable method for hippocampal subfields volume's measurement, showing a good agreement with manual segmentation (Tae, Kim, Lee, Nam, & Kim, 2008). It also shows a good test-retest reliability, especially in the tail, subiculum, presubiculum, CA1-4, dentate Gyrus, and molecular layer (E. M. Brown et al., 2020; Marizzoni et al., 2015; Quattrini et al., 2020; Worker et al., 2018). Moreover, its use has previously proved useful to provide insight into the neurobiological underpinnings of several brain related traits and disorders (Sämann et al., 2018). In this study, the hippocampal segmentation was implemented with the additional use of a FLAIR image (Multispectral segmentation)(J. Iglesias, 2020) which has shown to additionally improve the subfields segmentation reliability (J. E. Iglesias et al., 2015; Seiger et al., 2021). Further, all FreeSurfer output used in The Maastricht Study undergoes quality control through the exclusion of outliers based on Euler numbers, technique that shows similar quality control benefits than visual inspection for hippocampal subfields segmentation (Monereo-Sánchez et al., 2021), reinforcing the solidity of the data.

In conclusion, differences in hippocampal volumes of specific subfields, indicating hippocampal atrophy, were associated with prevalent depression, in particular with a chronic course. In longitudinal analyses we found some evidence that smaller volume in left HATA was associated with a risk of incident depression with a chronic course. Our results indicate that changes in hippocampus subfield volumes may co-occur or follow the onset of depressive symptoms, rather than precede it. We found limited evidence to support that specific volume changes could precede the onset of (chronic) depressive symptoms. Therefore, our results could be capturing a biological foundation for the development of chronic depression, and further stresses the need to discriminate between subtypes of depression to unravel its biological underpinnings.

References

American Psychiatric Association. (1994). DSM-IV: Diagnostic and statistical manual of mental disorders. In: Washington, DC: American Psychiatric Association.

Ballmaier, M., Narr, K. L., Toga, A. W., Elderkin-Thompson, V., Thompson, P. M., Hamilton, L., . . . Kumar, A. (2008). Hippocampal morphology and distinguishing late-onset from early-onset elderly depression. *American journal of Psychiatry, 165*(2), 229-237.

Brown, E. M., Pierce, M. E., Clark, D. C., Fischl, B. R., Iglesias, J. E., Milberg, W. P., . . . Salat, D. H. (2020). Test-retest reliability of FreeSurfer automated hippocampal subfield segmentation within and across scanners. *NeuroImage, 210*, 116563.

Brown, S., Rutland, J., Verma, G., Feldman, R., Alper, J., Schneider, M., . . . Balchandani, P. (2019). Structural MRI at 7T reveals amygdala nuclei and hippocampal subfield volumetric association with major depressive disorder symptom severity. *Scientific reports, 9*(1), 1-10.

Campbell, S., Marriott, M., Nahmias, C., & MacQueen, G. M. (2004). Lower hippocampal volume in patients suffering from depression: a meta-analysis. *American journal of Psychiatry, 161*(4), 598-607.

Cao, B., Luo, Q., Fu, Y., Du, L., Qiu, T., Yang, X., . . . Cho, R. Y. (2018). Predicting individual responses to the electroconvulsive therapy with hippocampal subfield volumes in major depression disorder. *Scientific reports, 8*(1), 1-8.

den Heijer, T., Tiemeier, H., Luijendijk, H. J., van der Lijn, F., Koudstaal, P. J., Hofman, A., & Breteler, M. M. (2011). A study of the bidirectional association between hippocampal volume on magnetic resonance imaging and depression in the elderly. *Biological psychiatry, 70*(2), 191-197.

Fanselow, M. S., & Dong, H.-W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron, 65*(1), 7-19.

Fischl, B. (2012). FreeSurfer. *NeuroImage, 62*(2), 774-781.

Geerlings, M. I., & Gerritsen, L. (2017). Late-life depression, hippocampal volumes, and hypothalamic-pituitary-adrenal axis regulation: a systematic review and meta-analysis. *Biological psychiatry, 82*(5), 339-350.

Han, K.-M., Won, E., Sim, Y., & Tae, W.-S. (2016). Hippocampal subfield analysis in medication-naive female patients with major depressive disorder. *Journal of affective disorders, 194*, 21-29.

Hickie, I., Naismith, S., Ward, P. B., Turner, K., Scott, E., Mitchell, P., . . . Parker, G. (2005). Reduced hippocampal volumes and memory loss in patients with early-and late-onset depression. *The British Journal of Psychiatry, 186*(3), 197-202.

Hu, X., Zhang, L., Hu, X., Lu, L., Tang, S., Li, H., . . . Huang, X. (2019). Abnormal hippocampal subfields may be potential predictors of worse early response to antidepressant treatment in drug-Naïve patients with major depressive disorder. *Journal of Magnetic Resonance Imaging, 49*(6), 1760-1768.

Iglesias, J. (2020, 2020-05-08). Hippocampal Subfields. Retrieved from https://surfer.nmr.mgh.harvard.edu/fswiki/HippocampalSubfields

Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., . . . Wald, L. L. (2015). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. *NeuroImage, 115*, 117-137.

Janssen, E. P., Köhler, S., Stehouwer, C. D., Schaper, N. C., Dagnelie, P. C., Sep, S. J., . . . Schram, M. T. (2016). The Patient Health Questionnaire-9 as a screening tool for depression in individuals with type 2 diabetes mellitus: The Maastricht study. *Journal of the American Geriatrics Society, 64*(11), e201-e206.

Kempton, M. J., Salvador, Z., Munafò, M. R., Geddes, J. R., Simmons, A., Frangou, S., & Williams, S. C. (2011). Structural neuroimaging studies in major depressive disorder: meta-analysis and comparison with bipolar disorder. *Archives of general psychiatry, 68*(7), 675-690.
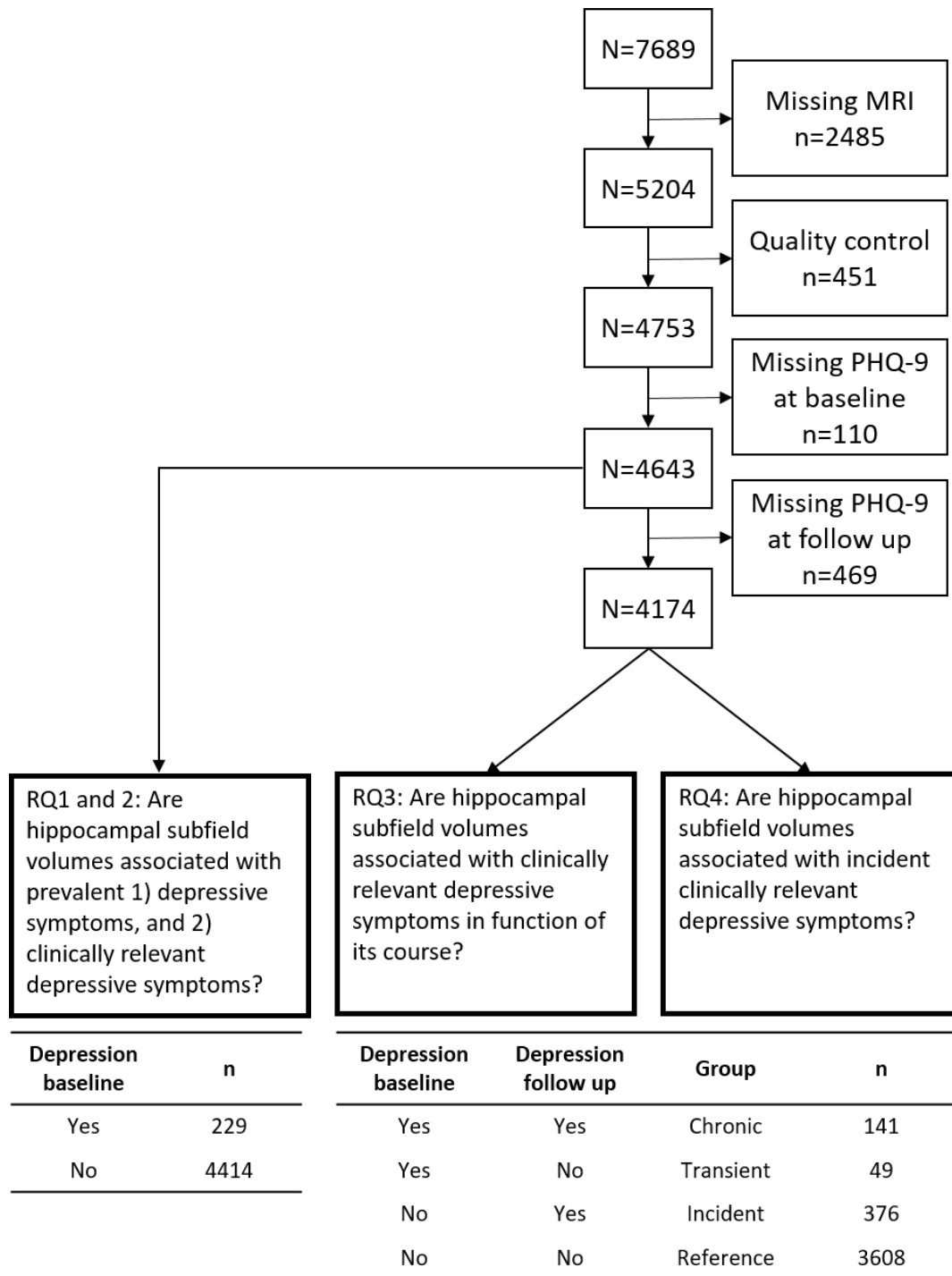
Kraus, C., Seiger, R., Pfabigan, D. M., Sladky, R., Tik, M., Paul, K., . . . Komorowski, A. (2019). Hippocampal subfields in acute and remitted depression—an ultra-high field magnetic resonance imaging study. *International Journal of Neuropsychopharmacology, 22*(8), 513-522.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine, 16*(9), 606-613.

Kronmüller, K.-T., Pantel, J., Köhler, S., Victor, D., Giesel, F., Magnotta, V. A., . . . Schröder, J. (2008). Hippocampal volume and 2-year outcome in depression. *The British Journal of Psychiatry, 192*(6), 472-473.

Lim, H. K., Hong, S. C., Jung, W. S., Ahn, K. J., Won, W. Y., Hahn, C., . . . Lee, C. U. (2012). Automated hippocampal subfields segmentation in late life depression. *Journal of affective disorders, 143*(1-3), 253-256.

Marizzoni, M., Antelmi, L., Bosch, B., Bartrés-Faz, D., Müller, B. W., Wiltfang, J., . . . Nobili, F. (2015). Longitudinal reproducibility of automatically segmented hippocampal subfields: A multisite E uropean 3T study on healthy elderly. *Human brain mapping, 36*(9), 3516-3527.

Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. *General hospital psychiatry, 28*(1), 71-77.

Monereo-Sánchez, J., de Jong, J. J., Drenthen, G. S., Beran, M., Backes, W. H., Stehouwer, C. D., . . . Jansen, J. F. (2021). Quality control strategies for brain MRI segmentation and parcellation: Practical approaches and recommendations-insights from the Maastricht study. *Neuroimage, 237*, 118174.

Muñoz-Navarro, R., Cano-Vindel, A., Medrano, L. A., Schmitz, F., Ruiz-Rodríguez, P., Abellán-Maeso, C., . . . Hermosilla-Pasamar, A. M. (2017). Utility of the PHQ-9 to identify major depressive disorder in adult patients in Spanish primary care centres. *BMC psychiatry, 17*(1), 1-9.

Negeri, Z. F., Levis, B., Sun, Y., He, C., Krishnan, A., Wu, Y., . . . Benedetti, A. (2021). Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. *bmj, 375*.

Nestler, E. J., Barrot, M., DiLeone, R. J., Eisch, A. J., Gold, S. J., & Monteggia, L. M. (2002). Neurobiology of depression. *Neuron, 34*(1), 13-25.

Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics, 74*(4), 765-769.

Pettersson, A., Boström, K. B., Gustavsson, P., & Ekselius, L. (2015). Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nordic journal of psychiatry, 69*(7), 497-508.

Quattrini, G., Pievani, M., Jovicich, J., Aiello, M., Bargalló, N., Barkhof, F., . . . Blin, O. (2020). Amygdalar nuclei and hippocampal subfields on MRI: Test-retest reliability of automated volumetry across different MRI sites and vendors. *NeuroImage, 218*, 116932.

Roddy, D. W., Farrell, C., Doolin, K., Roman, E., Tozzi, L., Frodl, T., . . . O'Hanlon, E. (2019). The hippocampus in depression: more than the sum of its parts? Advanced hippocampal substructure segmentation in depression. *Biological psychiatry, 85*(6), 487-497.

Rose, G. A. (1962). The diagnosis of ischaemic heart pain and intermittent claudication in field surveys. *Bulletin of the World Health Organization, 27*(6), 645.

Sämann, P. G., Iglesias, J. E., Gutman, B., Grotegerd, D., Leenings, R., Flint, C., . . . van Erp, T. G. (2018). FreeSurfer-based segmentation of hippocampal subfields: A review of methods and applications, with a novel quality control procedure for ENIGMA studies and other collaborative efforts. *Human Brain Mapping*.

Schram, M. T., Sep, S. J., van der Kallen, C. J., Dagnelie, P. C., Koster, A., Schaper, N., . . . Stehouwer, C. D. (2014). The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *European journal of epidemiology, 29*, 439-451.

Seiger, R., Hammerle, F. P., Godbersen, G. M., Reed, M. B., Spurny-Dworak, B., Handschuh, P., . . . Vanicek, T. (2021). Comparison and reliability of hippocampal subfield segmentations within FreeSurfer utilizing T1-and T2-weighted multispectral MRI data. *Frontiers in neuroscience, 15*.

Sexton, C. E., Le Masurier, M., Allan, C. L., Jenkinson, M., McDermott, L., Kalu, U. G., . . . Ebmeier, K. P. (2012). Magnetic resonance imaging in late-life depression: vascular and glucocorticoid cascade hypotheses. *The British Journal of Psychiatry, 201*(1), 46-51.

Stockmeier, C. A., Mahajan, G. J., Konick, L. C., Overholser, J. C., Jurjus, G. J., Meltzer, H. Y., . . . Rajkowska, G. (2004). Cellular changes in the postmortem hippocampus in major depression. *Biological psychiatry, 56*(9), 640-650.

Tae, W. S., Kim, S. S., Lee, K. U., Nam, E.-C., & Kim, K. W. (2008). Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology, 50*(7), 569.

Taylor, W. D., McQuoid, D. R., Payne, M. E., Zannas, A. S., MacFall, J. R., & Steffens, D. C. (2014). Hippocampus atrophy and the longitudinal course of late-life depression. *The American Journal of Geriatric Psychiatry, 22*(12), 1504-1512.

Team, R. C. (2013). R: A language and environment for statistical computing. In: Vienna, Austria.

Treadway, M. T., Waskom, M. L., Dillon, D. G., Holmes, A. J., Park, M. T. M., Chakravarty, M. M., . . . Fava, M. (2015). Illness progression, recent stress, and morphometry of hippocampal subfields and medial prefrontal cortex in major depression. *Biological psychiatry, 77*(3), 285-294.

Worker, A., Dima, D., Combes, A., Crum, W. R., Streffer, J., Einstein, S., . . . O'daly, O. (2018). Test–retest reliability and longitudinal analysis of automated hippocampal subregion volumes in healthy ageing and A lzheimer's disease populations. *Human Brain Mapping, 39*(4), 1743-1754.

World Health Organization. (2006). *Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation* (9241594934). Retrieved from https://apps.who.int/iris/handle/10665/43588

Zhang, F. F., Peng, W., Sweeney, J. A., Jia, Z. Y., & Gong, Q. Y. (2018). Brain structure alterations in depression: Psychoradiological evidence. *CNS neuroscience & therapeutics, 24*(11), 994-1003.

Chapter 4: Association of hippocampal subfield volumes with prevalence, course and incidence of depressive symptoms

Supplementary Figure S4.1

N=7689

Missing MRI
n=2485

N=5204

Quality control
n=451

N=4753

Missing PHQ-9
at baseline
n=110

N=4643

Missing PHQ-9
at follow up
n=469

N=4174

RQ1 and 2: Are hippocampal subfield volumes associated with prevalent 1) depressive symptoms, and 2) clinically relevant depressive symptoms?

RQ3: Are hippocampal subfield volumes associated with clinically relevant depressive symptoms in function of its course?

RQ4: Are hippocampal subfield volumes associated with incident clinically relevant depressive symptoms?

Depression baseline	n
Yes	229
No	4414

Depression baseline	Depression follow up	Group	n
Yes	Yes	Chronic	141
Yes	No	Transient	49
No	Yes	Incident	376
No	No	Reference	3608

Supplementary Figure S4.1 | Flowchart of the study population

Additional material and methods

From the initial 7,685 participants, baseline magnetic resonance imaging (MRI) data were available from n=5,204 participants. Quality control was then applied to the MRI data through the exclusion of outliers based on Euler Numbers, following guidelines (Monereo-Sánchez et al., 2021). Further, we excluded individuals with missing PHQ-9 data at baseline, resulting in a cross sectional study population of n=4,643. With this sample we answered the research questions (RQ) 1 and 2 We then excluded participants with missing follow up data, resulting in a longitudinal sample of n=4,174 participants. This sample is then divided by clinically relevant depressive symptoms at baseline and follow up to answer RQ3 and 4. Chronic depression contains those with depression at baseline and follow up, transient depression contains those with depression at baseline but not at follow up. Incident depression contains participants with no depression at baseline but at least one episode during follow up. The reference group contains participants with no depression at baseline nor at follow up. Subsequently, missing data among the covariates was excluded. Hence, Model 2 uses 76 less cases than Model 1 due to missing data in: education level (n=47), history of cardiovascular diseases (n=42), smoking and alcohol intake (n=25), cholesterol ratio (n=3), and waist circumference (n=2). Overall, during 18,328 person-years of follow-up, 517 (2.8%) participants developed clinically relevant depressive symptoms (PHQ-9 ≥ 10) during follow-up. Data was available among 88.0% (year 1), 79.6% (year 2), 74.0% (year 3), 65.4% (year 4), 57.9% (year 5), 35.8% (year 6), and 19.1% (year 7) of the participants. It is important to note that the lower percentages after the fifth year are a result of the ongoing annual follow-up from year 6 onwards.

Monereo-Sánchez, J., de Jong, J. J., Drenthen, G. S., Beran, M., Backes, W. H., Stehouwer, C. D., . . . Jansen, J. F. (2021). Quality control strategies for brain MRI segmentation and parcellation: Practical approaches and recommendations-insights from the Maastricht study. *Neuroimage, 237*, 118174.

Supplementary Table S4.1

In text name	Description (FreeSurfer name)	Left volume mm3 (mean ± SD)	Right volume mm3 (mean ± SD)
Hippocampus total volume	Total Hippocampal Volume (Whole_hippocampus)	3414.18 ± 354.78	3436.48 ± 356.99
Tail	Hippocampal tail (Hippocampal_tail)	533.86 ± 73.08	509.64 ± 68.47
Subiculum	Subiculum (Subiculum)	470.50 ± 53.94	468.95 ± 52.34
CA1	Cornu Ammonis 1 (CA1)	709.24 ± 86.80	740.29 ± 91.22
Fissure	Hippocampal fissure (Hippocampal-fissure)	77.85 ± 15.44	84.93 ± 17.22
Presubiculum	Presubiculum (Presubiculum)	307.61 ± 40.18	290.93 ± 38.54
Parasubiculum	Parasubiculum (Parasubiculum)	61.63 ± 13.12	61.19 ± 12.51
Molecular layer	Molecular layer of the hippocampus (Molecular_layer_HP)	351.56 ± 49.56	371.96 ± 49.80
Dentate gyrus	Granule cell and molecular cell layer of the dentate gyrus (GC_ML_DG)	333.98 ± 40.38	333.96 ± 40.84
CA2/3	Cornu Ammonis 2 and 3 (CA2/3)	231.63 ± 39.10	248.77 ± 39.84
CA4	Cornu Ammonis 4 (CA4)	273.66 ± 30.90	276.85 ± 32.03
Fimbria	Fimbria (Fimbria)	79.63 ± 23.69	71.12 ± 23.18
HATA	Hippocampus-amygdala-transition-area (HATA)	60.87 ± 10.88	62.83 ± 10.42

*Supplementary Table S4.1 | **Hippocampal subfields description.** Table shows the abbreviation and complete name, as well as the mean volume (mm3) ± standard deviation (SD) for total hippocampal volume and each hippocampal subfield of the cross sectional sample.*

Supplementary Figure S4.2
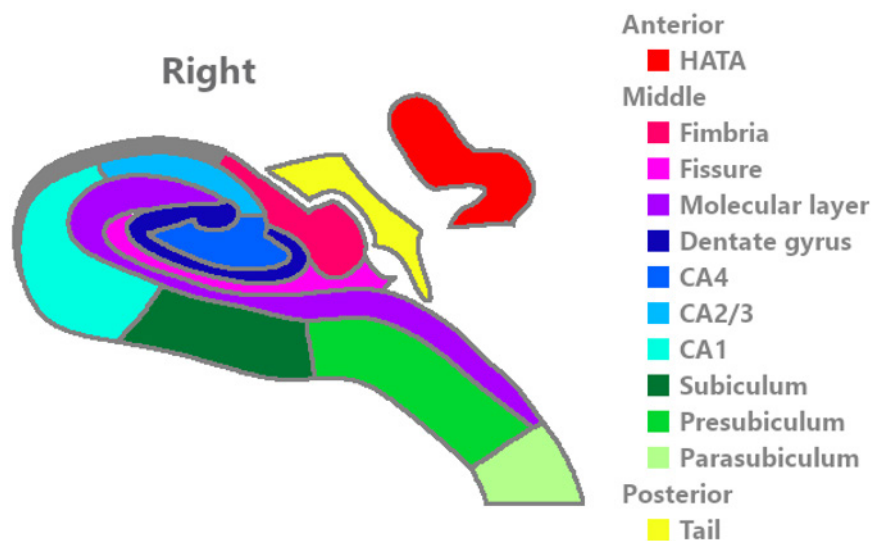
*Supplementary Figure S4.2 | **Hippocampal subfields map legend.** Figure shows a schematic representation of the right hippocampus in a coronal section. Hippocampal subfields are represented in different colors. Notice than the subfield HATA is transposed from an anterior coronal section, while Tail is transposed from a posterior coronal section. Abbreviations: HATA, Hippocampus-amygdala-transition-area; CA, Cornu Ammonis.*

Supplementary Table S4.2

Characteristic	Excluded n=3,046	Included n=4,643	p value
Age	66 ± 8	60 ± 8.6	**p<0.001**
Sex (% female)	46.4	51.8	**p<0.001**
BMI (Km/cm3)	28 ± 5	26.4 ± 4.1	**p<0.001**
Waist (cm)	98.8 ± 14.6	93 ± 12.6	**p<0.001**
Education (% Low, Medium, High)	41.3, 26.1, 32.6	30.4, 28.6, 40.9	**p<0.001**
Alcohol consumption (% None, Low, High)	21.2, 56.8, 22	16.8, 59, 24.3	**p<0.001**
Smoking status (% Never, Former, Current)	32.9, 51.4, 15.7	40, 48.1, 12	**p<0.001**
Partner (% yes)	80.6	85	**p<0.001**
T2DM (% yes)[α]	34.6	18.1	**p<0.001**
Cardiovascular disease (% yes)	26.1	11.2	**p<0.001**
Hypertension (% yes)	64.4	47	**p<0.001**
Cholesterol ratio	3.7 ± 1.2	3.6 ± 1.2	**p<0.001**
Cholesterol medication (% yes)	43.4	25.5	**p<0.001**
Antidepressants (% yes)	8.1	6.6	**p=0,015**
History of depression (% yes)	33.5	30.1	**p=0,0018**

Supplementary Table S4.2 | Characteristics of excluded vs. included participants. *Table shows the* demographics and group comparison results between included and excluded participants. Data is presented as mean ± SD or percentage as appropriate. Participants not included in this study due to unavailable MRI (n=2,485), low quality MRI (n=451), or missing baseline PHQ-9 data (n=110) were statistically significantly older, more often men, had a higher waist circumference, had lower education level, had a higher cardiovascular risk profile, and were more prone to take antidepressants than participants included in the analyses. Abbreviations: PHQ-9, patient health questionnaire; BMI, Body Mass Index; T2DM, Type 2 diabetes mellitus; CVD, cardiovascular disease. Bold shows p<0.05.

[α] *Study design is oversampled with individuals with type 2 diabetes for reasons of efficiency.*

Supplementary Table S4.3

Structure	MODEL 1		MODEL 2	
	RR(95%CI)	p value	RR(95%CI)	p value
Left hemisphere				
Total Hippocampus volume	0.98(0.93;1.03)	0.427	0.98(0.94;1.03)	0.477
Subfields				
Tail	1.03(0.98;1.08)	0.288	1.02(0.97;1.07)	0.441
Subiculum	1.03(0.97;1.10)	0.305	1.03(0.97;1.10)	0.279
Cornu Ammonis 1 (CA1)	0.99(0.92;1.08)	0.901	1.01(0.93;1.09)	0.779
Fissure	**1.05(1.01;1.09)**	**0.013**	1.04(1.00;1.08)	0.054
Presubiculum	0.98(0.93;1.02)	0.309	0.98(0.94;1.03)	0.526
Parasubiculum	0.97(0.94;1.01)	0.175	0.98(0.94;1.02)	0.287
Molecular layer	**1.11(1.05;1.17)**	**p<0.001** *	**1.06(1.01;1.12)**	**0.029**
Dentate gyrus	**0.92(0.85;0.99)**	**0.021**	**0.93(0.86;1.00)**	**0.040**
Cornu Ammonis 2/3 (CA2/3)	**0.94(0.90;0.98)**	**0.009**	0.96(0.91;1.00)	0.062
Cornu Ammonis 4 (CA4)	0.95(0.88;1.03)	0.221	0.96(0.88;1.04)	0.261
Fimbria	0.98(0.94;1.02)	0.390	0.98(0.94;1.03)	0.436
HATA	0.99(0.95;1.04)	0.756	1.00(0.96;1.05)	0.917
Right hemisphere				
Total Hippocampus volume	0.97(0.92;1.02)	0.214	0.97(0.92;1.01)	0.170
Subfields				
Tail	0.99(0.94;1.04)	0.649	0.99(0.94;1.05)	0.847
subiculum	1.03(0.97;1.10)	0.304	1.05(0.99;1.12)	0.133
Cornu Ammonis 1 (CA1)	**1.09(1.01;1.19)**	**0.040**	1.07(0.98;1.16)	0.120
Fissure	**1.06(1.02;1.11)**	**0.002** *	**1.04(1.00;1.08)**	**0.040**
Presubiculum	0.99(0.95;1.04)	0.744	1.00(0.96;1.05)	0.899
Parasubiculum	1.00(0.96;1.04)	0.955	1.00(0.96;1.04)	0.976
Molecular layer	**1.14(1.08;1.20)**	**p<0.001** *	**1.08(1.02;1.14)**	**0.006**
Dentate gyrus	**0.89(0.83;0.96)**	**0.002** *	**0.90(0.84;0.97)**	**0.005**
Cornu Ammonis 2/3 (CA2/3)	**0.91(0.87;0.96)**	**p<0.001** *	**0.94(0.89;0.98)**	**0.007**
Cornu Ammonis 4 (CA4)	**0.89(0.83;0.96)**	**0.002** *	**0.89(0.83;0.96)**	**0.002** *
Fimbria	0.99(0.95;1.04)	0.799	1.00(0.96;1.04)	0.898
HATA	0.99(0.95;1.04)	0.821	1.00(0.95;1.04)	0.939

*Supplementary Table S4.3 | **Results, prevalent depressive symptoms (PHQ-9 score).** Table shows the results for negative binomial regression on depressive symptoms load. Results are presented in rate ratio (RR) and confidence intervals (CI). Bold shows nominally significant (p<0.05); Star (*) indicate a significant result after multiple comparison correction (p<0.0039). Abbreviations; HATA, Hippocampus-amygdala-transition-area. Model 1 was, adjusted for brain total volume when analyzing hippocampal total volumes, or adjusted for left/right hippocampus total volume, when analyzing hippocampal subfields, MRI lag time, age and sex; Model 2 was additionally adjusted for potential confounders: T2DM status, education level, waist circumference, history of cardiovascular disease, cholesterol ratio, use of alcohol and smoking status.*

Supplementary Figure S4.3
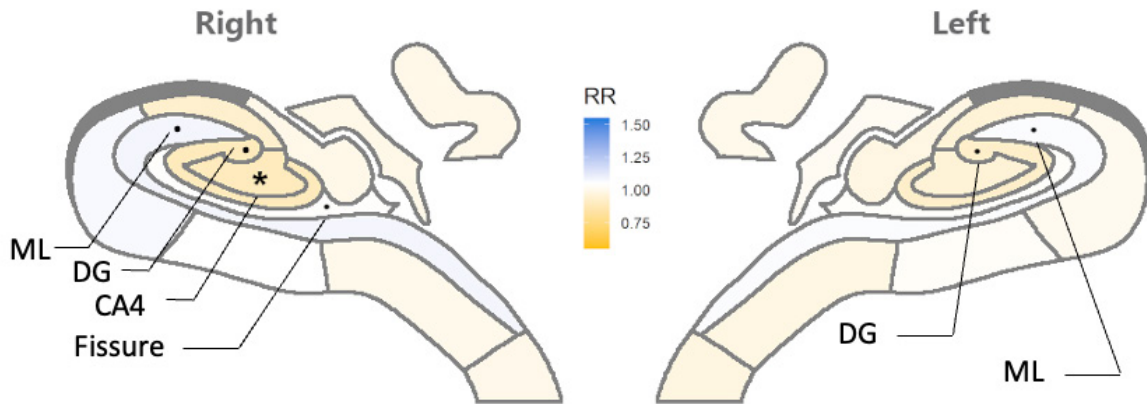
Supplementary Figure S4.3 | Hippocampal map: Prevalent depressive symptoms (PHQ-9 score). *The hippocampal map shows the associations between hippocampal subfield volumes and prevalent depressive symptoms on a continuous scale after full adjustment (Model 2). Blue color represents a positive association: a higher volume is associated with higher rate ratio (RR) for depression, while yellow represent a negative association. Dots show the hippocampal subfields with associations of p<0.05, stars show the subfields that are significant after multiple comparison correction (p<0.0039). See Supplementary figure S4.2 for a hippocampal map legend. Abbreviations: CA, Cornu Ammonis; DG, Dentate gyrus; ML, Molecular layer.*

Supplementary Table S4.4

Structure	MODEL 1		MODEL 2	
	OR(95%CI)	p value	OR(95%CI)	p value
Left hemisphere				
Total Hippocampus volume	0.97(0.79;1.17)	0.723	0.96(0.78;1.17)	0.659
Subfields				
Tail	**1.25(1.02;1.53)**	**0.032**	1.22(0.99;1.50)	0.063
Subiculum	1.16(0.91;1.47)	0.236	1.18(0.92;1.52)	0.187
Cornu Ammonis 1 (CA1)	0.92(0.67;1.28)	0.624	0.94(0.67;1.31)	0.704
Fissure	**1.23(1.06;1.42)**	**0.006**	**1.20(1.03;1.39)**	**0.017**
Presubiculum	0.90(0.74;1.09)	0.290	0.92(0.75;1.11)	0.376
Parasubiculum	**0.82(0.70;0.96)**	**0.012**	**0.82(0.70;0.96)**	**0.014**
Molecular layer	**1.27(1.04;1.55)**	**0.018**	1.17(0.95;1.44)	0.146
Dentate gyrus	**0.71(0.53;0.96)**	**0.025**	**0.74(0.54;1.00)**	**0.048**
Cornu Ammonis 2/3 (CA2/3)	0.86(0.71;1.03)	0.107	0.89(0.73;1.09)	0.262
Cornu Ammonis 4 (CA4)	0.85(0.61;1.17)	0.321	0.87(0.63;1.22)	0.425
Fimbria	0.90(0.76;1.07)	0.221	0.91(0.76;1.09)	0.323
HATA	0.85(0.71;1.01)	0.068	0.86(0.72;1.03)	0.110
Right hemisphere				
Total Hippocampus volume	0.93(0.76;1.13)	0.455	0.90(0.74;1.10)	0.318
Subfields				
Tail	1.05(0.85;1.30)	0.646	1.04(0.83;1.29)	0.751
subiculum	**1.30(1.01;1.67)**	**0.039**	**1.38(1.07;1.78)**	**0.014**
Cornu Ammonis 1 (CA1)	1.31(0.93;1.85)	0.125	1.29(0.91;1.83)	0.154
Fissure	**1.32(1.13;1.53)**	**p<0.001** *	**1.26(1.08;1.48)**	**0.003** *
Presubiculum	0.91(0.76;1.10)	0.342	0.94(0.78;1.14)	0.518
Parasubiculum	0.93(0.80;1.08)	0.353	0.93(0.79;1.09)	0.357
Molecular layer	**1.46(1.18;1.81)**	**p<0.001** *	**1.31(1.05;1.64)**	**0.015**
Dentate gyrus	**0.69(0.51;0.92)**	**0.012**	**0.69(0.51;0.93)**	**0.014**
Cornu Ammonis 2/3 (CA2/3)	**0.72(0.59;0.87)**	**p<0.001** *	**0.75(0.62;0.92)**	**0.006**
Cornu Ammonis 4 (CA4)	**0.70(0.52;0.94)**	**0.019**	**0.68(0.50;0.92)**	**0.013**
Fimbria	0.93(0.79;1.10)	0.394	0.95(0.80;1.13)	0.583
HATA	0.91(0.77;1.09)	0.321	0.91(0.75;1.09)	0.287

*Supplementary Table S4.4 | **Results, prevalent depression (PHQ-9≥10).** Table shows the results for logistic regression on prevalent depression. Results are presented in odds ratio (OR) and confidence intervals (CI). Bold shows nominally significant (p<0.05); Star (*) shows multiple comparison correction significant (p<0.0039). Abbreviations; HATA, Hippocampus-amygdala-transition-area. Model 1 was, adjusted for brain total volume when analyzing hippocampal total volumes, or adjusted for left/right hippocampus total volume, when analyzing hippocampal subfields, MRI lag time, age and sex; Model 2 was additionally adjusted for potential confounders: T2DM status, education level, waist circumference, history of cardiovascular disease, cholesterol ratio, use of alcohol and smoking status.*

Supplementary Table S4.5

	A. TRANSIENT COURSE				
	Model 1		Model 2		
Structure	OR(95%CI)	P value	OR(95%CI)	P value	
Left hemisphere					
Total Hippocampus volume	1.45(0.97;2.16)	0.068	1.46(0.96;2.21)	0.073	
Subfields					
Tail	1.11(0.74;1.69)	0.608	1.09(0.71;1.66)	0.698	
Subiculum	0.95(0.58;1.56)	0.849	0.92(0.55;1.55)	0.759	
Cornu Ammonis 1 (CA1)	1.27(0.65;2.46)	0.481	1.22(0.62;2.40)	0.569	
Fissure	0.95(0.68;1.33)	0.753	0.95(0.67;1.34)	0.758	
Presubiculum	0.88(0.59;1.30)	0.512	0.81(0.54;1.22)	0.320	
Parasubiculum	0.94(0.68;1.28)	0.680	0.90(0.65;1.25)	0.518	
Molecular layer	0.83(0.53;1.31)	0.429	0.76(0.47;1.23)	0.271	
Dentate gyrus	0.86(0.47;1.56)	0.612	0.91(0.49;1.70)	0.774	
Cornu Ammonis 2/3 (CA2/3)	1.05(0.72;1.53)	0.804	1.25(0.84;1.86)	0.281	
Cornu Ammonis 4 (CA4)	0.89(0.47;1.72)	0.735	0.97(0.49;1.89)	0.921	
Fimbria	1.15(0.81;1.62)	0.439	1.22(0.86;1.74)	0.270	
HATA	0.91(0.64;1.30)	0.605	0.90(0.62;1.30)	0.560	
Right hemisphere					
Total Hippocampus volume	1.26(0.85;1.88)	0.242	1.30(0.87;1.94)	0.206	
Subfields					
Tail	0.97(0.62;1.50)	0.877	0.98(0.63;1.53)	0.918	
Subiculum	1.04(0.63;1.72)	0.888	1.02(0.60;1.72)	0.953	
Cornu Ammonis 1 (CA1)	1.08(0.54;2.15)	0.824	1.12(0.55;2.26)	0.755	
Fissure	1.18(0.86;1.61)	0.318	1.14(0.82;1.58)	0.447	
Presubiculum	0.77(0.52;1.13)	0.181	0.75(0.50;1.11)	0.152	
Parasubiculum	1.11(0.82;1.51)	0.492	1.10(0.81;1.51)	0.534	
Molecular layer	1.29(0.83;2.00)	0.258	1.18(0.75;1.86)	0.477	
Dentate gyrus	0.99(0.56;1.75)	0.971	1.00(0.56;1.81)	0.989	
Cornu Ammonis 2/3 (CA2/3)	0.98(0.66;1.46)	0.940	1.08(0.72;1.64)	0.698	
Cornu Ammonis 4 (CA4)	1.09(0.61;1.94)	0.776	1.10(0.61;2.00)	0.749	
Fimbria	0.83(0.59;1.17)	0.276	0.83(0.58;1.18)	0.299	
HATA	1.25(0.88;1.79)	0.217	1.20(0.83;1.73)	0.340	

*Supplementary Table S4.5 | **Results, course of prevalent depression (chronic or transient).** Table shows the results for multinomial logistic regression analysis on prevalent depression when it presents A) a transient course or B) a chronic course. Reference group: No depression at baseline nor at follow up (n=3,608). Results are presented in odds ratios (OR) and confidence intervals (CI). Bold shows nominally significant (p<0.05); Star (*) shows multiple comparison correction significant (p<0.0039). Abbreviations: LH/RH, Left/right hemisphere; HC, Hippocampus; CA, Cornu Ammonis; HATA, Hippo-campus-amygdala-transition-area.*

	B. CHRONIC COURSE			
	Model 1		Model 2	
	OR(95%CI)	P value	OR(95%CI)	P value
	0.73(0.56;0.94)	**0.013**	**0.70(0.54;0.91)**	**0.009**
	1.29(0.99;1.68)	0.060	1.24(0.95;1.63)	0.114
	1.10(0.80;1.51)	0.562	1.17(0.84;1.61)	0.354
	0.99(0.65;1.50)	0.948	1.04(0.68;1.60)	0.849
	1.42(1.18;1.70)	**p<0.001** *	**1.37(1.14;1.64)**	**p<0.001** *
	0.85(0.66;1.09)	0.208	0.86(0.67;1.11)	0.258
	0.73(0.59;0.90)	**0.003** *	**0.74(0.60;0.91)**	**0.005**
	1.45(1.13;1.85)	**0.003** *	**1.34(1.04;1.74)**	**0.026**
	0.74(0.50;1.08)	0.120	0.75(0.51;1.12)	0.159
	0.83(0.65;1.06)	0.131	0.83(0.65;1.07)	0.158
	0.99(0.65;1.50)	0.953	1.02(0.67;1.58)	0.914
	0.77(0.62;0.97)	**0.025**	**0.78(0.61;0.98)**	**0.032**
	0.81(0.64;1.02)	0.071	0.81(0.64;1.04)	0.093
	0.70(0.54;0.90)	**0.006**	**0.68(0.52;0.87)**	**0.003** *
	1.16(0.88;1.52)	0.296	1.14(0.86;1.50)	0.375
	1.29(0.94;1.79)	0.118	1.37(0.99;1.91)	0.059
	1.29(0.83;2.00)	0.260	1.30(0.83;2.03)	0.252
	1.46(1.21;1.77)	**p<0.001** *	**1.40(1.15;1.71)**	**p<0.001** *
	1.01(0.79;1.28)	0.953	1.01(0.79;1.29)	0.938
	0.87(0.71;1.07)	0.181	0.87(0.71;1.07)	0.179
	1.66(1.27;2.19)	**p<0.001** *	**1.51(1.14;2.00)**	**0.004** *
	0.62(0.43;0.90)	**0.012**	**0.63(0.43;0.93)**	**0.019**
	0.60(0.47;0.77)	**p<0.001** *	**0.61(0.48;0.79)**	**p<0.001** *
	0.61(0.42;0.89)	**0.010**	**0.60(0.40;0.88)**	**0.010**
	0.97(0.78;1.20)	0.772	1.00(0.80;1.24)	0.973
	0.76(0.61;0.96)	**0.020**	**0.74(0.59;0.94)**	**0.014**

Supplementary Table S4.6

	Structure	Model 1		Model 2	
		HR(95%CI)	P value	HR(95%CI)	P value
LH	Total HC	0.96(0.83;1.12)	0.612	0.96(0.83;1.12)	0.644
	Tail	1.15(0.99;1.35)	0.066	1.15(0.99;1.35)	0.074
	Subiculum	**0.83(0.69;1.00)**	**0.047**	0.83(0.69;1.01)	0.057
	CA1	1.14(0.89;1.45)	0.305	1.09(0.85;1.39)	0.483
	Fissure	1.10(0.98;1.23)	0.118	1.05(0.94;1.18)	0.396
	Presubiculum	0.88(0.76;1.01)	0.070	0.88(0.76;1.02)	0.082
	Parasubiculum	1.03(0.92;1.15)	0.649	1.04(0.92;1.17)	0.536
	Molecular layer	1.15(0.99;1.33)	0.067	1.09(0.94;1.27)	0.264
	Dentate gyrus	**0.78(0.62;0.97)**	**0.028**	0.82(0.66;1.03)	0.092
	CA2/3	1.06(0.92;1.22)	0.398	1.10(0.95;1.27)	0.216
	CA4	0.99(0.77;1.26)	0.913	1.04(0.81;1.33)	0.771
	Fimbria	**0.85(0.75;0.97)**	**0.015**	**0.86(0.75;0.98)**	**0.026**
	HATA	0.88(0.77;1.01)	0.060	0.88(0.77;1.01)	0.074
RH	Total HC	1.05(0.91;1.22)	0.493	1.06(0.91;1.22)	0.467
	Tail	1.03(0.87;1.21)	0.754	1.03(0.88;1.22)	0.703
	Subiculum	0.84(0.69;1.01)	0.063	0.84(0.69;1.02)	0.071
	CA1	1.18(0.91;1.52)	0.213	1.18(0.91;1.53)	0.202
	Fissure	1.08(0.96;1.22)	0.194	1.05(0.93;1.18)	0.456
	Presubiculum	0.94(0.81;1.08)	0.365	0.94(0.82;1.09)	0.407
	Parasubiculum	0.99(0.88;1.11)	0.878	1.00(0.89;1.12)	0.992
	Molecular layer	**1.22(1.04;1.42)**	**0.015**	1.16(0.99;1.36)	0.073
	Dentate gyrus	0.92(0.74;1.13)	0.415	0.94(0.76;1.16)	0.567
	CA2/3	0.92(0.80;1.06)	0.264	0.94(0.81;1.09)	0.403
	CA4	0.99(0.80;1.22)	0.895	0.99(0.79;1.23)	0.897
	Fimbria	1.00(0.88;1.13)	0.993	0.99(0.87;1.12)	0.838
	HATA	0.97(0.85;1.11)	0.651	0.97(0.85;1.11)	0.689

*Supplementary Table S4.6 | **Results, incident depression (PHQ-9 ≥ 10).** Table shows the results of Cox regression analyses predicting risk of incident clinical depression. Results are presented in hazard ratios (HR) and confidence intervals (CI). Bold shows nominally significant (p<0.05); Star (*) shows multiple comparison correction significant (p<0.0039). Abbreviations: LH/RH, Left/right hemisphere; HC, Hippocampus; CA, Cornu Ammonis; Dentate gyrus, HATA, Hippocampus-amygdala-transition-area.*

Supplementary Table S4.7

	Structure	MAIN		A) ONE INCIDENT EVENT		B) ONE OR MORE INCIDENT EVENTS		
		HR(95%CI)	P value	HR(95%CI)	P value	HR(95%CI)	P value	
LH	Total HC	0.96(0.83;1.12)	0.644	0.87(0.72;1.04)	0.127	1.22(0.93;1.60)	0.161	
	Tail	1.15(0.99;1.35)	0.074	1.10(0.91;1.33)	0.324	1.29(0.98;1.69)	0.069	
	Subiculum	0.83(0.69;1.01)	0.057	0.95(0.75;1.19)	0.631	**0.62(0.44;0.87)**	**0.005**	
	CA1	1.09(0.85;1.39)	0.483	0.99(0.74;1.33)	0.955	1.36(0.89;2.10)	0.159	
	Fissure	1.05(0.94;1.18)	0.396	1.00(0.87;1.16)	0.967	1.17(0.96;1.43)	0.122	
	Presubiculum	0.88(0.76;1.02)	0.082	0.95(0.80;1.13)	0.573	**0.75(0.58;0.96)**	**0.024**	
	Parasubiculum	1.04(0.92;1.17)	0.536	1.01(0.88;1.16)	0.895	1.10(0.89;1.35)	0.366	
	Molecular layer	1.09(0.94;1.27)	0.264	1.08(0.90;1.31)	0.402	1.12(0.85;1.48)	0.418	
	Dentate gyrus	0.82(0.66;1.03)	0.092	0.80(0.61;1.06)	0.119	0.83(0.56;1.24)	0.366	
	CA2/3	1.10(0.95;1.27)	0.216	1.04(0.87;1.24)	0.660	1.22(0.95;1.57)	0.127	
	CA4	1.04(0.81;1.33)	0.771	0.95(0.70;1.28)	0.726	1.21(0.79;1.87)	0.381	
	Fimbria	**0.86(0.75;0.98)**	**0.026**	0.91(0.77;1.06)	0.230	**0.76(0.61;0.97)**	**0.024**	
	HATA	0.88(0.77;1.01)	0.074	0.99(0.84;1.17)	0.938	**0.70(0.55;0.89)**	**0.003**	*
RH	Total HC	1.06(0.91;1.22)	0.467	1.00(0.84;1.20)	0.986	1.19(0.92;1.54)	0.187	
	Tail	1.03(0.88;1.22)	0.703	1.01(0.83;1.23)	0.930	1.09(0.82;1.45)	0.563	
	Subiculum	0.84(0.69;1.02)	0.071	0.89(0.71;1.13)	0.344	0.72(0.52;1.01)	0.057	
	CA1	1.18(0.91;1.53)	0.202	1.04(0.76;1.42)	0.817	**1.60(1.02;2.52)**	**0.041**	
	Fissure	1.05(0.93;1.18)	0.456	1.00(0.86;1.16)	0.989	1.17(0.95;1.44)	0.137	
	Presubiculum	0.94(0.82;1.09)	0.407	0.98(0.82;1.16)	0.800	0.87(0.68;1.11)	0.249	
	Parasubiculum	1.00(0.89;1.12)	0.992	0.95(0.83;1.09)	0.466	1.11(0.91;1.34)	0.306	
	Molecular layer	1.16(0.99;1.36)	0.073	1.18(0.98;1.44)	0.087	1.12(0.84;1.49)	0.432	
	Dentate gyrus	0.94(0.76;1.16)	0.567	0.96(0.74;1.25)	0.760	0.87(0.60;1.28)	0.488	
	CA2/3	0.94(0.81;1.09)	0.403	0.95(0.79;1.13)	0.538	0.93(0.71;1.21)	0.592	
	CA4	0.99(0.79;1.23)	0.897	1.03(0.79;1.34)	0.847	0.90(0.61;1.32)	0.581	
	Fimbria	0.99(0.87;1.12)	0.838	1.01(0.86;1.17)	0.937	0.94(0.75;1.17)	0.573	
	HATA	0.97(0.85;1.11)	0.689	0.96(0.81;1.13)	0.620	1.00(0.79;1.26)	0.986	

*Supplementary Table S4.7 | **Results, course of incident depression (chronic or transient).** Table shows the results of Cox regression analyses predicting risk of incident clinical depression stratified by number of incident events as A) transient (only one incident event), and B) chronic (more than one incident events) clinically relevant depressive symptoms. Results are presented in hazard ratios (HR) and confidence intervals (CI). Bold shows nominally significant (p<0.05); Star (*) shows multiple comparison correction significant (p<0.0039). Abbreviations: LH/RH, Left/right hemisphere; HC, Hippocampus; CA, Cornu Ammonis; Dentate gyrus, HATA, Hippocampus-amygdala-transition-area.*

Supplementary Table S4.8

	Structure	MAIN		A. EXCLUSION T2DM		
		RR(95%CI)	p value	RR(95%CI)	p value	
LH	Total HC	0.98(0.94;1.03)	0.477	0.99(0.94;1.04)	0.676	
	Tail	1.02(0.97;1.07)	0.441	1.02(0.97;1.08)	0.463	
	Subiculum	1.03(0.97;1.10)	0.279	1.02(0.95;1.09)	0.577	
	CA1	1.01(0.93;1.09)	0.779	1.03(0.94;1.12)	0.575	
	Fissure	1.04(1.00;1.08)	0.054	1.03(0.98;1.07)	0.237	
	Presubiculum	0.98(0.94;1.03)	0.526	0.98(0.93;1.03)	0.426	
	Parasubiculum	0.98(0.94;1.02)	0.287	0.97(0.93;1.01)	0.181	
	Molecular layer	**1.06(1.01;1.12)**	**0.029**	1.06(1.00;1.12)	0.050	
	Dentate gyrus	**0.93(0.86;1.00)**	**0.040**	0.92(0.85;1.00)	0.055	
	CA2/3	0.96(0.91;1.00)	0.062	0.97(0.92;1.02)	0.178	
	CA4	0.96(0.88;1.04)	0.261	0.96(0.87;1.04)	0.308	
	Fimbria	0.98(0.94;1.03)	0.436	0.99(0.94;1.03)	0.564	
	HATA	1.00(0.96;1.05)	0.917	1.01(0.96;1.06)	0.780	
RH	Total HC	0.97(0.92;1.01)	0.170	0.97(0.92;1.02)	0.227	
	Tail	0.99(0.94;1.05)	0.847	1.00(0.94;1.06)	0.980	
	Subiculum	1.05(0.99;1.12)	0.133	1.04(0.97;1.11)	0.233	
	CA1	1.07(0.98;1.16)	0.120	1.06(0.97;1.16)	0.223	
	Fissure	**1.04(1.00;1.08)**	**0.040**	1.05(1.00;1.09)	0.052	
	Presubiculum	1.00(0.96;1.05)	0.899	0.99(0.94;1.04)	0.726	
	Parasubiculum	1.00(0.96;1.04)	0.976	1.00(0.96;1.04)	0.945	
	Molecular layer	1.08(1.02;1.14)	0.006	**1.08(1.01;1.15)**	**0.016**	
	Dentate gyrus	0.90(0.84;0.97)	0.005	**0.92(0.85;0.99)**	**0.026**	
	CA2/3	0.94(0.89;0.98)	0.007	0.95(0.90;1.01)	0.080	
	CA4	**0.89(0.83;0.96)**	**0.002** *	**0.92(0.85;1.00)**	**0.048**	
	Fimbria	1.00(0.96;1.04)	0.898	0.98(0.94;1.03)	0.476	
	HATA	1.00(0.95;1.04)	0.939	0.98(0.93;1.03)	0.445	

*Supplementary Table S4.8 | **Sensitivity analysis, prevalent depressive symptoms (PHQ-9 score).** Sensitivity analyses are performed over Model 2. First column shows main results of Model 2. Panel A) shows the results after the exclusion of participants with T2DM (in this case, there is no adjustment for T2DM; panel B) After additionally adjusting Model 2 for antidepressant medication; panel C) after excluding participants using antidepressant medication; and panel D) after excluding participants with a lifetime of major depressive disorder diagnosis. Results are presented in rate ratio (RR) and confidence intervals (CI). Bold **shows nominally significant (p<0.05); Star (*) shows multiple comparison correction significant (p<0.0039). Abbreviations: LH/RH, Left/right hemisphere; HC, Hippocampus; CA, Cornu Ammonis; HATA, Hippocampus-amygdala-transition-area.***

	B. ADJUSTING FOR ANTIDEPRESSANT MEDICATION		C. EXCLUSION ANTIDEPRESSANT MEDICATION		D. EXCLUSION PAST DEPRESSION	
	RR(95%CI)	p value	RR(95%CI)	p value	RR(95%CI)	p value
	0.99(0.95;1.04)	0.751	1.00(0.95;1.05)	0.974	0.98(0.92;1.04)	0.492
	1.02(0.97;1.07)	0.469	1.01(0.96;1.06)	0.714	1.00(0.94;1.06)	0.916
	1.03(0.97;1.09)	0.404	1.03(0.96;1.09)	0.441	1.05(0.97;1.13)	0.216
	1.02(0.94;1.10)	0.638	1.02(0.94;1.11)	0.569	1.06(0.96;1.17)	0.274
	1.03(1.00;1.07)	0.082	1.03(0.99;1.07)	0.143	**1.06(1.01;1.11)**	**0.023**
	0.98(0.94;1.03)	0.439	0.99(0.94;1.03)	0.564	0.99(0.93;1.05)	0.691
	0.98(0.94;1.02)	0.288	0.98(0.94;1.02)	0.367	0.99(0.95;1.04)	0.686
	1.05(1.00;1.11)	0.053	1.04(0.99;1.10)	0.112	1.05(0.98;1.11)	0.153
	0.93(0.86;1.00)	**0.044**	0.94(0.87;1.01)	0.111	0.94(0.85;1.03)	0.147
	0.97(0.92;1.01)	0.136	0.97(0.92;1.02)	0.166	0.95(0.90;1.01)	0.075
	0.96(0.88;1.04)	0.268	0.96(0.88;1.05)	0.354	0.97(0.88;1.07)	0.501
	0.99(0.95;1.03)	0.528	0.99(0.95;1.03)	0.656	0.97(0.92;1.02)	0.236
	1.00(0.96;1.05)	0.948	1.01(0.96;1.05)	0.823	1.01(0.96;1.07)	0.678
	0.98(0.93;1.03)	0.391	0.99(0.94;1.04)	0.574	0.97(0.91;1.03)	0.276
	0.99(0.94;1.04)	0.743	0.98(0.93;1.04)	0.569	1.02(0.95;1.09)	0.574
	1.04(0.98;1.10)	0.217	1.04(0.97;1.10)	0.277	1.05(0.98;1.14)	0.185
	1.05(0.97;1.14)	0.226	1.05(0.96;1.14)	0.316	1.04(0.94;1.15)	0.443
	1.04(1.00;1.08)	0.054	1.04(0.99;1.08)	0.099	**1.05(1.00;1.11)**	**0.037**
	1.01(0.96;1.05)	0.828	1.01(0.96;1.06)	0.669	1.03(0.97;1.09)	0.278
	1.00(0.96;1.03)	0.863	0.99(0.95;1.03)	0.744	1.03(0.98;1.08)	0.221
	1.07(1.01;1.13)	**0.020**	1.06(1.00;1.12)	0.064	1.05(0.98;1.13)	0.137
	0.92(0.86;0.99)	**0.019**	0.94(0.87;1.01)	0.090	**0.88(0.81;0.96)**	**0.005**
	0.95(0.90;0.99)	**0.023**	0.95(0.91;1.00)	0.057	**0.93(0.88;0.99)**	**0.018**
	0.91(0.85;0.98)	**0.009**	**0.93(0.86;1.00)**	**0.045**	**0.89(0.81;0.96)**	**0.006**
	1.01(0.97;1.05)	0.751	1.01(0.96;1.05)	0.753	0.99(0.94;1.04)	0.633
	1.00(0.96;1.05)	0.876	1.01(0.96;1.06)	0.763	0.98(0.93;1.04)	0.559

Supplementary Table S4.9

	Structure	MAIN		A. EXCLUSION T2DM		
		OR(95%CI)	p value	OR(95%CI)	p value	
LH	Total HC	0.96(0.78;1.17)	0.659	1.01(0.80;1.28)	0.930	
	Tail	1.22(0.99;1.50)	0.063	1.19(0.93;1.51)	0.161	
	Subiculum	1.18(0.92;1.52)	0.187	1.13(0.85;1.51)	0.408	
	CA1	0.94(0.67;1.31)	0.704	0.98(0.66;1.45)	0.920	
	Fissure	**1.20(1.03;1.39)**	**0.017**	1.14(0.95;1.37)	0.169	
	Presubiculum	0.92(0.75;1.11)	0.376	0.88(0.70;1.11)	0.295	
	Parasubiculum	**0.82(0.70;0.96)**	**0.014**	**0.79(0.65;0.95)**	**0.013**	
	Molecular layer	1.17(0.95;1.44)	0.146	1.19(0.94;1.51)	0.157	
	Dentate gyrus	**0.74(0.54;1.00)**	**0.048**	0.79(0.55;1.12)	0.179	
	CA2/3	0.89(0.73;1.09)	0.262	0.92(0.73;1.15)	0.443	
	CA4	0.87(0.63;1.22)	0.425	0.93(0.64;1.37)	0.732	
	Fimbria	0.91(0.76;1.09)	0.323	0.89(0.73;1.10)	0.292	
	HATA	0.86(0.72;1.03)	0.110	0.93(0.75;1.15)	0.499	
RH	Total HC	0.90(0.74;1.10)	0.318	0.93(0.74;1.18)	0.568	
	Tail	1.04(0.83;1.29)	0.751	1.01(0.79;1.30)	0.937	
	Subiculum	**1.38(1.07;1.78)**	**0.014**	1.27(0.95;1.71)	0.113	
	CA1	1.29(0.91;1.83)	0.154	1.41(0.93;2.14)	0.104	
	Fissure	**1.26(1.08;1.48)**	**0.003** *	**1.28(1.06;1.54)**	**0.011**	
	Presubiculum	0.94(0.78;1.14)	0.518	0.84(0.67;1.05)	0.126	
	Parasubiculum	0.93(0.79;1.09)	0.357	0.91(0.75;1.09)	0.294	
	Molecular layer	1.31(1.05;1.64)	0.015	**1.38(1.06;1.79)**	**0.016**	
	Dentate gyrus	0.69(0.51;0.93)	0.014	0.78(0.56;1.11)	0.167	
	CA2/3	0.75(0.62;0.92)	0.006	0.79(0.63;1.00)	0.053	
	CA4	0.68(0.50;0.92)	0.013	0.79(0.56;1.13)	0.196	
	Fimbria	0.95(0.80;1.13)	0.583	0.92(0.75;1.13)	0.427	
	HATA	0.91(0.75;1.09)	0.287	0.89(0.72;1.09)	0.261	

*Supplementary Table S4.9 | **Sensitivity analysis, prevalent depression (PHQ-9≥10).** Sensitivity analyses are performed over Model 2. First column shows main results of Model 2. Panel A) shows the results after the exclusion of participants with T2DM (in this case, there is no adjustment for T2DM; panel B) After additionally adjusting Model 2 for antidepressant medication; panel C) after excluding participants using antidepressant medication; and panel D) after excluding participants with a lifetime of major depressive disorder diagnosis. Results are presented in odds ratio (OR) and confidence intervals (CI). Bold shows nominally significant (p<0.05); Star (*) shows multiple comparison correction significant (p<0.0039). Abbreviations: LH/RH, Left/right hemisphere; HC, Hippocampus; CA, Cornu Ammonis; HATA, Hippocampus-amygdala-transition-area.*
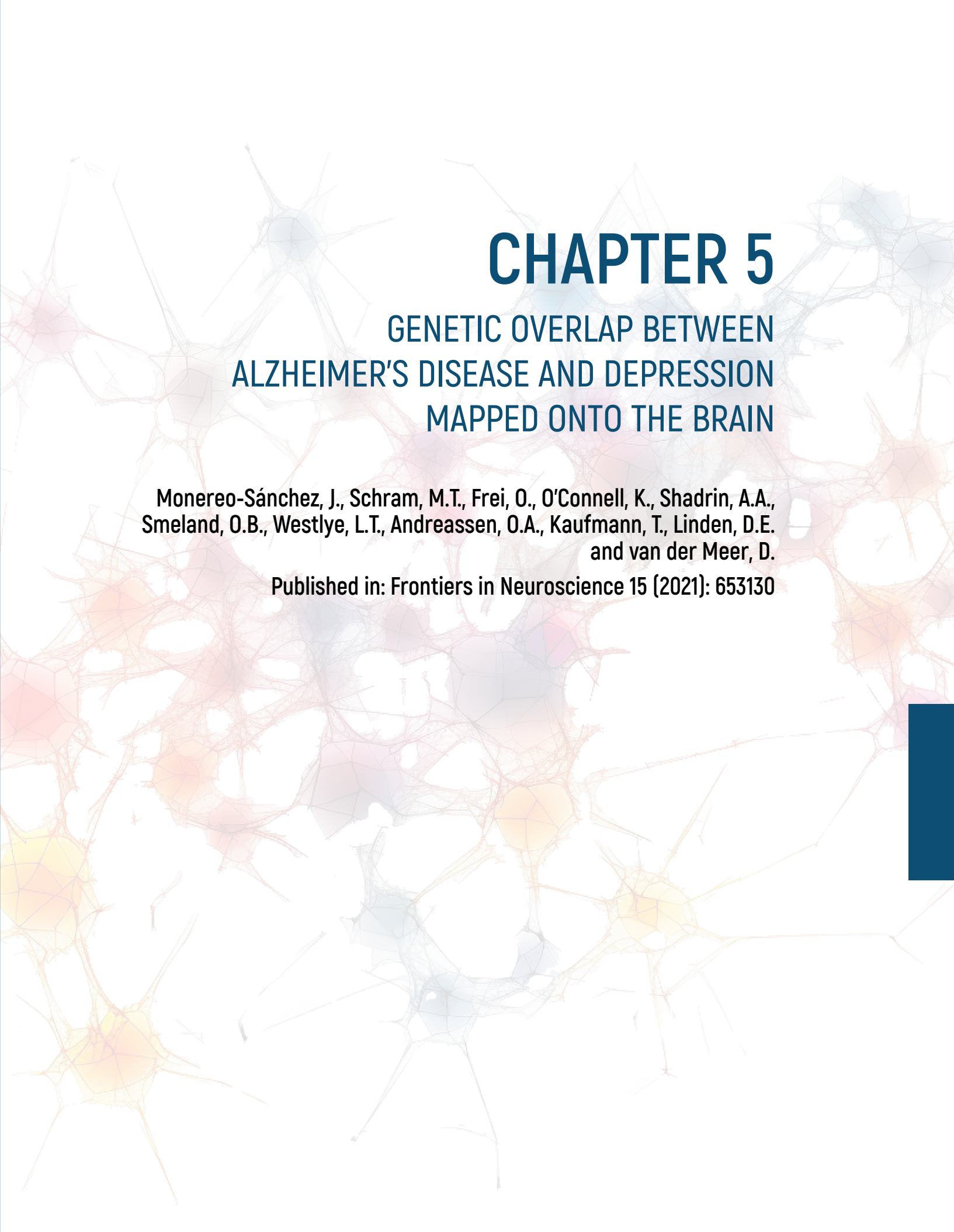
	B. ADJUSTING FOR ANTIDEPRESSANT MEDICATION		C. EXCLUSION ANTIDEPRESSANT MEDICATION		D. EXCLUSION PAST DEPRESSION	
	OR(95%CI)	p value	OR(95%CI)	p value	OR(95%CI)	p value
	0.97(0.79;1.19)	0.787	1.07(0.85;1.35)	0.568	1.11(0.75;1.62)	0.606
	1.23(1.00;1.52)	0.052	1.18(0.93;1.49)	0.183	0.92(0.62;1.36)	0.680
	1.16(0.90;1.50)	0.244	1.17(0.88;1.55)	0.281	1.53(0.97;2.40)	0.068
	0.97(0.69;1.36)	0.857	0.96(0.66;1.41)	0.848	1.06(0.56;2.01)	0.865
	1.19(1.02;1.39)	**0.023**	**1.20(1.01;1.42)**	**0.039**	**1.30(1.00;1.68)**	**0.049**
	0.89(0.73;1.09)	0.272	0.95(0.76;1.19)	0.666	0.91(0.63;1.32)	0.609
	0.81(0.69;0.95)	**0.011**	**0.78(0.65;0.94)**	**0.008**	0.79(0.58;1.06)	0.114
	1.16(0.94;1.44)	0.175	1.09(0.86;1.38)	0.488	1.12(0.76;1.65)	0.565
	0.72(0.53;0.98)	**0.037**	0.77(0.54;1.09)	0.142	0.96(0.53;1.72)	0.879
	0.92(0.75;1.12)	0.383	0.96(0.77;1.20)	0.702	0.86(0.60;1.24)	0.427
	0.87(0.62;1.21)	0.398	0.92(0.63;1.34)	0.654	1.11(0.59;2.08)	0.741
	0.91(0.76;1.10)	0.329	0.92(0.75;1.13)	0.409	0.96(0.68;1.34)	0.800
	0.85(0.71;1.03)	0.094	0.87(0.70;1.07)	0.186	0.99(0.70;1.42)	0.977
	0.93(0.76;1.14)	0.465	1.02(0.81;1.28)	0.881	1.09(0.75;1.57)	0.661
	1.05(0.84;1.31)	0.684	0.96(0.75;1.23)	0.760	1.19(0.77;1.82)	0.432
	1.34(1.04;1.74)	**0.026**	**1.37(1.03;1.83)**	**0.033**	1.48(0.92;2.38)	0.108
	1.24(0.87;1.77)	0.233	1.28(0.86;1.91)	0.228	1.18(0.62;2.25)	0.619
	1.26(1.07;1.48)	**0.005**	**1.25(1.04;1.49)**	**0.015**	**1.38(1.04;1.83)**	**0.026**
	0.93(0.77;1.13)	0.468	0.96(0.77;1.19)	0.710	0.96(0.67;1.38)	0.845
	0.93(0.79;1.09)	0.363	0.88(0.74;1.05)	0.166	1.09(0.81;1.45)	0.573
	1.28(1.02;1.60)	**0.036**	1.17(0.91;1.51)	0.230	1.28(0.85;1.92)	0.238
	0.71(0.52;0.96)	**0.025**	0.80(0.57;1.13)	0.206	**0.53(0.30;0.94)**	**0.029**
	0.78(0.63;0.95)	**0.015**	0.83(0.66;1.04)	0.111	0.73(0.50;1.06)	0.094
	0.70(0.52;0.95)	**0.024**	0.79(0.56;1.11)	0.178	**0.56(0.32;0.99)**	**0.048**
	0.98(0.82;1.17)	0.817	0.95(0.78;1.15)	0.597	0.90(0.65;1.26)	0.546
	0.92(0.77;1.11)	0.386	0.96(0.78;1.19)	0.725	0.90(0.63;1.27)	0.542

Supplementary Table S4.10

| | Structure | MAIN | | A. EXCLUSION T2DM | | |
		OR(95%CI)	P value	OR(95%CI)	P value	
LH	Total HC	**0.70(0.54;0.91)**	**0.009**	0.75(0.56;1.02)	0.063	
	Tail	1.24(0.95;1.63)	0.114	1.19(0.87;1.62)	0.282	
	Subiculum	1.17(0.84;1.61)	0.354	1.22(0.84;1.77)	0.293	
	CA1	1.04(0.68;1.60)	0.849	1.03(0.62;1.71)	0.907	
	Fissure	**1.37(1.14;1.64)**	**p<0.001** *	**1.25(1.00;1.58)**	**0.055**	
	Presubiculum	0.86(0.67;1.11)	0.258	0.85(0.63;1.14)	0.287	
	Parasubiculum	**0.74(0.60;0.91)**	**0.005**	**0.74(0.58;0.95)**	**0.017**	
	Molecular layer	**1.34(1.04;1.74)**	**0.026**	**1.36(1.01;1.82)**	**0.042**	
	Dentate gyrus	0.75(0.51;1.12)	0.159	0.81(0.51;1.27)	0.350	
	CA2/3	0.83(0.65;1.07)	0.158	0.81(0.60;1.08)	0.151	
	CA4	1.02(0.67;1.58)	0.914	1.03(0.63;1.70)	0.894	
	Fimbria	**0.78(0.61;0.98)**	**0.032**	0.79(0.60;1.03)	0.082	
	HATA	0.81(0.64;1.04)	0.093	0.89(0.68;1.17)	0.410	
RH	Total HC	**0.68(0.52;0.87)**	**0.003** *	**0.67(0.50;0.91)**	**0.010**	
	Tail	1.14(0.86;1.50)	0.375	1.14(0.82;1.57)	0.431	
	Subiculum	1.37(0.99;1.91)	0.059	1.32(0.90;1.93)	0.157	
	CA1	1.30(0.83;2.03)	0.252	1.36(0.80;2.33)	0.255	
	Fissure	**1.40(1.15;1.71)**	**p<0.001** *	**1.37(1.08;1.74)**	**0.009**	
	Presubiculum	1.01(0.79;1.29)	0.938	0.89(0.67;1.19)	0.436	
	Parasubiculum	0.87(0.71;1.07)	0.179	0.85(0.67;1.07)	0.168	
	Molecular layer	**1.51(1.14;2.00)**	**0.004** *	**1.61(1.16;2.23)**	**0.005**	
	Dentate gyrus	**0.63(0.43;0.93)**	**0.019**	0.70(0.45;1.10)	0.120	
	CA2/3	**0.61(0.48;0.79)**	**p<0.001** *	**0.65(0.48;0.87)**	**0.004**	
	CA4	**0.60(0.40;0.88)**	**0.010**	0.68(0.43;1.07)	0.093	
	Fimbria	1.00(0.80;1.24)	0.973	0.94(0.73;1.22)	0.648	
	HATA	**0.74(0.59;0.94)**	**0.014**	**0.73(0.55;0.96)**	**0.022**	

*Supplementary Table S4.10 | **Sensitivity analysis, chronic depressive symptoms.** Sensitivity analyses are performed over Model 2. First column shows main results of Model 2. Panel A) shows the results after the exclusion of participants with T2DM (in this case, there is no adjustment for T2DM; panel B) After additionally adjusting Model 2 for antidepressant medication; panel C) after excluding participants using antidepressant medication; and panel D) after excluding participants with a lifetime of major depressive disorder diagnosis. Results are presented in odds ratio (OR) and confidence intervals (CI). Bold shows nominally significant (p<0.05); Star (*) shows multiple comparison correction significant (p<0.0039). Abbreviations: LH/RH, Left/right hemisphere; HC, Hippocampus; CA, Cornu Ammonis; HATA, Hippocampus-amygdala-transition-area.*

	B. ADJUSTING FOR ANTIDEPRESSANT MEDICATION		C. EXCLUSION ANTIDEPRESSANT MEDICATION		D. EXCLUSION PAST DEPRESSION	
	OR(95%CI)	P value	OR(95%CI)	P value	OR(95%CI)	P value
	0.71(0.54;0.93)	**0.012**	0.76(0.56;1.03)	0.079	0.74(0.43;1.26)	0.266
	1.27(0.97;1.67)	0.088	1.19(0.86;1.65)	0.282	1.13(0.64;2.00)	0.684
	1.13(0.81;1.57)	0.471	1.08(0.73;1.58)	0.704	**1.95(1.01;3.76)**	**0.047**
	1.10(0.71;1.71)	0.671	1.07(0.64;1.78)	0.798	1.14(0.45;2.87)	0.778
	1.38(1.14;1.66)	**p<0.001** *	**1.39(1.13;1.71)**	**0.002** *	**1.49(1.07;2.08)**	**0.018**
	0.84(0.65;1.09)	0.189	0.92(0.68;1.24)	0.580	0.77(0.45;1.32)	0.350
	0.72(0.58;0.90)	**0.004** *	**0.68(0.53;0.87)**	**0.003** *	0.66(0.43;1.01)	0.058
	1.35(1.03;1.77)	**0.029**	1.34(1.00;1.80)	0.053	1.59(0.99;2.54)	0.054
	0.73(0.49;1.09)	0.122	0.74(0.46;1.18)	0.200	0.69(0.29;1.63)	0.398
	0.85(0.66;1.10)	0.220	0.85(0.63;1.15)	0.295	0.65(0.38;1.10)	0.106
	1.02(0.66;1.58)	0.933	1.02(0.61;1.69)	0.951	1.18(0.47;2.96)	0.728
	0.77(0.61;0.97)	**0.027**	0.82(0.62;1.08)	0.151	**0.57(0.34;0.94)**	**0.029**
	0.80(0.63;1.02)	0.074	0.90(0.68;1.20)	0.487	0.91(0.55;1.53)	0.731
	0.69(0.53;0.90)	**0.007**	0.74(0.55;1.01)	0.056	0.80(0.47;1.36)	0.417
	1.15(0.87;1.54)	0.329	1.05(0.76;1.46)	0.764	1.38(0.76;2.51)	0.290
	1.34(0.95;1.88)	0.096	1.27(0.86;1.88)	0.222	**2.08(1.05;4.13)**	**0.036**
	1.25(0.79;1.97)	0.333	1.32(0.77;2.24)	0.309	1.57(0.63;3.87)	0.331
	1.42(1.16;1.75)	**p<0.001** *	**1.45(1.15;1.82)**	**0.002** *	**1.73(1.18;2.53)**	**0.005**
	1.01(0.79;1.29)	0.939	1.01(0.76;1.35)	0.930	1.02(0.61;1.71)	0.929
	0.87(0.71;1.07)	0.176	0.80(0.63;1.02)	0.066	1.17(0.77;1.77)	0.459
	1.47(1.10;1.97)	**0.009**	**1.40(1.01;1.94)**	**0.046**	1.66(0.95;2.91)	0.074
	0.64(0.44;0.95)	**0.027**	0.72(0.46;1.14)	0.164	**0.33(0.15;0.76)**	**0.009**
	0.63(0.49;0.82)	**p<0.001** *	**0.68(0.51;0.92)**	**0.013**	**0.49(0.29;0.82)**	**0.007**
	0.61(0.41;0.91)	**0.015**	0.68(0.43;1.08)	0.099	**0.40(0.18;0.90)**	**0.027**
	1.03(0.82;1.29)	0.814	1.04(0.80;1.35)	0.775	**0.65(0.41;1.05)**	**0.076**
	0.77(0.61;0.98)	**0.036**	0.80(0.60;1.06)	0.119	0.70(0.43;1.16)	0.168

CHAPTER 5

GENETIC OVERLAP BETWEEN ALZHEIMER'S DISEASE AND DEPRESSION MAPPED ONTO THE BRAIN

Monereo-Sánchez, J., Schram, M.T., Frei, O., O'Connell, K., Shadrin, A.A., Smeland, O.B., Westlye, L.T., Andreassen, O.A., Kaufmann, T., Linden, D.E. and van der Meer, D.

Abstract

Alzheimer's disease (AD) and depression are debilitating brain disorders that are often comorbid. Shared brain mechanisms have been implicated, yet findings are inconsistent, reflecting the complexity of the underlying pathophysiology. As both disorders are (partly) heritable, characterizing their genetic overlap may provide etiological clues. While previous studies have indicated negligible genetic correlations, this study aims to expose the genetic overlap that may remain hidden due to mixed directions of effects.

We applied Gaussian mixture modelling, through MiXeR, and conjunctional false discovery rate (cFDR) analysis, through pleioFDR, to genome-wide association study (GWAS) summary statistics of AD (n = 79,145) and depression (n = 450,619). The effects of identified overlapping loci on AD and depression were tested in 403,029 participants of the UK Biobank (UKB) (mean age 57.21, 52.0% female), and mapped onto brain morphology in 30,699 individuals with brain MRI data.

MiXer estimated 98 causal genetic variants overlapping between the 2 disorders, with 0.44 concordant directions of effects. Through pleioFDR, we identified a SNP in the TMEM106B gene, which was significantly associated with AD (B = -0.002, p = 9.1 x 10^{-4}) and depression (B = 0.007, p = 3.2 x 10^{-9}) in the UKB. This SNP was also associated with several regions of the corpus callosum volume anterior (B > 0.024, p < 8.6 x 10^{-4}), third ventricle volume ventricle (B = -0.025, p = 5.0 x 10^{-6}), and inferior temporal gyrus surface area (B = 0.017, p = 5.3 x 10^{-4}).

Our results indicate there is substantial genetic overlap, with mixed directions of effects, between AD and depression. These findings illustrate the value of biostatistical tools that capture such overlap, providing insight into the genetic architectures of these disorders.

Introduction

Alzheimer's disease (AD) is a highly disabling neurodegenerative disease characterized by memory loss and a gradual cognitive, functional, and behavioral decline (Reitz & Mayeux, 2014). Its prevalence increases rapidly with age, affecting 13% of the population at age 80, and 37% of the population at age 90 (von Strauss, Viitanen, De Ronchi, Winblad, & Fratiglioni, 1999). Individuals with AD often have comorbid major depressive disorder (MDD), present in 22– 59% of cases (Starkstein, Jorge, Mizrahi, & Robinson, 2005; Zubenko et al., 2003), while MDD has an estimated lifetime prevalence of 11–15% in the general population (Bromet et al., 2011). MDD is a heterogeneous disorder; in addition to the core symptoms of low mood, anhedonia, and loss of energy, it comprises behavioral, physiological, and psychological signs and symptoms that include changes in appetite, sleeping, and psychomotor patterns, fatigue, lack of concentration, feelings of worthlessness or guilt, and suicidal ideation (American Psychiatric Association, 2013).

It has been long discussed whether a history of depressive symptoms is a risk factor for later development of AD, or rather an early prodromal manifestation of AD (Chen, Ganguli, Mulsant, & DeKosky, 1999; Devanand et al., 1996). While bidirectional effects between the two disorders is likely, there is more evidence that midlife onset depressive symptoms and/or MDD are a risk factor for AD than vice versa (Barnes et al., 2012; Gracia-García et al., 2015; Green et al., 2003; Jorm, 2001; Ownby, Crocco, Acevedo, John, & Loewenstein, 2006; R. S. Wilson et al., 2002). Furthermore, AD patients with depressive symptoms show accelerated cognitive decline and neurodegeneration, with significantly more plaques and tangles in the hippocampus than non-depressed individuals with AD (Rapp et al., 2006), while AD symptom count (Verkaik, Nuyen, Schellevis, & Francke, 2007) or tau pathology (Kramberger et al., 2012) does not appear to contribute to the incidence or severity of depressive disorders.

Neuroimaging studies have provided scattered evidence that AD and depressive disorders share neurobiological pathways. Early stage AD is associated with atrophy of the hippocampus, para-hippocampal regions (Jack et al., 1999), and temporo-parietal cortex (Acharya et al., 2019), with atrophy becoming generalized in later stages of the disease, including cortical thinning in primary motor and sensory regions (Fox et al., 2001; Sabuncu et al., 2011). Similarly, MDD and recurrent major depression (MD) are related to smaller hippocampal volumes(Bremner et al., 2000; Mervaala et al., 2000; Sheline, Sanghavi, Mintun, & Gado, 1999), amygdala, and parahippocampal areas (Andreescu et al., 2008) as well as lower cortical thickness in medial orbitofrontal cortex, fusiform gyrus, insula, rostral, and caudal anterior and posterior cingulate cortex, temporal lobe in MDD (Schmaal et al., 2020), many of these changes correlating positively with the duration of the disease (Andreescu et al., 2008). In AD patients with comorbid symptoms of depression, MRI studies have shown specifically thinner cortex in temporal and parietal areas when compared to non-depressed AD patients (Lebedeva et al., 2014). Conjunction analysis on the brain morphological changes that overlap between AD and late-life onset depression has shown that, in addition to the previously mentioned structures, both conditions are associated with hippocampal atrophy(Boccia, Acierno, & Piccardi, 2015). Yet, the risk of developing AD in MDD does not seem to be mediated by hippocampal or amygdala volumes (Geerlings, den Heijer, Koudstaal, Hofman, & Breteler, 2008).

Both AD and depressive disorders are heritable, with twin studies indicating 37% broad heritability for MDD (Sullivan, Neale, & Kendler, 2000) and 74% for AD (Gatz et al., 1997). Molecular genetics studies show that both disorders have complex genetic architectures. AD has recently been characterized as oligogenic, with estimates indicating the involvement of relatively few genetic variants, in addition to the well-known, strong APOE-e4 risk variant (Holland et al., 2020; Zhang et al., 2020). MDD on the other hand has been estimated to be the most polygenic of all major brain disorders (Holland et al., 2020), involving many genetic variants with small effects that explain a small amount of its heritability (Wray et al., 2018). Regardless, given the high comorbidity and indications of shared neurobiological pathways, substantial genetic overlap is to be expected, which may be leveraged to better understand these disorders (Smeland et al., 2020). Indeed, several candidate gene studies have identified shared genetic risk factors (Ye, Bai, & Zhang, 2016) that implicate hypothesized shared mechanisms, such as chronic neuroinflammatory changes in the brain (Leszek et al., 2018). While negligible genetic overlap between AD and MDD has been reported (Gibson et al., 2017; Lutz, Sprague, Barrera, & Chiba-Falek, 2020), substantial genetic overlap may remain hidden from measures of global genetic correlation due to mixed directions of effects. Here, we assess the genetic overlap between AD and depression across the genome through tools that capture the extent of overlap or specific loci, regardless of directions of effect. This was followed-up by analyses of the associations between shared loci and regional brain morphology in the UK Biobank (UKB) population study, providing valuable insights into their shared neurobiology.

Materials and Methods

GWAS Summary Statistics

To investigate the genetics of AD, we made use of the phase 1 summary statistics from a recent genome-wide association study (GWAS) that combined samples from the Psychiatric Genomics Consortium (PGC), the International Genomics of Alzheimer's Project (IGAP), and the Alzheimer's Disease Sequencing Project (ADSP) (Jansen et al., 2019). The phase 1 sample of this GWAS was chosen as it did not include any UKB participants, thereby preventing sample overlap with our follow-up analyses in the UKB. The summary statistics contained 9,862,739 SNPs and was based on 24,087 late-onset AD cases and 55,058 controls with European ancestry.

For the depression phenotype of the GWAS data, we obtained the summary statistics from the PGC MDD GWAS from 2019, including the 23andMe cohort (Wray et al., 2018). The construct of depression here is based on data from cohorts with MDD as well as self-reported depression, thereby closely aligning to the measure of depression that we constructed from the UKB data. We used a version of the meta-analyzed summary statistics where the UKB sample was left out, to prevent sample overlap in downstream analyses. This version contained 15,507,882 SNPs for 121,198 individuals with depression and 329,421 controls. For the post-GWA analyses, we excluded the major histocompatibility complex (MHC) region (chr6: 26–34MB) from both summary statistics, as well as the APOE locus (chr19: 45–45.8 MB) from the AD GWAS, in accordance with recommendations (Smeland et al., 2020).

For the estimate of rg, we applied cross-trait linkage disequilibrium score regression (LDSR) (Bulik-Sullivan et al., 2015). We further applied Gaussian mixture modelling, as implemented in the MiXeR tool, to the GWAS summary statistics, to estimate distributions of causal genetic variants, i.e., unobserved functional genetic variants that influence the phenotypes under investigation (Frei et al., 2019). MiXeR achieves this by fitting Gaussian curves to the GWAS summary statistics to optimally model null and non-null effects. The shapes of these Gaussians are then used to estimate the polygenicity (the number of causal genetic variants involved) and discoverability (average effect size of the causal variants, as h2) of AD and depression. We further estimated the genetic overlap between AD and depression, as the number of causal variants shared regardless of direction of effects, through bivariate MiXeR. For the calculations of the MiXeR parameters, we made use of 9,997,231 SNPs from the 1000 Genomes Phase 3 data. Please see the MiXeR design paper for more details (Frei et al., 2019)).

We conducted conjunctional false discovery rate (cFDR) analysis through the pleioFDR tool using default settings (Smeland et al., 2020). We set an FDR threshold of 0.05 as whole genome significance, in accordance with recommendations.

Participants

We made use of data from participants of the UKB population cohort, under accession number 27412. The composition, setup, and data gathering protocols of UKB have been extensively described elsewhere (Sudlow et al., 2015). We selected all individuals with White European ancestry, as determined by a combination of self-identification as "White British" and similar genetic ancestry based on genetic principal components (UKB field code 22006), with good quality genetic data.

We constructed a proxy measure of AD case-control status, combining information on International Classification of Disease, version 10 (ICD-10) diagnoses of dementia of the participants together with parental age and parental AD status, as described previously (Jansen et al., 2019). Based on lifetime hospital inpatient records linked to the UKB data, we made use of the ICD-10 to assign a score of 2 to any participants with a diagnosis of AD (F00 and/or G30; n=782). All other participants received a 1-U increase for each biological parent reported to have (had) AD. Further, the contribution for each unaffected parent to the score was inversely weighted by the parent's age/age at death, namely (100-age)/100, giving us an approximate score between 0 and 2. This approach was taken in order to account for possible late-life onset AD, i.e., to minimize the labelling of individuals that will develop AD as controls. This proxy measure has been shown to be highly genetically correlated to AD status (r_g=0.81) (Jansen et al., 2019). Participants with missing data on any of the relevant questions were excluded from these analyses (n=19,332). The final sample size was n=390,284, with a mean age of 57.33 years (SD=7.49), and 52.02% was female.

The depression phenotype utilized in this study was constructed by assigning case status to any UKB participant with an ICD10 diagnosis of depression (F32–34, F38–39), n=15,238, as well as any additional participants that answered affirmative o the question whether they had ever seen a general practitioner or psychiatrist for nerves, anxiety, tension, or depression (UKB field codes 2090 and 2010), during any UKB testing visit (n=159,063). Control

status was assigned to anyone who had answered "no" to these questions at all testing visits. This definition of depression is identical to the "broad depression phenotype" described by Howard in 2018 (Howard et al., 2018), based on a GWAS of depression in the UKB, which reported that this definition led to the largest number of genome-wide significant hits, while still being highly genetically correlated with GWAS using a strict clinical definition of MDD, rg=0.85 (Howard et al., 2018). We excluded anyone with any missing data on these questions (n=6,587). The final sample size was n=403,029, with a mean age of 57.21 years (SD=7.49), and 52.02% was female.

Our sample size for the neuroimaging analyses, following preprocessing as described below and excluding individuals with brain disorders, was n=30,699. As the neuroimaging data collection took place several years after the initial data collection, this subsample had a mean age of 64.32 years (SD=7.48), and 52.06% was female.

Genetic Data Pre-processing

We made use of the UKB v3 imputed data, which has undergone extensive quality control procedures as described by the UKB genetics team (Bycroft et al., 2018). After converting the BGEN format to PLINK binary format, we additionally carried out standard quality check procedures, including filtering out individuals with more than 10% missingness, SNPs with more than 5% missingness, SNPs with an INFO score below 0.8, and SNPs failing the Hardy–Weinberg equilibrium test at $p=1 \_ 10^{-9}$. We further set a minor allele frequency threshold of 0.001, leaving 12,245,112 SNPs.

Image Acquisition

For the analyses involving neuroimaging data, we made use of MRI data from UKB released up to March 2020. T1-weighted scans were collected from four scanning sites throughout the United Kingdom, all on identically configured Siemens Skyra 3T scanners, with 32-channel receive head coils. The UKB core neuroimaging team has published extensive information on the applied scanning protocols and procedures, which we refer to for more details (Miller et al., 2016).

The T1-weighted scans were stored locally at the secure computing cluster of the University of Oslo. We applied the standard "recon-all-all" processing pipeline of FreeSurfer v5.3, performing automated surface-based morphometry and subcortical segmentation (Desikan et al., 2006; Fischl et al., 2002). From the output, we extracted all commonly studied global, subcortical, and cortical morphology measures, as listed in Supplementary Table S5.1. For each of these, we summed the left and right hemisphere measure, if applicable, leaving a total of 96 brain measures.

We excluded individuals with bad structural scan quality as indicated by an age and sex-adjusted Euler number [a measure of segmentation quality based on surface reconstruction complexity (Rosen et al., 2018)] more than three standard deviations lower than the scanner site mean, or with a global brain measure more than five standard deviations from the sample mean, n=717.

Statistical Analyses

All downstream analyses were carried out in R v3.6.1. In all follow-up analyses, involving UKB data, we adjusted for age, sex, and the first 20 genetic principal components to control for population stratification. For the neuroimaging analyses, we additionally adjusted for scanner site, Euler number (Rosen et al., 2018), and a measure-specific global estimate for the regional measures (total surface area, mean cortical thickness, or intracranial volume). The latter was done to ensure that we are studying associations with regional brain morphology rather than global effects.

To correct for multiple comparisons, we applied spectral decomposition to the Pearson's correlation matrix of the 96 regional brain measures (Nyholt, 2004). Based on the observed eigenvalues, we estimated the effective number of independent traits in our neuroimaging analyses to be 51. We therefore set an alpha of.001 for these analyses.

Graphs were created through ggplot2 (Wickham, 2009), and brain maps through ggseg (Mowinckel & Vidal-Piñeiro, 2020). The code for running pleioFDR and MiXeR is available via GitHub, https://github.com/precimed/.

Results

Global genetic overlap

Eighteen loci were genome-wide significant in the AD GWAS, which had an estimated SNP-based heritability, $h2$, of 0.05 (SE=0.01). The depression GWAS summary statistics contained 33 significant loci, with an $h2$ of 0.05 (SE=0.002), see Figure 5.1A. These numbers are in line with the results from the original GWAS studies (Jansen et al., 2019; Wray et al., 2018). Using LDSC, the two disorders showed a negligible genetic correlation of -0.03 (SE=0.06, p=0.60).

Through univariate mixture modelling, we found that AD has an estimated 261 causal genetic variants, with a discoverability of 2.1 x10-4. Depression was estimated to involve 15,228 variants, with a discoverability of 6.8 x 10-6. In other words, depression was estimated to be over 50 times more polygenic and its genetic determinants were estimated to be approximately 30 times less discoverable than AD. Expected sample sizes needed to explain half of the genetic variance for AD was 0.5 million, for depression 10 million, see Figure 5.1B.

Bivariate mixture modelling indicated that there were 98 causal variants overlapping between the two traits, i.e., 38% of all variants for AD and 1% of all variants for depression, see Figure 5.1C. Given the size of the reference genome, we estimate that by chance the overlap would be approximately four variants. The fraction of concordant directions of effects for the shared variants was 0.44. The bivariate density plot, Figure 5.1D, illustrates the presence of mixed directions of associations for many SNPs; some SNPs have the same direction of association for both traits, while others are positively associated with AD and negatively associated with

depression or vice versa. The net result of this is a negligible negative correlation, despite a large proportion of AD's causal variants overlapping with depression.
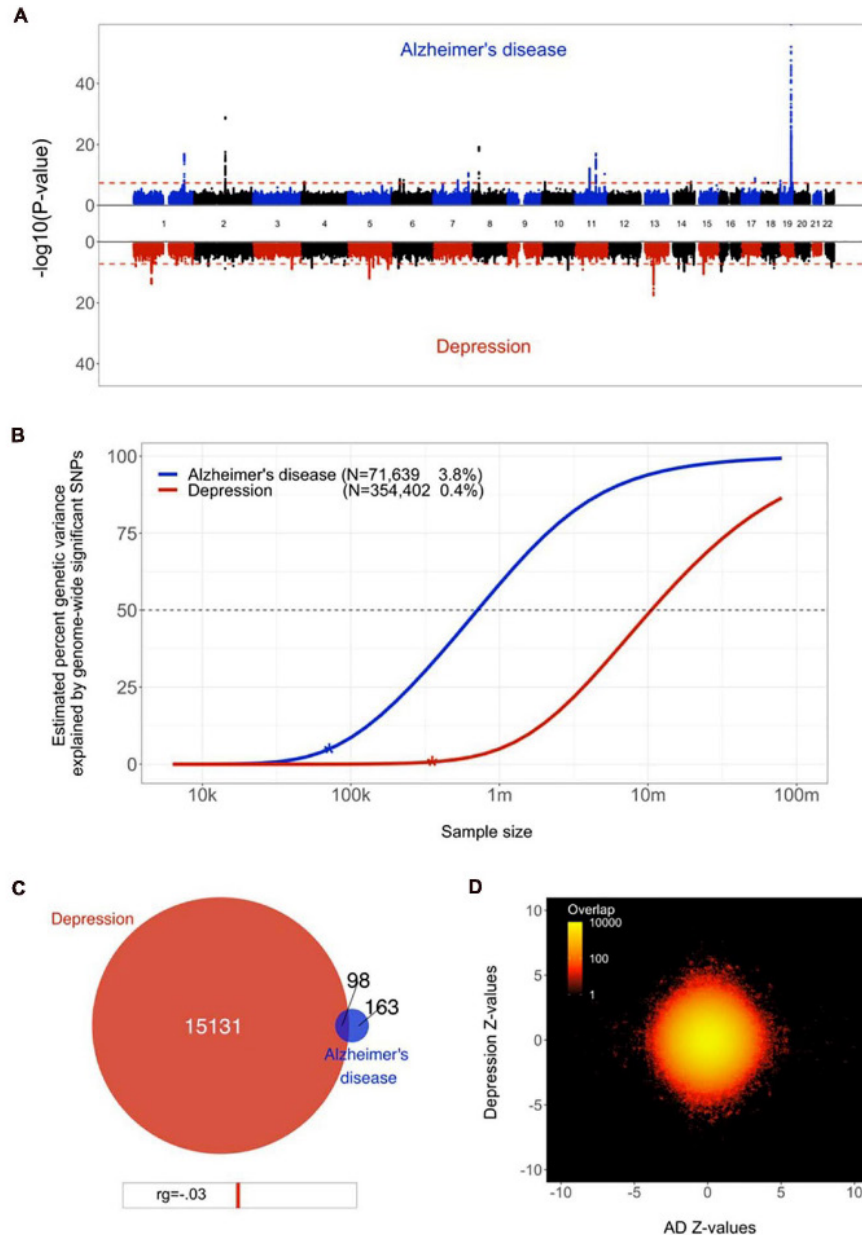
Figure 5.1 | Genetic overlap between Alzheimer's disease (AD) and depression. A) Miami plot, contrasting the observed ⊠log10 (p-values), shown on the y-axis, of each SNP for AD (top half, blue) with depression (bottom half, red). The x-axis shows the relative genomic location, grouped by chromosome, and the red dashed lines indicate the whole-genome significance threshold of 5 10⊠8. B) Estimated percent of genetic variance explained by SNPs surpassing the genome-wide significance threshold, on the y axis, as a function of sample size, depicted on the x axis on a log10 scale, for AD and depression. Current sample sizes and percentages of genetic variance explained by discovered SNPs are shown in parentheses. C) Venn diagram depicting the estimated number of causal variants shared between AD and depression and unique to either of them. Below the diagram, we show the estimated genetic correlation. D) Bivariate density plot, illustrating the relationship between the observed GWAS Z-values for AD (on the x-axis) and depression (on the y-axis).

Locus Overlap

Through conjunctional FDR analysis, we discovered a SNP at chromosome 7, rs5011436, located at an intron of the TMEM106B gene, that was significantly associated with both traits. We replicated this association with both traits using UKB data; for AD, we found a negative relation with the number of copies of the C allele (B=-0.002, SE=6.5 x 10-4, p=9.1 x 10-4), whereas for depression we found a positive relation (B=0.007, SE=0.001, p=3.2 x 10-9), in accordance with the directions of effects as reported in the two original GWAS. We subsequently calculated the association of rs5011436 with cortical and subcortical brain morphology, using the neuroimaging subset of the UKB. As shown in Figure 5.2, we found that the C allele of this SNP is significantly associated with higher volume of the posterior (B=0.035, SE=7.6 x 10-3, p=3.4 x 10-6), mid posterior (B=0.026, SE=7.5 x 10-3, p=6.6 x 10-4), and anterior (B=0.024, SE=7.3 x 10-3, p=8.6 x 10-4) sections of the corpus callosum, lower volume of the third ventricle (B=-0.025, SE=6.1 x 10-3, p=5.0 x 10-6), as well as larger area of the inferior temporal gyrus (B=0.017, SE=4.8 x 10-3, p=5.3 x 10-4).
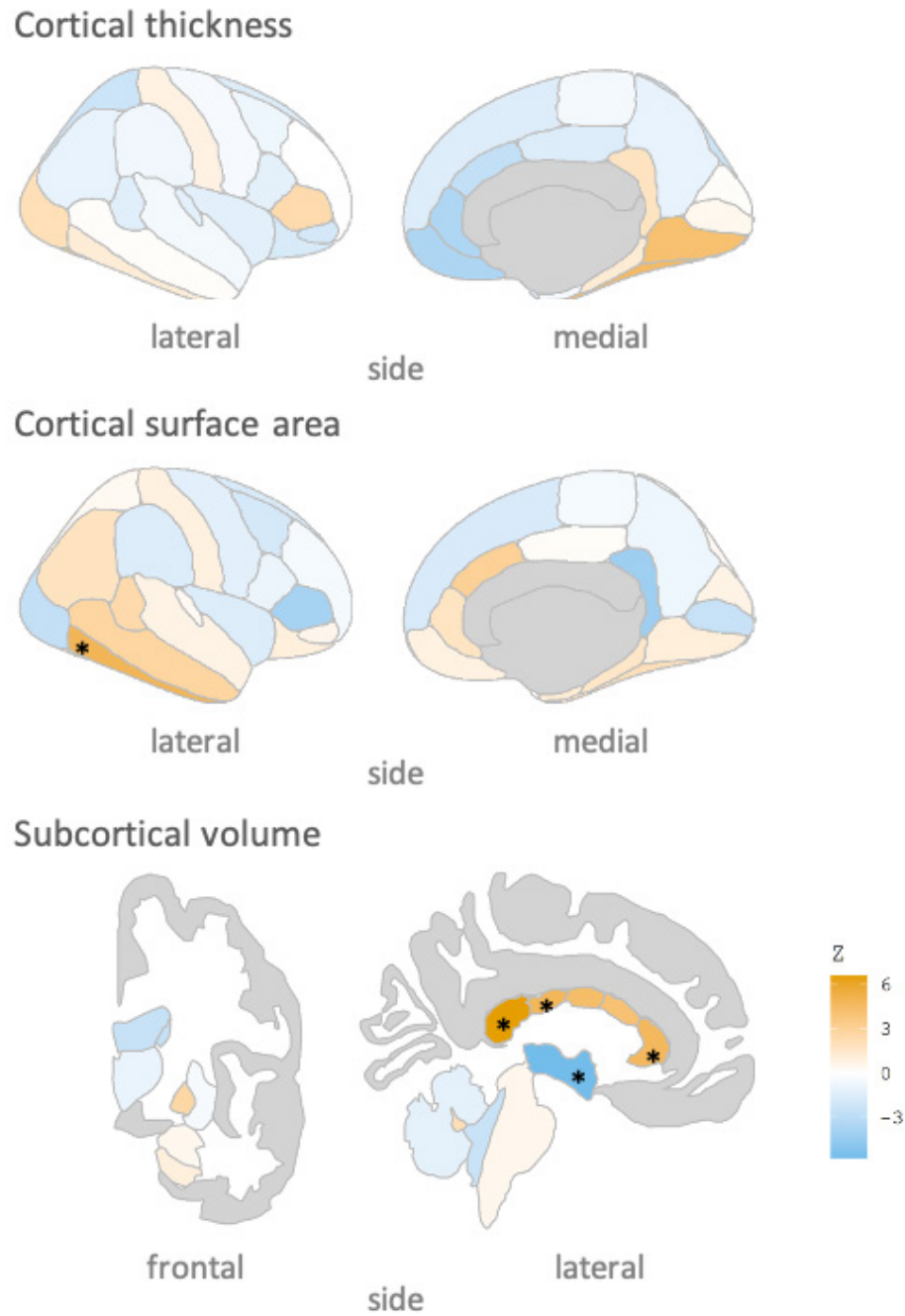
Figure 5.2 | rs5011436 C allele relation to brain morphology. *Brain maps showing the spatial distribution of Z scores. Legend color's intensity shows strength in correlation. Positive correlation in orange, negative correlations in blue. Stars mark the regions that remain significant after multiple comparisons correction (p<.001). Cortical thickness: no significant regions; Cortical surface area: inferior temporal gyrus; Subcortical volume: anterior, mid posterior and posterior corpus callosum, and third ventricle.*

Discussion

Here we employed state-of-the-art biostatistical tools to improve our knowledge of the genetic underpinnings of the relation between AD and depression. In line with previous reports we have identified large differences in the genetic architecture of these disorders. We add new knowledge by revealing the presence of genetic overlap between them. We further illustrated how conjunctional analysis may be used to discover specific shared genetic loci, and substantially expanded on previous efforts by mapping the effects onto the brain in order to identify neurobiological mechanisms that contribute to the relation between these disorders.

We found that many of the causal variants for AD are overlapping with depression. This partly contradicts the previously reported negligible genetic correlation between AD and MD (Gibson et al., 2017), as well as the overall low genetic correlation reported between neurologic and psychiatric disorders (Consortium et al., 2018). However, whereas genetic correlations rely on globally consistent directions of effects between the two traits under investigation, bivariate Gaussian mixture modelling estimates the number of causal variants that have an effect on both, regardless of directions of effects. High levels of mixed directions of effects is likely to be commonplace for complex traits such as brain disorders. This can be seen in, for instance, another psychiatric disorder like schizophrenia, which has been estimated to share virtually all causal variants with educational attainment, despite a near-zero genetic correlation (Frei et al., 2019).

The strong heterogeneity of depressive disorders is likely to contribute to the mixed directions of effects between the two traits, which appears in contradiction to the high levels of reported comorbidity with AD. The heterogeneity of the depression phenotype is evident from its wide range of signs and symptoms and the likely existence of several depression subtypes. This may explain how the extent of genetic overlap can be large for the less polygenic AD, yet very small for depression, which fits with numerous reports that depressive disorders are an important predictor of AD pathology while the opposite is less true (Kramberger et al., 2012; Rapp et al., 2006; Verkaik et al., 2007). We speculate that some depressive disorder subtypes will be shown as genetically more concordant with AD than others, in line with indications that depressive disorders subtypes have significantly different genetic architectures (Jang, Livesley, Taylor, Stein, & Moon, 2004; Milaneschi et al., 2016). The wide range of reported levels of comorbidity with AD across studies (Gracia-García et al., 2015; Mega, Masterman, O'Connor, Barclay, & Cummings, 1999; Starkstein et al., 2005; Verkaik et al., 2007; Zubenko et al., 2003) may also be due to this heterogeneity, as they differ in defining and subtyping of depression. A direct investigation of the relation of AD comorbidity and depressive disorders subtypes, coupled to neurobiological data, would be valuable. We postulate that studies using more narrow depressive disorder subtypes would find lower polygenicity and more concordant directions of effects with AD for specific subtypes.

Our use of cFDR to identify a specific locus shared by AD and depression is an example of how we may use genetics to improve our understanding of the neurobiology underlying the relation between these two disorders. The SNP rs5011436 is located in an intron of the gene TMEM106B, which encodes the transmembrane protein 106B. TMEM106B was the first genetic risk factor to be identified for fronto-temporal lobar degeneration

(FTLD, Van Deerlin et al., 2010). Since then, it has also been reported in GWAS of both AD (Jun et al., 2016) and MD (Howard et al., 2019). The protein TMEM106b is thought to regulate lysosomal function, with a role in the clearance of TPD-43 (Nicholson & Rademakers, 2016). Both lysosomal function and specifically TDP-43 are highly related with the pathogenesis of AD (Nixon et al., 1992; A. C. Wilson, Dugger, Dickson, & Wang, 2011) and MD (Modrego & Ferrández, 2004). TMEM106B expression has been shown to be downregulated in brains of individuals with AD (Satoh et al., 2014), while it has been found to be upregulated in individuals with MD (Dall'Aglio, Lewis, & Pain, 2021).

The evidence for involvement of TMEM106B in both disorders is further substantiated by our neuroimaging analyses, indicating effects on several brain regions that have been tied to both AD and depression. In particular, the corpus callosum was implicated by our analyses, in line with previous neuroimaging findings on TMEM106B (Adams et al., 2014), with higher volume of several callosal subregions for carriers of the rs5011436 c-allele, the allele that we found to convey risk for depression and to be protective for AD. MDD is associated with abnormal cerebral lateralization, and individuals with familial MDD have been found to have significantly larger callosal volume than individuals with non-familial forms of MDD (Lacerda et al., 2005), while AD has been repeatedly linked to degeneration of the corpus callosum (Di Paola, Spalletta, & Caltagirone, 2010). Thus, specific, genetically mediated, forms of depression have been found to have opposing directions of effects on the corpus callosum than other forms of depression and AD.

Our estimates of heritability, polygenicity, and discoverability highlight the complexity of the genetic architecture of both disorders. While twin studies have indicated high broad heritability of AD (Gatz et al., 1997) and MD (Sullivan et al., 2000), we replicated previous findings of low SNP-based heritability, as captured by GWAS data (Jansen et al., 2019; Lambert et al., 2013; Wray et al., 2018). This possibly implicates an important role for rare variants, as well as a high degree of genetic and environmental interaction effects. Contrasting the two disorders, it is clear that AD polygenicity is relatively low, with a recent study even qualifying late-onset AD as oligogenic (Zhang et al., 2020). Depressive symptoms on the other hand are highly polygenic, partly reflecting the substantial clinical heterogeneity which may capture a broad range of conditions, each likely with partly distinct genetic determinants (Jang et al., 2004). Regardless of these differences in genetic architectures, our analyses made clear that, with current approaches, GWAS sample sizes will need to reach millions of individuals to uncover a substantial fraction of the common genetic variance influencing both disorders.

Our findings once again reiterate the complexity of the genetic architectures of brain disorders, highlighting the limitations of the GWAS approach. Our power analyses suggest that, despite tremendous efforts from worldwide consortia to bring together large samples, we are only at the very beginning of uncovering the genetic determinants of AD and depression through the standard GWAS approach. Clearly, more powerful biostatistical tools are needed, ones that better match this complexity and that leverage genetic signal shared across traits of interest (Frei et al., 2019; Smeland et al., 2020; van der Meer et al., 2020) in order to lower the required sample sizes and provide more meaningful metrics. While approaches like Gaussian mixture modelling are a step in the right direction, the current implementation does still suffer from oversimplified

assumptions about the nature of the genetic architecture of brain disorders; AD is enriched for rare variants, while the MiXeR analysis focuses on common variants only. Further, low polygenicity implies a handful of large genetic effects – there is a bigger chance that the distribution of those effect sizes won't follow a Gaussian distribution, violating model assumptions. We are developing extensions of this method that will handle such characteristics.

To conclude, in this study we provided further insights into the genetic relationship between AD and depression, providing evidence of significant genetic overlap, and neuropathological effects reflected in brain morphological changes, warranting further genetic research. However, it seems that the complex relation between AD and depression will require future research to employ larger sample sizes, cleaner phenotype definitions and further improvements of biostatistical tools. It will also be important to study interaction effects between genetic variants and between genetic and environmental factors, as well as the dynamic interplay between relevant factors over the lifespan. These all will influence the underlying biological mechanisms that account for the complex relationship between these disorders. Ultimately this knowledge may provide a path toward more effective treatments, thereby reducing the enormous burden that AD and depression place on patients and their care-givers.

References

Acharya, U. R., Fernandes, S. L., WeiKoh, J. E., Ciaccio, E. J., Fabell, M. K. M., Tanik, U. J., . . . Yeong, C. H. (2019). Automated detection of Alzheimer's disease using brain MRI images–a study with various feature extraction techniques. *Journal of Medical Systems, 43*, 1-14.

Adams, H. H., Verhaaren, B. F., Vrooman, H. A., Uitterlinden, A. G., Hofman, A., van Duijn, C. M., . . . Ikram, M. A. (2014). TMEM106B influences volume of left-sided temporal lobe and interhemispheric structures in the general population. *Biological psychiatry, 76*(6), 503-508.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders, 5th edn (DSM-5). 2013. *Disponible sur:[www. psych. org/practice/dsm/dsm5]. Accessed, 1*.

Andreescu, C., Butters, M. A., Begley, A., Rajji, T., Wu, M., Meltzer, C. C., . . . Aizenstein, H. (2008). Gray matter changes in late life depression—a structural MRI analysis. *Neuropsychopharmacology, 33*(11), 2566-2572.

Barnes, D. E., Yaffe, K., Byers, A. L., McCormick, M., Schaefer, C., & Whitmer, R. A. (2012). Midlife vs late-life depressive symptoms and risk of dementia: differential effects for Alzheimer disease and vascular dementia. *Archives of general psychiatry, 69*(5), 493-498.

Boccia, M., Acierno, M., & Piccardi, L. (2015). Neuroanatomy of Alzheimer's disease and late-life depression: a coordinate-based meta-analysis of MRI studies. *Journal of Alzheimer's Disease, 46*(4), 963-970.

Bremner, J. D., Narayan, M., Anderson, E. R., Staib, L. H., Miller, H. L., & Charney, D. S. (2000). Hippocampal volume reduction in major depression. *American Journal of Psychiatry, 157*(1), 115-118.

Bromet, E., Andrade, L. H., Hwang, I., Sampson, N. A., Alonso, J., De Girolamo, G., . . . Iwata, N. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC medicine, 9*(1), 1-16.

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., . . . Duncan, L. (2015). An atlas of genetic correlations across human diseases and traits. *Nature genetics, 47*(11), 1236-1241.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., . . . O'Connell, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature, 562*(7726), 203-209.

Chen, P., Ganguli, M., Mulsant, B. H., & DeKosky, S. T. (1999). The temporal relationship between depressive symptoms and dementia: a community-based prospective study. *Archives of general psychiatry, 56*(3), 261-266.

Consortium, B., Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., . . . Gormley, P. (2018). Analysis of shared heritability in common disorders of the brain. *Science, 360*(6395), eaap8757.

Dall'Aglio, L., Lewis, C. M., & Pain, O. (2021). Delineating the genetic component of gene expression in major depression. *Biological psychiatry, 89*(6), 627-636.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Hyman, B. T. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage, 31*(3), 968-980.

Devanand, D. P., Sano, M., Tang, M.-X., Taylor, S., Gurland, B. J., Wilder, D., . . . Mayeux, R. (1996). Depressed mood and the incidence of Alzheimer's disease in the elderly living in the community. *Archives of general psychiatry, 53*(2), 175-182.

Di Paola, M., Spalletta, G., & Caltagirone, C. (2010). In vivo structural neuroanatomy of corpus callosum in Alzheimer's disease and mild cognitive impairment using different MRI techniques: a review. *Journal of Alzheimer's Disease, 20*(1), 67-95.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Klaveness, S. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron, 33*(3), 341-355.

Fox, N. C., Crum, W. R., Scahill, R. I., Stevens, J. M., Janssen, J. C., & Rossor, M. N. (2001). Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *The Lancet, 358*(9277), 201-205.

Frei, O., Holland, D., Smeland, O. B., Shadrin, A. A., Fan, C. C., Maeland, S., . . . Thompson, W. K. (2019). Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nature communications, 10*(1), 2417.

Gatz, M., Pedersen, N. L., Berg, S., Johansson, B., Johansson, K., Mortimer, J. A., . . . Ahlbom, A. (1997). Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 52*(2), M117-M125.

Geerlings, M., den Heijer, T., Koudstaal, P., Hofman, A., & Breteler, M. (2008). History of depression, depressive symptoms, and medial temporal lobe atrophy and the risk of Alzheimer disease. *Neurology, 70*(15), 1258-1264.

Gibson, J., Russ, T., Adams, M. J., Clarke, T.-K., Howard, D. M., Hall, L., . . . Davies, G. (2017). Assessing the presence of shared genetic architecture between Alzheimer's disease and major depressive disorder using genome-wide association data. *Translational psychiatry, 7*(4), e1094-e1094.

Gracia-García, P., De-La-Cámara, C., Santabárbara, J., Lopez-Anton, R., Quintanilla, M. A., Ventura, T., . . . Lyketsos, C. (2015). Depression and incident Alzheimer disease: the impact of disease severity. *The American journal of geriatric psychiatry, 23*(2), 119-129.

Green, R. C., Cupples, L. A., Kurz, A., Auerbach, S., Go, R., Sadovnick, D., . . . Edeki, T. (2003). Depression as a risk factor for Alzheimer disease: the MIRAGE Study. *Archives of neurology, 60*(5), 753-759.

Holland, D., Frei, O., Desikan, R., Fan, C.-C., Shadrin, A. A., Smeland, O. B., . . . Dale, A. M. (2020). Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genetics, 16*(5), e1008612.

Howard, D. M., Adams, M. J., Clarke, T.-K., Hafferty, J. D., Gibson, J., Shirali, M., . . . Wigmore, E. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience, 22*(3), 343-352.

Howard, D. M., Adams, M. J., Shirali, M., Clarke, T.-K., Marioni, R. E., Davies, G., . . . Barbu, M. C. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature communications, 9*(1), 1470.

Jack, C. R., Petersen, R. C., Xu, Y. C., O'Brien, P. C., Smith, G. E., Ivnik, R. J., . . . Kokmen, E. (1999). Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology, 52*(7), 1397-1397.

Jang, K. L., Livesley, W. J., Taylor, S., Stein, M. B., & Moon, E. C. (2004). Heritability of individual depressive symptoms. *Journal of affective disorders, 80*(2-3), 125-133.

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., . . . Athanasiu, L. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature genetics, 51*(3), 404-413.

Jorm, A. F. (2001). History of depression as a risk factor for dementia: an updated review. *Australian & New Zealand Journal of Psychiatry, 35*(6), 776-781.

Jun, G., Ibrahim-Verbaas, C. A., Vronskaya, M., Lambert, J.-C., Chung, J., Naj, A. C., . . . Bellenguez, C. (2016). A novel Alzheimer disease locus located near the gene encoding tau protein. *Molecular psychiatry, 21*(1), 108-117.

Kramberger, M. G., Jelic, V., Kåreholt, I., Enache, D., Eriksdotter Jönhagen, M., Winblad, B., & Aarsland, D. (2012). Cerebrospinal fluid Alzheimer markers in depressed elderly subjects with and without Alzheimer's disease. *Dementia and geriatric cognitive disorders extra, 2*(1), 48-56.

Lacerda, A. L., Brambilla, P., Sassi, R. B., Nicoletti, M. A., Mallinger, A. G., Frank, E., . . . Soares, J. C. (2005). Anatomical MRI study of corpus callosum in unipolar depression. *Journal of psychiatric research, 39*(4), 347-354.

Lambert, J.-C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., . . . Beecham, G. W. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics, 45*(12), 1452-1458.

Lebedeva, A., Westman, E., Lebedev, A. V., Li, X., Winblad, B., Simmons, A., . . . Initiative, A. s. D. N. (2014). Structural brain changes associated with depressive symptoms in the elderly with Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry, 85*(8), 930-935.

Leszek, J., Trypka, E., Koutsouraki, E., Michmizos, D., Yarla, N. S., Tarasov, V. V., . . . Aliev, G. (2018). Late-life depression and Alzheimer disease: a potential synergy of the underlying mechanisms. *Current Medicinal Chemistry, 25*(39), 5389-5394.

Lutz, M. W., Sprague, D., Barrera, J., & Chiba-Falek, O. (2020). Shared genetic etiology underlying Alzheimer's disease and major depressive disorder. *Translational psychiatry, 10*(1), 88.

Mega, M. S., Masterman, D. M., O'Connor, S. M., Barclay, T. R., & Cummings, J. L. (1999). The spectrum of behavioral responses to cholinesterase inhibitor therapy in Alzheimer disease. *Archives of neurology, 56*(11), 1388-1393.

Mervaala, E., Föhr, J., Könönen, M., Valkonen-Korhonen, M., Vainio, P., Partanen, K., . . . Karjalainen, A.-K. (2000). Quantitative MRI of the hippocampus and amygdala in severe depression. *Psychological medicine, 30*(1), 117-125.

Milaneschi, Y., Lamers, F., Peyrot, W. J., Abdellaoui, A., Willemsen, G., Hottenga, J. J., . . . Lu, C. (2016). Polygenic dissection of major depression clinical heterogeneity. *Molecular psychiatry, 21*(4), 516-522.

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., . . . Andersson, J. L. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience, 19*(11), 1523-1536.

Modrego, P. J., & Ferrández, J. (2004). Depression in patients with mild cognitive impairment increases the risk of developing dementia of Alzheimer type: a prospective cohort study. *Archives of neurology, 61*(8), 1290-1293.

Mowinckel, A. M., & Vidal-Piñeiro, D. (2020). Visualization of brain statistics with R packages ggseg and ggseg3d. *Advances in Methods and Practices in Psychological Science, 3*(4), 466-483.

Nicholson, A. M., & Rademakers, R. (2016). What we know about TMEM106B in neurodegeneration. *Acta neuropathologica, 132*, 639-651.

Nixon, R. A., Cataldo, A. M., Paskevich, P. A., Hamilton, D. J., Wheelock, T. R., & Kanaley-Andrews, L. (1992). The lysosomal system in neurons. Involvement at multiple stages of Alzheimer's disease pathogenesis. *Annals of the New York Academy of Sciences, 674*, 65-88.

Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics, 74*(4), 765-769.

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. *Archives of general psychiatry, 63*(5), 530-538.

Rapp, M. A., Schnaider-Beeri, M., Grossman, H. T., Sano, M., Perl, D. P., Purohit, D. P., . . . Haroutunian, V. (2006). Increased hippocampal plaques and tangles in patients with Alzheimer disease with a lifetime history of major depression. *Archives of general psychiatry, 63*(2), 161-167.

Reitz, C., & Mayeux, R. (2014). Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochemical pharmacology, 88*(4), 640-651.

Rosen, A. F., Roalf, D. R., Ruparel, K., Blake, J., Seelaus, K., Villa, L. P., . . . Elliott, M. A. (2018). Quantitative assessment of structural image quality. *Neuroimage, 169*, 407-418.

Sabuncu, M. R., Desikan, R. S., Sepulcre, J., Yeo, B. T. T., Liu, H., Schmansky, N. J., . . . Sperling, R. A. (2011). The dynamics of cortical and hippocampal atrophy in Alzheimer disease. *Archives of neurology, 68*(8), 1040-1048.

Satoh, J.-i., Kino, Y., Kawana, N., Yamamoto, Y., Ishida, T., Saito, Y., & Arima, K. (2014). TMEM106B expression is reduced in Alzheimer's disease brains. *Alzheimer's Research & Therapy, 6*, 1-14.

Schmaal, L., Pozzi, E., C. Ho, T., Van Velzen, L. S., Veer, I. M., Opel, N., . . . Aleman, A. (2020). ENIGMA MDD: seven years of global neuroimaging studies of major depression through worldwide data sharing. *Translational psychiatry, 10*(1), 172.

Sheline, Y. I., Sanghavi, M., Mintun, M. A., & Gado, M. H. (1999). Depression duration but not age predicts hippocampal volume loss in medically healthy women with recurrent major depression. *Journal of Neuroscience, 19*(12), 5034-5043.

Smeland, O. B., Frei, O., Shadrin, A., O'Connell, K., Fan, C.-C., Bahrami, S., . . . Dale, A. M. (2020). Discovery of shared genomic loci using the conditional false discovery rate approach. *Human genetics, 139*(1), 85-94.

Starkstein, S. E., Jorge, R., Mizrahi, R., & Robinson, R. G. (2005). The construct of minor and major depression in Alzheimer's disease. *American Journal of Psychiatry, 162*(11), 2086-2093.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine, 12*(3), e1001779.

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry, 157*(10), 1552-1562.

Van Deerlin, V. M., Sleiman, P. M., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.-S., Graff-Radford, N. R., . . . Grossman, M. (2010). Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nature genetics, 42*(3), 234-239.

van der Meer, D., Frei, O., Kaufmann, T., Shadrin, A. A., Devor, A., Smeland, O. B., . . . Westlye, L. T. (2020). Understanding the genetic determinants of the brain with MOSTest. *Nature communications, 11*(1), 3512.

Verkaik, R., Nuyen, J., Schellevis, F., & Francke, A. (2007). The relationship between severity of Alzheimer's disease and prevalence of comorbid depressive symptoms and depression: a systematic review. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences, 22*(11), 1063-1086.

von Strauss, E., Viitanen, M., De Ronchi, D., Winblad, B., & Fratiglioni, L. (1999). Aging and the occurrence of dementia: findings from a population-based cohort with a large sample of nonagenarians. *Archives of neurology, 56*(5), 587-592.

Wickham, H. (2009). ggplot2: elegant graphics for data analysis New York. *NY: Springer*.

Wilson, A. C., Dugger, B. N., Dickson, D. W., & Wang, D.-S. (2011). TDP-43 in aging and Alzheimer's disease-a review. *International journal of clinical and experimental pathology, 4*(2), 147.

Wilson, R. S., Barnes, L., De Leon, C. M., Aggarwal, N., Schneider, J., Bach, J., . . . Evans, D. (2002). Depressive symptoms, cognitive decline, and risk of AD in older persons. *Neurology, 59*(3), 364-370.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., . . . Andlauer, T. M. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics, 50*(5), 668-681.

Ye, Q., Bai, F., & Zhang, Z. (2016). Shared genetic risk factors for late-life depression and Alzheimer's disease. *Journal of Alzheimer's Disease, 52*(1), 1-15.

Zhang, Q., Sidorenko, J., Couvy-Duchesne, B., Marioni, R. E., Wright, M. J., Goate, A. M., . . . Laws, S. M. (2020). Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. *Nature communications, 11*(1), 4799.

Zubenko, G. S., Zubenko, W. N., McPherson, S., Spoor, E., Marin, D. B., Farlow, M. R., . . . Petersen, R. C. (2003). A collaborative study of the emergence and clinical features of the major depressive syndrome of Alzheimer's disease. *American Journal of Psychiatry, 160*(5), 857-866.

Chapter 5: Genetic overlap between Alzheimer's disease and depression mapped onto the brain

Supplementary Table S5.1. List of the 96 brain measures included in the analyses

Global measures and subcortical volumes	Regional cortical thickness and surface area
Estimated intracranial volume	1. Bankssts
Cerebrospinal fluid	2. Caudalanteriorcingulate
Total white surface area	3. Caudalmiddlefrontal
Mean cortical thickness	4. Cuneus
Lateral Ventricle	5. Entorhinal
Inferior Lateral Ventricle	6. Fusiform
Cerebellum White Matter	7. Inferiorparietal
Cerebellum Cortex	8. Inferiortemporal
Thalamus Proper	9. Isthmuscingulate
Caudate	10. Lateraloccipital
Putamen	11. Lateralorbitofrontal
Pallidum	12. Lingual
3rd Ventricle	13. Medialorbitofrontal
4th Ventricle	14. Middletemporal
Brain Stem	15. Parahippocampal
Hippocampus	16. Paracentral
Amygdala	17. Parsopercularis
Accumbens Area	18. Parsorbitalis
Ventral Diencephalon	19. Parstriangularis
Vessel	20. Pericalcarine
Choroid Plexus	21. Postcentral
5th Ventricle	22. Posteriorcingulate
Optic Chiasm	23. Precentral
Corpus Callosum Posterior	24. Precuneus
Corpus Callosum Mid Posterior	25. Rostralanteriorcingulate
Corpus Callosum Central	26. Rostralmiddlefrontal
Corpus Callosum Mid Anterior	27. Superiorfrontal
Corpus Callosum Anterior	28. Superiorparietal
	29. Superiortemporal
	30. Supramarginal
	31. Frontalpole
	32. Temporalpole
	33. Transversetemporal
	34. Insula

Supplementary Table S5.1. List of the 96 brain measures included in the analyses

CHAPTER 6
GENERAL DISCUSSION

Outline & key findings

Brain imaging in population-based cohort studies plays a pivotal role in unveiling the structural changes and underlying pathophysiological mechanisms associated with various degenerative, psychiatric, and metabolic diseases. The primary objective of this thesis is to leverage high-quality neuroimaging data from two distinguished population-based cohorts to investigate methodological considerations and clinical applications of this invaluable resource.

Chapter 2 investigates the effects of manual editing on brain segmentations, aiming to assess the impact on morphological estimates. Additionally, it strives to identify a reliable and time-efficient quality control strategy that can serve as an alternative to the current gold standard, visual inspection. Results demonstrate that manual editing of brain surfaces induces significant changes in brain estimates, particularly within subcortical structures. Manual quality control exhibits the highest reliability, while the exclusion of outliers based on automatically generated Euler numbers in FreeSurfer emerges as the most reliable time-efficient alternative among the automated quality control strategies.

Chapter 3 focuses the associations between prediabetes, type 2 diabetes mellitus (T2DM), and continuous measures of hyperglycemia with the volumes of hippocampal subfields. Notably, no significant associations were observed between prediabetes and hippocampal subfields, whereas T2DM was associated with a generalized lower volume of most subfields. Linear trend analysis and the use of continuous measures of hyperglycemia, show a dose-response relationship between hyperglycemia and lower hippocampal subfield volumes.

Chapter 4 examines the associations between the incidence, prevalence, and course of depression with the volumes of hippocampal subfields. The results indicate that changes in hippocampus subfield volumes may co-occur or follow the onset of depressive symptoms, rather precede it. We found limited evidence to support that specific volume changes could precede the onset of (chronic) depressive symptoms.

Chapter 5 explores the genetic overlap between depression and Alzheimer's disease. There is genetic overlap with varying directions of effect between both diseases. Notably, a single nucleotide polymorphism (SNP) located in the TMEM106B gene is shared by both Alzheimer's disease and depression. The presence of this SNP correlates with higher volumes of the corpus callosum and higher surface area in the inferior temporal gyrus.

Overall, this thesis utilizes robust neuroimaging data from recognized population-based cohorts to unravel crucial insights into structural changes, pathophysiological mechanisms, and clinical applications in the context of degenerative, psychiatric, and metabolic diseases.

Neuroimaging in large population-based cohorts

This thesis utilizes data from two large population-based cohorts that include neuroimaging data, namely, The Maastricht Study and the United Kingdom (UK) Biobank, to investigate the neurobiological underpinnings of complex diseases such as T2DM, depression, and Alzheimer's disease. As the brain is a complex organ where small changes may have a considerable impact, large-scale approaches are needed to capture the subtle variations in brain structure associated with these conditions. The use of population-based cohorts is beneficial to identify early markers before the onset of a disease, helping understand its etiological process. Further, large population-based cohorts increase the statistical power needed to detect meaningful differences in brain structure associated with these diseases. Additionally, large cohorts enable the use of more complex biological models, facilitating the formulation of more specific research questions. For instance, multiple regression analysis can add covariates to a statistical model that might be related to the outcome, controlling for the effect that a specific covariate has on the dependent variable. This approach helps to rule out whether a covariate is driving the association detected between a dependent and an independent variable. By carefully considering the role of covariates, the associations between diseases and brain structure can be studied while holding other potentially influential factors constant, thereby improving the validity and reliability of our findings.

The use of large population-based cohorts also improves the generalizability of the results, increasing the probability that the findings are representative of the entire population. As in any cohort study, healthier people are more willing to participate, creating a selection bias. However, as long as the variation in determinants and outcomes is large enough, this should not interfere too much with the results. Further, it is important to take into consideration any oversampling of specific phenotypes, as this may affect the generality of the results. In this regard, The Maastricht Study was designed with the specific purpose to study T2DM, its complications and comorbidities. Therefore, The Maastricht Study has an oversampling for T2DM, which increases the available sample size for studying this phenotype that is otherwise scarce in the general population. This allows for greater statistical power in certain analyses, but may affect the generalizability of the results. **Chapter 3** investigates the associations between prediabetes, T2DM, and continuous measures of hyperglycemia and the hippocampal subfields using data from The Maastricht Study. The oversampling for T2DM proved to be extremely useful in this case. The study uses a sample of n=4636 participants with available data on glucose metabolism status, neuroimaging, and all covariates. According to recent epidemiological studies, the prevalence of T2DM in the Netherlands ranges between 2.2% and 3.2% (Baan & Feskens, 2001; Ubink-Veltmaat et al., 2003). In our sample, this would imply a total of n=148 cases with T2DM in the scenario with the highest prevalence. However, due to the oversampling for T2DM in The Maastricht Study, a total of n=869 participants with T2DM were available, considerably increasing the statistical power and enabling the formulation of a very specific research question about the association between T2DM and hippocampal subfields.

Despite this oversampling, the cohort can be used as population sample as well. However, additional analyses are then needed to take this into account. **Chapter 4** investigates the associations between prevalent and incident depression and the volume of the hippocampal subfields. For this particular research question, T2DM was first included as a covariate, enabling to account for its potential confounding effect on the association between depression and the volume of the hippocampal subfields. However, the inclusion of a covariate is not always a perfect solution. For the inclusion of T2DM in the model, one has to take into consideration that if a covariate is highly correlated with another covariate, the model may present multi-collinearity, which can result in unstable estimates, inflated standard errors, and ultimately decrease the power of the model. Furthermore, when a covariate is not actually related to the outcome variable, and does not explain any additional variance beyond what is already accounted for by other variables in the model, including such a covariate could increase the model complexity and decrease degrees of freedom, ultimately reducing the power of the model. In the case of **Chapter 4**, however, the inclusion of T2DM as a covariate revealed no influence of T2DM on the association between depression and hippocampal subfields. In other words, correcting for T2DM allows to claim that the associations found between depression and hippocampal subfields were independent of T2DM. Nevertheless, the cohort still does not represent the entire population, and caution should be exercised when generalizing the results, even when accounting for the oversampled covariate in the model. To ensure the robustness of the findings, sensitivity analyses were conducted where participants with T2DM were excluded from the analysis. This alternative approach verified that participants with T2DM did not drive the associations, thereby increasing the confidence in the reliability of our results.

The use of large population-based neuroimaging cohorts also presents numerous challenges, including the acquisition and storage of vast amounts of data that require significant resources in infrastructure and personnel. In addition, ensuring the accuracy, reliability, and validity of the data and results while minimizing errors and biases is critical in neuroimaging, and particularly in structural brain segmentation. Traditional manual quality control of the acquired phenotypes is complex due to the high numbers and subsequent labor involved. **Chapter 2** assesses the effects of manual editing brain segmentations on the phenotype-derived information used to answer subsequent research questions. Further, and to find a reliable and time-efficient quality control strategy for The Maastricht Study, several available and commonly used quality control strategies were compared to manual quality control, currently considered the gold standard. The obtained results identified a time-efficient and reproducible automated quality control strategy suitable to be applied in The Maastricht Study brain segmentation data. This process enhanced the quality of the data used in the subsequent chapters.

Neuroimaging of the hippocampus in T2DM and depression

The hippocampus is a complex brain structure located in the medial temporal lobe, composed of different subfields with distinct cellular and molecular properties. It is associated with memory formation, consolidation and retrieval, emotional processing, and spatial navigation (Knierim, 2015).

This thesis shows that both T2DM and chronic depression are associated with lower hippocampal subfields volume. Given the different cellular composition and physiology of the hippocampal subfields (Fanselow & Dong, 2010), we hypothesized that the subfields would be differently associated with hyperglycemia. However, results in **Chapter 3** indicate that T2DM is associated with a generalized atrophy of the hippocampus. Hyperglycemia has previously been linked to neuronal loss and impaired connectivity (Vergoossen et al., 2020; Zhang, Shaw, & Cherbuin, 2022), and our findings suggest that this is occurring throughout the hippocampus, with no specificity across subfields. In contrast, findings in **Chapter 4** showed that depression exhibited some specificity for particular hippocampal subfields. These are lower volumes in the fimbria, dentate gyrus, and Cornu Ammonis structures, as well as a higher volume in the molecular layer. These results raise the question of whether it is better to study the hippocampus as a whole or to divide it into subfields. On the one hand, using different subfields entails repeating the statistical analysis in each subfield, increasing the possibility of a type I error. On the other hand, studying the hippocampus as a whole might obscure the association due to mixed directions of effect of its parts, increasing the possibility of a type II error. Moreover, studying the hippocampus as a whole might overlook relevant associations with subfields that have specific functions. **Chapter 4** illustrates this concept. While the analysis of the entire hippocampus revealed a lower volume to be associated with a chronic course of depression, studying the subfields further uncovered a higher volume of the molecular layer. The study of the hippocampal subfields volume in addition to the total hippocampus volume shows associations otherwise hidden, providing further insights into the underlying pathophysiology of depression and comorbid conditions such as cognitive impairment.

Both T2DM and depression are associated with lower volumes in fimbria, dentate gyrus, Cornu Ammonis, and subicular structures. Furthermore, T2DM is associated with lower volumes in the subiculum and tail, which are not found in depression. As mentioned before, T2DM seems to be associated with generalized atrophy. However, depression is additionally associated with a higher volume in the molecular layer. This is a remarkable result that has never been shown before. A higher volume could be due to inflammatory processes, cellular degeneration, or the associated edema. In fact, oxidative stress and inflammation have been proposed to explain the structural changes observed throughout the illness progression in depression (Bellau, 2019).

Advanced phenotyping

Diabetes

Chapter 3 delves into the associations of T2DM, prediabetes, and continuous measures of hyperglycemia with the volume of hippocampal subfields. Glucose metabolism status assessment in The Maastricht Study was conducted through an oral glucose tolerance test following an overnight fasting, as described by Schram et al. (2014). This method is currently considered the gold standard, ensuring highly accurate assessment of participants' glucose metabolism status. The classification into T2DM, prediabetes, or normal glucose metabolism status was determined following the World Health Organization criteria (World Health Organization, 2006), which is widely used in research in Europe. In the United States, levels of HbA1c and the American Diabetes Association criteria (Gavin III, Alberti, Davidson, & DeFronzo, 1997) are more often used. However, both measures and criteria are widely recognized and trusted in the literature (Barzilay et al., 1999; Gabir et al., 2000). The oral glucose tolerance test measurement also enabled the use of continuous measures of hyperglycemia (fasting blood glucose, 2 hours post load blood glucose, and HbA1c) directly into the analysis, which is not possible in studies relying on self-reported glucose metabolism status. This feature holds significant value in understanding the nature of the associations between hyperglycemia and brain damage. Notably, leveraging this feature, **Chapter 3** uncovered a dose-response relationship between hyperglycemia and differences in hippocampal volume, thereby highlighting the significance of employing deep phenotyping with high-quality data to advance our understanding of diabetes.

Depression

Depression is a multifactorial disease, with a complex neurobiological correlate widely studied in neuroscience. Several tools and definitions can be used to classify depression. **Chapter 4** examines the relationship between depressive symptoms and hippocampal volumes. Depressive symptoms were assessed using the 9-item Patient Health Questionnaire (PHQ-9) developed by Kroenke, Spitzer, and Williams (2001). The PHQ-9 is a self-administered questionnaire that measures the presence of nine symptoms associated with major depressive disorder, as defined by the DSM-IV criteria (American Psychiatric Association, 1994). The PHQ-9 has demonstrated high validity and reliability (Kroenke et al., 2001; Martin, Rief, Klaiberg, & Braehler, 2006). Further, to define "clinically relevant depressive symptoms", a cutoff score of ≥10 was utilized. This has shown high sensitivity and specificity in identifying depression (Pettersson, Boström, Gustavsson, & Ekselius, 2015). However, despite the good sensitivity and specificity, the PHQ-9 does not equal a diagnostic of major depressive disorder that is used in clinical care (Negeri et al., 2021). Hence, it is important to note that although elevated scores on the PHQ-9 may strongly indicate the presence of major depressive disorder, they alone do not provide a definitive confirmation of its diagnosis. Yet, the PHQ-9 is an easy tool to apply as it is a short questionnaire that participants can complete on their own, whereas a diagnostic interview requires the participant to meet

in person with an instructed professional, which makes the assessment more time and resource consuming. Using the PHQ-9 increases adherence to the study as it reduces the burden for the participant, and further allows for an annual evaluation of depressive symptoms, together with a set of tests and questionnaires that are sent to the participants of The Maastricht Study every year. Another advantage of utilizing the PHQ-9, and specifically PHQ-9 scores, is that it allows for the examination of depressive symptoms as a continuum. This continuous measure of depressive symptoms is highly relevant, and by considering depressive symptoms on a spectrum rather than a categorical diagnosis, one could gain a more nuanced understanding of the association between depression and the brain.

Chapter 5 employs data from the UK Biobank to explore the genetic overlap between Alzheimer's disease and depression. In this case, the "broad depression phenotype" described by Howard et al. (2018) was utilized. Depressive status was determined based on clinical records and self-report. Participants were classified as having depression if a clinical diagnosis according to ICD-10 was documented in their clinical records, or if they reported seeking professional help for nerves, anxiety, or depression. The creation of a composite in this case helps mitigate potential inaccuracies stemming from incomplete or outdated clinical records. However, it is important to acknowledge that seeking help for depression does not necessarily equate to a clinical diagnosis of depression. Nevertheless, Howard et al. (2018) conducted a genome-wide association study (GWAS) using this composite measure in the UK Biobank, and the results demonstrated a high genetic correlation with a GWAS that employed a strict clinical definition of major depressive disorder, making this phenotype a good proxy for major depressive disorder.

Early-life depression has been associated with an increased risk for the later development of Alzheimer's disease (Andersen, Lolk, Kragh-Sørensen, Petersen, & Green, 2005; Ownby, Crocco, Acevedo, John, & Loewenstein, 2006), even when depressive symptoms were observed 25 years before the onset of Alzheimer's disease (Green et al., 2003). The ways in which these pathologies are associated are complex, and some studies suggest that depression may not be a risk factor but a prodromal manifestation of Alzheimer's disease (Alexopoulos et al., 2002; Chen, Ganguli, Mulsant, & DeKosky, 1999; Geerlings & Gerritsen, 2017; Sun et al., 2008). **Chapter 5** aimed to elucidate the potential genetic overlap between both diseases. We found a single nucleotide polymorphism located in the TMEM106B gene to be associated with both depression and Alzheimer's disease. Consistent with our findings, previous literature has shown the TMEM106B gene is upregulated in the brains of individuals with depression (Dall'Aglio et al., 2020), while being downregulated in the brains of individuals with Alzheimer's disease (Satoh et al., 2014). However, this genetic overlap only explains a small part of the complex relationship between these diseases.

Overall, the study of the associations between depression and brain volumes is complex. The use of self-reported tools and proxies when clinical diagnoses are not available, add another layer of complexity in an already difficult question. Also, depression is a heterogeneous disease, characterized by a wide range of, sometimes contrary, signs and symptoms, with different courses and severity. This further complicates the analysis of its relationship with brain structures. It is important to note that a current report of depression,

whether self-reported or diagnosed by a healthcare professional, does not differentiate between depression resulting from external stressors and depression as an intrinsic chronic condition. This is a significant consideration since a depressive response to a stressful and challenging event can be specific to the situation, and once the stressor diminishes, the depressive symptoms can in some occasions dissipate. Distinguishing between transient and persistent depression is crucial for a comprehensive understanding of the underlying mechanisms and brain correlates. **Chapter 4** addresses this topic through the use of longitudinal data on depression. Depression was assessed in The Maastricht Study annually, over a period of 7 years. This allowed the distinction between chronic or transient depression, and results show a significant association with some hippocampal subfields in chronic but not in transient depression. This result emphasizes the need for cautious interpretation when studying the associations between brain structure and depression, and potentially, provide a hint to address this heterogeneity by defining subtypes of depression. Overall, these results highlight the importance of considering various factors, such as the nature and duration of depressive symptoms, the presence of external stressors, and the potential for adaptive responses to challenging life circumstances, for the better understanding of depression's neurological processes. Further research needs a comprehensive approach to disentangle the intricate interplay between brain structure and depression in its various manifestations.

Alzheimer's disease

The use of phenotypes by proxy in research is a valuable approach, particularly when direct measurement of a specific phenotype poses challenges or is not feasible. This method involves leveraging related proxy measures or indicators to make inferences about the presence of a particular phenotype of interest. **Chapter 5** investigates the genetic overlap between depression and Alzheimer's disease using data from the UK Biobank. To achieve this, an Alzheimer's disease by proxy measure was employed, following the methodology previously described by Jansen et al. (2019). This measure was created by gathering information from clinical records, including diagnoses of dementia, parental age at diagnosis, at death or at assessment, and parental Alzheimer's disease status. In essence, participants with a confirmed diagnosis of Alzheimer's disease were assigned a score of 2. For other participants, a score of 1 was given for each biological parent reported to have had Alzheimer's disease. Additionally, the contribution of unaffected parents to the score was inversely weighted by their age or age at death, employing the formula (100-age)/100. It is important to note that, as expected in a population-based cohort, the number of cases with a current diagnosis of Alzheimer's disease was small. Therefore, most cases of Alzheimer's disease by proxy used in this thesis were actually based on parental Alzheimer's disease information.

The utilization of Alzheimer's disease by proxy in our study holds significant advantages, primarily due to its robust correlation with familial risk of Alzheimer's disease, as demonstrated by Jansen et al. (2019). This correlation presented a valuable opportunity for the objectives outlined in **Chapter 5**, where the main focus was to investigate the genetic overlap between Alzheimer's disease and depression. One crucial

aspect was the identification of potential cases of Alzheimer's disease before the onset of symptoms, which was essential to capture individuals who may carry a higher genetic risk for the disease. This approach allowed us to avoid misclassifying those with high risk burden as controls and enabled us to gain a deeper understanding of the genetic overlap between Alzheimer's disease and depression. Additionally, it provided the opportunity to map the associations between these genetic factors and the structural changes in the brain that precede the manifestation of clinical symptoms. By identifying individuals at-risk prior to symptom onset, one could uncover valuable insights into the genetic, environmental and lifestyle factors involved in the development of both Alzheimer's disease and depression, as well as their structural brain correlates. This early identification not only contributes to advancing our knowledge in the field but also opens avenues for potential early interventions and personalized approaches in managing these conditions. Although the Alzheimer's disease by proxy construct shows a strong genetic correlation with Alzheimer's disease (r_g=0.81) (Jansen et al., 2019), this approach cannot predict who will develop the symptoms associated to the disease, which may result in some misclassification and subsequent dilution of the associations being studied. Overall, despite some potential misclassification, the use of Alzheimer's disease by proxy in our study allowed us to explore the complex interplay between genetics, structural brain changes, and the development of Alzheimer's disease and depression, providing a valuable foundation for future research and potential clinical implications.

Hippocampal subfields

MRI is an invaluable tool for investigating the morphology of the brain in living subjects. This noninvasive technique offers high spatial resolution, enabling precise examination of brain structures. By employing segmentation algorithms, it becomes possible to measure the different components that constitute the brain. **Chapters 3** and **4** explore the relationships of T2DM and depression, with the hippocampal subfields using neuroimaging data from The Maastricht Study, with a comprehensive brain segmentation using FreeSurfer v6.0 (Fischl, 2012). An additional specialized segmentation technique for the segmentation of hippocampal subfields was used, as described by J. Iglesias (2020).

The automatic segmentation of hippocampal subfields, particularly at 3T, may initially raise concerns due to the complexity and small size of this structures. In The Maastricht Study, T1-w images have a 1.0 mm^3 reconstructed voxel size. However, images with similar acquisition parameters demonstrated favorable test-retest reliability in several studies (Brown et al., 2020; Marizzoni et al., 2015; Quattrini et al., 2020; Worker et al., 2018). Among the larger subfields, such as the tail, subiculum, presubiculum, CA1-4, dentate gyrus, and molecular layer, the highest reproducibility has been observed (Marizzoni et al., 2015). Additionally, a strong agreement of FreeSurfer's hippocampal segmentation with manual segmentation has also been reported (Tae, Kim, Lee, Nam, & Kim, 2008). However, the anterior portion of the hippocampal formation, specifically the CA1 and subiculum boundary, exhibits lower agreement. FreeSurfer tends to include parts of CA1 within the subiculum, resulting in an overestimation of this subfield

volume (de Flores et al., 2015; Marizzoni et al., 2015; Yushkevich et al., 2015). To improve segmentation quality, multispectral segmentation was used in our sample (J. E. Iglesias et al., 2015; Seiger et al., 2021). This method employs the intensities of the main T1 volume to improve the segmentation reliability. Furthermore, all segmentations obtained in The Maastricht Study underwent rigorous quality control, which involved excluding of outliers based on Euler numbers. This technique offers similar benefits in quality control compared to visual inspection for hippocampal subfield segmentation (Monereo-Sánchez et al., 2021), thus ensuring the robustness of the data.

While automated segmentation of hippocampal subfields has proven to be a valuable tool in gaining insight into the neurobiological foundations of various brain-related traits and disorders (Sämann et al., 2018), it is important to interpret the results with caution, particularly when considering the smaller subfields such as the fimbria or hippocampal fissure. Despite employing various strategies to enhance the reliability of hippocampal subfield segmentation, the inherent complexity and intricacies of these smaller subfields pose challenges that require careful consideration.

GWAS

Genome Wide Association Studies (GWAS) constitute a very powerful tool to identify genes associated with a specific trait. **Chapter 5** explores the genetic relationship between depression and Alzheimer's disease using GWAS summary statistics. This allowed to identify shared genetic factors between these two conditions. Additionally, **Chapter 5** studies the connection between a single nucleotide polymorphism found to be significantly associated with both Alzheimer's disease and depression, and its association with brain structure. Findings revealed a correlation with higher volume in the corpus callosum, lower volume in the third ventricle and larger surface area in inferior temporal gyrus. Reverse causation or confounding is unlikely due to the nature of genetic data. However, it is crucial to recognize that establishing a direct causal relationship between genetic variants and brain structure is challenging due to the intricate interplay of genes, behavior, and environmental factors. Genetic variants can indeed contribute to variations in brain structure. However, it is essential to interpret their impact with caution. Genetic factors alone do not provide a comprehensive explanation for the complexities of brain structure, as they interact with a wide range of behavioral and environmental influences. Thus, a holistic understanding of brain structure and its underlying determinants requires consideration of multiple factors beyond genetics alone.

Statistical considerations

The interpretation of the findings in this dissertation requires thoughtful consideration of several methodological factors. It is important to note that the reported associations are derived from observational data, which can be susceptible to various biases.

Confounding

Confounding refers to a situation where the association between an independent and a dependent variable is obscured by the influence of a third variable (Rothman, Greenland, & Lash, 2008). Given the intricate nature of the associations between brain morphology and the diverse and complex pathologies addressed in this thesis, it was essential to account for the potential influences that confounding variables might have had in our analysis.

To address confounding, a comprehensive set of covariates was incorporated to the models. The covariates were specific for each chapter, yet, when relevant they included demographic factors (age, sex and educational level), lifestyle and cardiovascular risk factors (such, waist circumference, smoking status, alcohol intake, hypertension, or total-to-HDL cholesterol ratio), and relevant covariates related to brain MRI (such as lag time between assessment of variables and MRI scans). By including these covariates in the models, the potential confounding effects that these variables could introduce were minimized. Furthermore, sensitivity analyses were conducted to assess the robustness of our findings. Specifically, sensitivity analyses where participants with specific characteristics were excluded from the analysis, such as participants with T2DM or participants with a history of major depressive disorder. This allowed to examine the impact of these covariates on the observed associations and limit their potential impact as confounders. This strategy was aimed to strengthen the internal validity of the study by exploring the consistency of the results across different scenarios. However, it is important to acknowledge that despite our efforts, residual confounding may still exist, as confounding variables can be difficult to fully capture and account for in any study. Nevertheless, by employing rigorous methodologies, incorporating a range of covariates, and conducting sensitivity analyses, significant steps were taken towards minimizing the impact of confounding and enhancing the robustness of the findings.

Selection bias

Selection bias is a systematic error that arises when the selection of participants in a sample is not representative of the entire population, leading to a disparity in the characteristics compared to those who are not selected. This can result in over- or under-estimations of the true associations under investigation (Szklo & Nieto, 2014). Cohort studies usually exhibit a selection bias towards healthier participants, which can manifest at several stages of the study. During the recruitment process, individuals who are willing to participate in research studies often exhibit better health. For instance, responders in the UK Biobank tend

to be older, female, and reside in less socioeconomically deprived areas compared to non-responders (Fry et al., 2017). Additionally, compared to the general population, participants in these cohorts demonstrate a more favorable cardio-metabolic profile and healthier lifestyle (Fry et al., 2017). Another potential source of selection bias is the selection of specific groups for the study. For example, in **Chapters 3 and 4**, data from The Maastricht Study included only 4,643 out of 7,689 participants. Most excluded cases had missing MRI data or missed some covariate, and these participants tended to be older, have a worse cardiovascular risk profile, and more frequently possess lower educational attainment compared to those included in the study. A third source of selection bias is the retention of participants throughout the study, specifically in longitudinal studies. Loss to follow-up can introduce biases if the attrition is not random, which might be the case with depression. However, The Maastricht Study preserves a high response rate towards depression questionnaires of around 80% in the first years. The lower numbers in later years are a result of the still ongoing annual follow up. Although population-based cohorts may be subject to a healthy volunteer bias, which can limit their representativeness, this limitation is not relevant unless the objective is to study the absolute prevalence or incidence of a specific phenotype. When examining associations between brain volumes and specific diseases, as long as there is adequate variation in the determinant and outcome, representativeness is not essential. However, it is crucial to take it into consideration when interpreting the results of the analysis.

Information bias

Information bias refers to a systematic error that occurs when there are inaccuracies in the measurements of variables in a study. Such inaccuracies can cause the collected data to deviate from reflecting the true values of the variables of interest, consequently leading to biased estimates of associations (Rothman et al., 2008).

Chapter 2 focuses exclusively on investigating strategies aimed at detecting and preventing segmentation inaccuracies to minimize information bias, thereby ensuring the validity and reliability of subsequent study findings. Segmentation inaccuracies occur when the boundary of a brain structure, as proposed by our used segmentation tool, FreeSurfer v6.0 (Fischl, 2012), fails to align with the actual anatomical boundary. These inaccuracies tend to be more prevalent in MRI scans of lower quality or with diminished parenchymal integrity, which commonly arise among older and less healthy participants. The accuracy of the segmentation process may vary across specific subgroups, potentially leading to either an overestimation or underestimation of the true associations in our analysis. However, by employing automated segmentation techniques, biases associated with subjective interpretation of data, recall bias, and observer bias were avoided.

To effectively mitigate information bias in these scenarios, the implementation of quality control strategies becomes crucial. **Chapter 2** shows that employing visual inspection with exclusion of inaccurate brain segmentations yields the most favorable outcomes in reducing measurement errors. However, given the impracticality of this approach for large cohorts such as The Maastricht Study, the objective in **Chapter 2** was to identify an alternative (semi)automated strategy that was both time and resource efficient. Our investigation

led to determine that the exclusion of outliers based on Euler numbers was the most effective strategy within our sample. Euler numbers, a measure of surface reconstruction complexity (Dale, Fischl, & Sereno, 1999), showed a strong correlation with visual inspection, as previously established (Rosen et al., 2018). Conversely, the UK Biobank employs Qoala-T (Klapwijk, Van De Kamp, Van Der Meulen, Peters, & Wierenga, 2019) in combination with visual inspection of segmented cases to identify and exclude inaccurate segmentations (UK Biobank, 2022). In **Chapter 2** Qoala-T was also evaluated, exhibiting good performance, particularly in assessing measures of cortical surface area. Yet, the exclusion of outliers based on Euler numbers outperformed this tool when used in The Maastricht Study cohort.

Multiple comparison correction

The analysis of segmented brain structures entails examining multiple brain region estimates, which necessitates conducting multiple analyses to address the research question at hand. **Chapter 3** explores the associations between hyperglycemia and 13 distinct hippocampal volumes. **Chapter 4** focused on investigating the relationship between depression and 26 hippocampal volumes (13 volumes for each hemisphere). Furthermore, **Chapter 5** mapped the genetic overlap between depression and Alzheimer's disease into the brain across 95 brain volumes. Performing analyses on each individual brain region increases the likelihood of encountering type I errors. Therefore, correction for multiple comparisons is necessary. However, caution needs to be exercised as an overly strict adjustment may lead to type II errors (Rothman, 1990). To address this concern, raw p-values were reported, and a significance α threshold was defined by employing Matrix Spectral Decomposition (Nyholt, 2004). Matrix Spectral Decomposition allows for corrections of multiple comparisons while accounting for the intersecting nature of brain regions. Matrix Spectral Decomposition generates a matrix that determines the true independence among each brain region, providing an accurate count of the genuinely independent analyses being conducted. This results in a strategy that incorporates multiple comparison correction while being less stringent than a conservative Bonferroni correction.

Implications and directions for future research

Chapter 2 of this thesis offers a comprehensive comparison of various quality control strategies, providing a foundation for the selection of suitable quality control measures in large-scale neuroimaging studies. This comparison aims to serve as a reference for researchers in the field seeking to ensure the quality and reliability of their MRI data volumetric estimates. It is imperative to continue improving segmentation tools and developing new, time-efficient quality control strategies to ensure the use of high-quality data. The implications of poor data quality are far-reaching and can undermine the reliability of research findings, and affecting clinical and scientific decision-making processes. Hence, ensuring the highest possible data quality should be a top priority in neuroimaging research.

Contrary to our initial hypothesis, the findings presented in **Chapter 3** show that hyperglycemia is associated with lower volumes in hippocampus as a whole rather than in specific subfields. Interestingly, the absence of observable hippocampal differences during the prediabetes stage suggests the existence of a critical window of opportunity for interventions aimed at preventing neuronal degeneration. To gain a deeper understanding of when these changes occur and the precipitating factors involved, further research is necessary. Replicating the results using longitudinal data and conducting a more detailed assessment of the dose-response patterns observed in hyperglycemia could greatly enhance our understanding of the effects of both prediabetes and T2DM on brain structure. By investigating the timing and underlying factors associated with hippocampal changes, clinicians might be able to contribute to the development of strategies to prevent hippocampal degeneration and mitigate the resulting cognitive impairments. Continued research in this area holds promise for optimizing preventive measures and improving patient outcomes in the context of hippocampal degeneration.

The results presented in **Chapter 4** provide insights into the associations between depression and hippocampal subfield volumes. The findings suggest that the observed associations may either co-occur with, or follow, the onset of depressive symptoms, while no evidence supports the notion that differences in hippocampal volumes precede the onset of symptoms. Notably, the study highlights that the presence of chronic depression, rather than transient depression, is associated with differences in hippocampal volume. This underscores the importance of considering the course of depression, in addition to its mere presence, when designing future studies. Considering the longitudinal trajectory of depression can provide a more comprehensive understanding of the brain correlates associated with depressive disorders, thereby aiding in the resolution of the multitude of inconsistent results within this field of study. The heterogeneity of signs and symptoms that constitute the pathophysiology of depression, gives rise to the hypothesis that a more comprehensive classification of depression subtypes would be beneficial for a better understanding of the correlations between depression and brain structure. This, in turn, could lead to improved diagnostics and personalized treatment. Findings in **Chapter 4** serve as a valuable addition to the existing body of knowledge, bringing us closer to unraveling the complexities surrounding the brain correlates of depressive disorders.

The study of the genetic overlap between depression and Alzheimer's disease, conducted in **Chapter 5**, sheds additional light on the potential mechanisms linking these diseases to alterations in brain morphology. The identification of a shared genetic locus adds evidence to intriguing possibilities that are now being questioned, such as the existence of a distinct subtype of depression associated with Alzheimer's disease or even depression being a prodromal stage of this condition. These findings contribute to our evolving understanding of the intricate relationship between depression and Alzheimer's disease, opening avenues for further investigation into their shared etiological factors and potential clinical implications.

References

Alexopoulos, G. S., Buckwalter, K., Olin, J., Martinez, R., Wainscott, C., & Krishnan, K. R. R. (2002). Comorbidity of late life depression: an opportunity for research on mechanisms and treatment. *Biological psychiatry, 52*(6), 543-558.

American Psychiatric Association. (1994). DSM-IV: Diagnostic and statistical manual of mental disorders. In: Washington, DC: American Psychiatric Association.

Andersen, K., Lolk, A., Kragh-Sørensen, P., Petersen, N. E., & Green, A. (2005). Depression and the risk of Alzheimer disease. *Epidemiology*, 233-238.

Baan, C., & Feskens, E. (2001). Disease burden of diabetes mellitus type II in the Netherlands: incidence, prevalence and mortality. *Nederlands Tijdschrift Voor Geneeskunde, 145*(35), 1681-1685.

Barzilay, J. I., Spiekerman, C. F., Wahl, P. W., Kuller, L. H., Cushman, M., Furberg, C. D., . . . Savage, P. J. (1999). Cardiovascular disease in older adults with glucose disorders: comparison of American Diabetes Association criteria for diabetes mellitus with WHO criteria. *The Lancet, 354*(9179), 622-625.

Brown, E. M., Pierce, M. E., Clark, D. C., Fischl, B. R., Iglesias, J. E., Milberg, W. P., . . . Salat, D. H. (2020). Test-retest reliability of FreeSurfer automated hippocampal subfield segmentation within and across scanners. *NeuroImage, 210*, 116563.

Chen, P., Ganguli, M., Mulsant, B. H., & DeKosky, S. T. (1999). The temporal relationship between depressive symptoms and dementia: a community-based prospective study. *Archives of general psychiatry, 56*(3), 261-266.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage, 9*(2), 179-194.

de Flores, R., La Joie, R., Landeau, B., Perrotin, A., Mézenge, F., de La Sayette, V., . . . Chételat, G. (2015). Effects of age and Alzheimer's disease on hippocampal subfields: comparison between manual and FreeSurfer volumetry. *Human brain mapping, 36*(2), 463-474.

Fanselow, M. S., & Dong, H.-W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron, 65*(1), 7-19.

Fischl, B. (2012). FreeSurfer. *NeuroImage, 62*(2), 774-781.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., . . . Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American journal of epidemiology, 186*(9), 1026-1034.

Gabir, M. M., Hanson, R. L., Dabelea, D., Imperatore, G., Roumain, J., Bennett, P. H., & Knowler, W. C. (2000). The 1997 American Diabetes Association and 1999 World Health Organization criteria for hyperglycemia in the diagnosis and prediction of diabetes. *Diabetes care, 23*(8), 1108-1112.

Gavin III, J. R., Alberti, K., Davidson, M. B., & DeFronzo, R. A. (1997). Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes care, 20*(7), 1183.

Geerlings, M. I., & Gerritsen, L. (2017). Late-life depression, hippocampal volumes, and hypothalamic-pituitary-adrenal axis regulation: a systematic review and meta-analysis. *Biological psychiatry, 82*(5), 339-350.

Green, R. C., Cupples, L. A., Kurz, A., Auerbach, S., Go, R., Sadovnick, D., . . . Edeki, T. (2003). Depression as a risk factor for Alzheimer disease: the MIRAGE Study. *Archives of neurology, 60*(5), 753-759.

Howard, D. M., Adams, M. J., Shirali, M., Clarke, T.-K., Marioni, R. E., Davies, G., . . . Barbu, M. C. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature communications, 9*(1), 1470.

Iglesias, J. (2020, 2020-05-08). Hippocampal Subfields. Retrieved from https://surfer.nmr.mgh.harvard.edu/fswiki/HippocampalSubfields

Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., . . . Wald, L. L. (2015). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. *NeuroImage, 115*, 117-137.

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., . . . Athanasiu, L. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature genetics, 51*(3), 404-413.

Klapwijk, E. T., Van De Kamp, F., Van Der Meulen, M., Peters, S., & Wierenga, L. M. (2019). Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data. *Neuroimage, 189*, 116-129.

Knierim, J. J. (2015). The hippocampus. *Current Biology, 25*(23), R1116-R1121.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine, 16*(9), 606-613.

Marizzoni, M., Antelmi, L., Bosch, B., Bartrés-Faz, D., Müller, B. W., Wiltfang, J., . . . Nobili, F. (2015). Longitudinal reproducibility of automatically segmented hippocampal subfields: A multisite E uropean 3T study on healthy elderly. *Human brain mapping, 36*(9), 3516-3527.

Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. *General hospital psychiatry, 28*(1), 71-77.

Monereo-Sánchez, J., de Jong, J. J., Drenthen, G. S., Beran, M., Backes, W. H., Stehouwer, C. D., . . . Jansen, J. F. (2021). Quality control strategies for brain MRI segmentation and parcellation: Practical approaches and recommendations-insights from the Maastricht study. *Neuroimage, 237*, 118174.

Negeri, Z. F., Levis, B., Sun, Y., He, C., Krishnan, A., Wu, Y., . . . Benedetti, A. (2021). Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. *bmj, 375*.

Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics, 74*(4), 765-769.

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. *Archives of general psychiatry, 63*(5), 530-538.

Pettersson, A., Boström, K. B., Gustavsson, P., & Ekselius, L. (2015). Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. *Nordic journal of psychiatry, 69*(7), 497-508.

Quattrini, G., Pievani, M., Jovicich, J., Aiello, M., Bargalló, N., Barkhof, F., . . . Blin, O. (2020). Amygdalar nuclei and hippocampal subfields on MRI: Test-retest reliability of automated volumetry across different MRI sites and vendors. *NeuroImage, 218*, 116932.

Rosen, A. F., Roalf, D. R., Ruparel, K., Blake, J., Seelaus, K., Villa, L. P., . . . Elliott, M. A. (2018). Quantitative assessment of structural image quality. *Neuroimage, 169*, 407-418.

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology*, 43-46.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (Vol. 3): Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia.

Sämann, P. G., Iglesias, J. E., Gutman, B., Grotegerd, D., Leenings, R., Flint, C., . . . van Erp, T. G. (2018). FreeSurfer-based segmentation of hippocampal subfields: A review of methods and applications, with a novel quality control procedure for ENIGMA studies and other collaborative efforts. *Human Brain Mapping*.

Schram, M. T., Sep, S. J., van der Kallen, C. J., Dagnelie, P. C., Koster, A., Schaper, N., . . . Stehouwer, C. D. (2014). The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *European journal of epidemiology, 29*, 439-451.

Seiger, R., Hammerle, F. P., Godbersen, G. M., Reed, M. B., Spurny-Dworak, B., Handschuh, P., . . . Vanicek, T. (2021). Comparison and reliability of hippocampal subfield segmentations within FreeSurfer utilizing T1-and T2-weighted multispectral MRI data. *Frontiers in neuroscience, 15*.

Sun, X., Steffens, D. C., Au, R., Folstein, M., Summergrad, P., Yee, J., . . . Qiu, W. Q. (2008). Amyloid-associated depression: a prodromal depression of Alzheimer disease? *Archives of general psychiatry, 65*(5), 542-550.

Szklo, M., & Nieto, F. J. (2014). *Epidemiology: beyond the basics*: Jones & Bartlett Publishers.

Tae, W. S., Kim, S. S., Lee, K. U., Nam, E.-C., & Kim, K. W. (2008). Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology, 50*(7), 569.

Ubink-Veltmaat, L., Bilo, H., Groenier, K., Houweling, S., Rischen, R., & Meyboom-de Jong, B. (2003). Prevalence, incidence and mortality of type 2 diabetes mellitus revisited: a prospective population-based study in The Netherlands (ZODIAC-1). *European journal of epidemiology, 18*, 793-800.

UK Biobank. (2022, September 2022). UK Biobank Brain Imaging Documentation version 1.9. Retrieved from https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf

Vergoossen, L. W., Schram, M. T., de Jong, J. J., Stehouwer, C. D., Schaper, N. C., Henry, R. M., . . . Eussen, S. J. (2020). White matter connectivity abnormalities in prediabetes and type 2 diabetes: The Maastricht Study. *Diabetes care, 43*(1), 201-208.

Worker, A., Dima, D., Combes, A., Crum, W. R., Streffer, J., Einstein, S., . . . O'daly, O. (2018). Test–retest reliability and longitudinal analysis of automated hippocampal subregion volumes in healthy ageing and A lzheimer's disease populations. *Human Brain Mapping, 39*(4), 1743-1754.

World Health Organization. (2006). *Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation* (9241594934). Retrieved from https://apps.who.int/iris/handle/10665/43588

Yushkevich, P. A., Amaral, R. S., Augustinack, J. C., Bender, A. R., Bernstein, J. D., Boccardi, M., . . . Chakravarty, M. M. (2015). Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: towards a harmonized segmentation protocol. *Neuroimage, 111*, 526-541.

Zhang, T., Shaw, M., & Cherbuin, N. (2022). Association between type 2 diabetes mellitus and brain atrophy: a meta-analysis. *Diabetes & metabolism journal, 46*(5), 781-802.

ADDENDUM

SUMMARY

Brain MRI is an extraordinary non-invasive approach that enables the examination of the intricate structure of the human brain. The study of brain structure under various metabolic conditions such as type 2 diabetes mellitus (T2DM), psychiatric disorders like depression, or neurodegenerative diseases such as Alzheimer's may help to unveil the underlying biological causes and deepen our comprehension of their pathophysiology. However, this avenue of research poses an inherent complexity, requiring the inclusion of large sample sizes. In recent years, the integration of neuroimaging techniques into population-based cohorts has provided a powerful tool to enhance our understanding of neurobiological diseases. The study of large numbers of neuroimaging data offers an opportunity to address key research questions. Nonetheless, the inclusion of large datasets brings its own challenges, including the quality control of such a vast amount of data. Therefore, the first objective of this dissertation is to identify a suitable and time-efficient solution for ensuring good quality of brain segmentation data in large samples. Once this challenge is solved, high-quality, population-based large cohorts is utilized to conduct a comprehensive study of brain structure under specific conditions. This research aims to deepen our understanding of the intricate relationship between brain structure and various pathological conditions.

Chapter 1 presents a general introduction to the current state of neuroimaging in population-based cohorts. It provides an overview of the field, emphasizing the significance of studying brain structure within large cohorts. Two prominent cohorts utilized in this thesis, namely The Maastricht Study and the United Kingdom (UK) Biobank, are introduced, offering valuable resources for investigating the relationship between neuroimaging data and various health conditions. The importance of ensuring data quality is emphasized, as it is crucial for drawing reliable conclusions from analyses. Additionally, a comprehensive overview of the current understanding of neuroimaging in relation to the three conditions under investigation in this thesis (T2DM, depression, and Alzheimer's disease) is provided, serving as the foundation for subsequent chapters that delve deeper into the study of brain structure under these specific conditions.

Chapter 2 explores quality control strategies for brain segmentation data. Recognizing the impracticality of manual quality control for large cohorts, the objective is to establish an efficient and reproducible automated quality control approach to ensure the accuracy of brain segmentation data in The Maastricht Study dataset. With this aim, a subsample of The Maastricht Study dataset is carefully selected. Visual inspection and manual editing is performed to rectify any inaccurate segmentations. Furthermore, multiple automated quality control strategies are implemented. Both manual and automated quality control strategies are tested on their ability to reduce the unexplained variance generated in a regression analysis. Among the strategies tested, manual intervention yields the most significant reduction in unexplained variance, making it the preferred strategy for ensuring the highest quality of brain segmentation. However, an automated strategy emerges as a viable alternative. Specifically, the exclusion of outliers based on Euler numbers exhibits superior performance compared to other automated approaches. This strategy is selected as the optimal solution

for maintaining high-quality brain segmentations when time efficiency is a priority. The chosen automated strategy is subsequently applied to the full The Maastricht Study sample, ensuring uniformity and accuracy of brain segmentation data used in subsequent chapters.

Chapter 3 utilizes The Maastricht Study dataset to investigate the associations between prediabetes, T2DM, continuous measures of hyperglycemia, and the volume of hippocampal subfields. The findings reveal that T2DM is associated with reduced hippocampal volumes, independent of demographic, lifestyle, and cardiovascular risk factors. The association appears to be widespread, affecting most hippocampal subfields. However, no significant associations between prediabetes and hippocampal volumes are observed. Nonetheless, analysis using continuous measures of hyperglycemia and trend analysis indicate a continuous dose-response association between glucose metabolism and lower hippocampal volumes.

In **Chapter 4**, the focus shifts to examining the links between hippocampal subfield volumes and the prevalence, course, and incidence of depressive symptoms. Longitudinal data on depressive symptoms is obtained through the PHQ-9 questionnaire. The results show specific hippocampal subfields are associated with prevalent depression, specifically with a chronic course. Longitudinal analyses find some evidence that smaller volume in left HATA is associated with a risk of incident depression with a chronic course. This results could be capturing a biological foundation for the development of chronic depression, and further stress the need to discriminate between subtypes of depression to unravel its biological underpinnings.

Deepening in the study of depression, **Chapter 5** investigates the genetic links between depression and Alzheimer's disease using UK Biobank data. The analysis shows 98 overlapping causal genetic variants, identifying a shared single nucleotide polymorphism located in the TMEM106B gene. Subsequently, the associations of this gene with brain morphometry are assessed. The findings indicate that the presence of this gene is associated with higher surface area in the inferior temporal gyrus, larger volumes in the corpus callosum, and a lower volume of the third ventricle. These results revealed a previously undisclosed genetic overlap between depression and Alzheimer's disease, which was previously obscured by mixed directions of effect.

Chapter 6 provides a comprehensive discussion of the key findings, limitations of the statistical approaches, and the clinical relevance of the content presented in this thesis.

SAMENVATTING

Brein MRI is een niet-invasieve beeldvormende techniek die het mogelijk maakt de ingewikkelde structuur van het menselijk brein te onderzoeken. Het bestuderen van de hersenstructuur bij verschillende metabole aandoeningen, zoals diabetes mellitus type 2 (T2DM), psychiatrische stoornissen zoals depressie of neurodegeneratieve ziekten zoals de ziekte van Alzheimer, kan helpen om de onderliggende biologische oorzaken te vinden en ons begrip van hun pathofysiologie te verdiepen. Dit type onderzoek brengt echter een inherente complexiteit met zich mee, waardoor een grote steekproefomvang vereist is. In de afgelopen jaren heeft de toepassing van neuroimaging-technieken in populatie gebaseerde cohortonderzoeken een krachtig middel opgeleverd om ons begrip van neurobiologische ziekten te vergroten. Het bestuderen van grootschalige neuroimaging-gegevens biedt de mogelijkheid om belangrijke onderzoeksvragen te beantwoorden. Desalniettemin brengt het gebruik van grote datasets zijn eigen uitdagingen met zich mee, waaronder de kwaliteitscontrole van de grote hoeveelheid data. Daarom is het eerste doel van dit proefschrift het identificeren van een geschikte en tijdbesparende oplossing om een goede kwaliteit van hersensegmentaties in een grote steekproef te garanderen. Hierna wordt de gevonden oplossing toegepast op grote cohorten met hoogkwalitatieve data om uitgebreid onderzoek van de hersenstructuur onder specifieke condities mogelijk te maken. Dit onderzoek heeft tot doel ons begrip van de ingewikkelde relatie tussen hersenstructuur en verschillende pathologische aandoeningen te verdiepen.

Hoofdstuk 1 geeft een algemene inleiding tot de huidige stand van zaken op het gebied van neuroimaging in bevolkingscohorten. Het geeft een overzicht van het onderzoeksveld en benadrukt het belang van het bestuderen van de hersenstructuur binnen grote cohorten. Twee prominente cohorten die in dit proefschrift worden gebruikt, namelijk De Maastricht Studie en de UKBiobank in het Verenigd Koninkrijk, worden geïntroduceerd en bieden waardevolle bronnen voor het onderzoeken van de relatie tussen neuroimaging-gegevens en verschillende gezondheidsaandoeningen. Het belang van het waarborgen van datakwaliteit wordt benadrukt, aangezien dit cruciaal is voor het trekken van betrouwbare conclusies. Daarnaast wordt een uitgebreid overzicht gegeven van het huidige begrip van neuroimaging in relatie tot de drie aandoeningen die in dit proefschrift worden onderzocht (T2DM, depressie en de ziekte van Alzheimer), dat als basis dient voor volgende hoofdstukken die dieper ingaan op de relatie tussen hersenstructuur en deze aandoeningen.

Hoofdstuk 2 verkent strategieën voor kwaliteitscontrole van hersensegmentaties. Omdat handmatige kwaliteitscontrole voor grote cohorten praktisch niet uitvoerbaar is, is het doel om een efficiënte en reproduceerbare geautomatiseerde kwaliteitscontrole te vinden, die de nauwkeurigheid van hersensegmentaties in de dataset van De Maastricht Studie kan waarborgen. Hiervoor werd een subset van De Maastricht Studie zorgvuldig geselecteerd. Visuele inspectie en handmatige bewerking werden uitgevoerd om onnauwkeurige segmentaties te corrigeren. Bovendien werden meerdere geautomatiseerde kwaliteitscontrolestrategieën geïmplementeerd. Zowel handmatige als geautomatiseerde kwaliteitscontrolestrategieën werden getest op hun vermogen om de onverklaarde variantie die wordt gegenereerd in een regressieanalyse te verminderen.

Van de geteste strategieën levert handmatige interventie de meest significante vermindering van onverklaarde variantie op, waardoor het de voorkeursstrategie is om de hoogste kwaliteit van hersensegmentatie te waarborgen. Een geautomatiseerde strategie komt echter naar voren als een goed alternatief. Met name de exclusie van uitschieters op basis van Euler-getallen vertoont superieure prestaties in vergelijking met andere geautomatiseerde benaderingen. Deze strategie is een optimale oplossing voor het behouden van hersensegmentaties van hoge kwaliteit wanneer tijdsefficiëntie een prioriteit is. De gekozen geautomatiseerde strategie is vervolgens toegepast op de volledige De Maastricht Studie dataset, waardoor de uniformiteit en nauwkeurigheid van de hersensegmentaties die in volgende hoofdstukken worden gebruikt, wordt gegarandeerd.

Hoofdstuk 3 gebruikt de dataset van De Maastricht Studie om de associaties tussen prediabetes, T2DM, continue maten van hyperglycemie en het volume van verschillende onderdelen van de hippocampus te bestuderen. De bevindingen onthullen dat T2DM geassocieerd is met kleinere hippocampusvolumes, onafhankelijk van demografische, levensstijl- en cardiovasculaire risicofactoren. De associatie lijkt generiek te zijn en betreft de meeste onderdelen van de hippocampus. Er worden echter geen significante associaties tussen prediabetes en volumes van onderdelen van de hippocampus gevonden. Desalniettemin wijzen analyses met behulp van continue metingen van hyperglycemie en trendanalyse op een continue dosis-respons relatie tussen glucosemetabolisme en lagere hippocampusvolumes.

In **Hoofdstuk 4** worden de verbanden tussen de volumes van de onderdelen van de hippocampus en de prevalentie, het beloop en de incidentie van depressieve symptomen onderzocht. Longitudinale gegevens over depressieve symptomen worden verkregen via de PHQ-9-vragenlijst. De resultaten tonen aan dat volumes van specifieke onderdelen van de hippocampus geassocieerd zijn met prevalente depressie, in het bijzonder met een chronisch beloop. Longitudinale analyses tonen aan dat een kleiner volume in de linker HATA geassocieerd is met een risico op incidente depressie met een chronisch beloop. Deze resultaten geven mogelijk een biologische verklaring voor de ontwikkeling van chronische depressie, en benadrukken de noodzaak om onderscheid te maken tussen subtypes van depressie om het biologische mechanisme ervan te ontrafelen.

Hoofdstuk 5, verdiept zich in depressie en onderzoekt de genetische verbanden tussen depressie en de ziekte van Alzheimer met behulp van UK Biobank-gegevens. De analyse toont 98 overlappende oorzakelijke genetische varianten, die een gemeenschappelijk single-nucleotide polymorfisme in het TMEM106B-gen identificeren. Vervolgens worden de associaties van dit gen met breinsegmentaties onderzocht. De bevindingen geven aan dat de aanwezigheid van dit gen geassocieerd is met een groter volume van de inferieure temporale gyrus, grotere volumes van het corpus callosum en een kleiner volume van het derde ventrikel. Deze resultaten laten een voorheen niet bekende genetische overlap zien tussen depressie en de ziekte van Alzheimer, die eerder verborgen bleef door de tegengestelde richting van de afzonderlijke effecten.

Hoofdstuk 6 geeft een uitgebreid discussie van de belangrijkste bevindingen, beperkingen van de statistische benaderingen en de klinische relevantie van dit proefschrift.

RESUMEN

La resonancia magnética cerebral es una extraordinaria técnica no invasiva que permite el examen de la estructura del cerebro humano. El estudio de la morfología cerebral bajo diversas condiciones metabólicas como la diabetes mellitus tipo 2 (DM2), trastornos psiquiátricos como la depresión, o enfermedades neurodegenerativas como el Alzheimer, puede ayudar a desvelar las causas biológicas subyacentes de estas patologías y profundizar en nuestra comprensión de su fisiopatología. Sin embargo, debido a la complejidad de esta vía de investigación se requiere el uso de bases de datos de grandes dimensiones. En los últimos años, la integración de técnicas de neuroimagen en cohortes poblacionales ha proporcionado una herramienta muy útil para mejorar nuestra comprensión de las enfermedades neurobiológicas, ya que el uso de un gran número de datos en este campo ofrece la oportunidad de abordar cuestiones de investigación más complejas. No obstante, el uso de grandes bases de datos trae desafíos específicos, entre los cuales se incluye el control de calidad de una cantidad de información tan grande. Por lo tanto, el primer objetivo de esta tesis es identificar una solución reproducible y eficiente para garantizar una buena calidad de la segmentación de imágenes cerebrales en muestras de gran tamaño. Una vez resuelto este desafío, se utilizan cohortes poblacionales de grandes dimensiones que cuentan con una neuroimagen de alta calidad para realizar un estudio integral de la estructura del cerebro en condiciones específicas. Esta investigación tiene como objetivo profundizar en nuestra comprensión de la compleja relación entre la estructura del cerebro y las diversas patologías.

El **Capítulo 1** presenta una introducción general al estado actual de la neuroimagen en cohortes poblacionales. Proporciona una descripción general del campo, y subraya la importancia de estudiar la estructura cerebral mediante el uso de grandes bases de datos. Se presentan dos cohortes destacadas y utilizadas en esta tesis, a saber, *The Maastricht Study* y el *United Kingdom (UK) biobank*, que ofrecen recursos valiosísimos para la investigación del cerebro bajo diferentes condiciones. Se enfatiza la importancia de asegurar la calidad de los datos, ya que es crucial para sacar conclusiones fiables de los análisis. Además, se proporciona una descripción general de los conocimientos actuales sobre neuroimagen en las tres patologías que se investigan en esta tesis (DM2, depresión y enfermedad de Alzheimer), lo que sirve de base para los capítulos posteriores. Estos profundizan en el estudio de la estructura cerebral bajo dichas condiciones.

El **Capítulo 2** explora las estrategias de control de calidad para la segmentación de imágenes cerebrales. Tras reconocer la impracticabilidad del control de calidad manual para grandes bases de datos, el objetivo es establecer un enfoque automatizado, eficiente y reproducible, para garantizar la precisión de los datos de segmentación de neuroimagen en la base de datos de *The Maastricht Study*. Con este objetivo, se selecciona cuidadosamente una muestra a la que se realiza una inspección visual y una edición manual para corregir cualquier segmentación inexacta. Además, se implementan múltiples estrategias de control de calidad automatizado. Tanto la estrategia de control de calidad manual como las automatizadas se ponen a prueba en cuanto a su capacidad para reducir la varianza no explicada generada en un análisis de regresión. Entre las estrategias probadas, la intervención manual produce la reducción más significativa en la varianza no

explicada, lo que la convierte en la estrategia preferida para garantizar la máxima calidad de segmentación. Sin embargo, la exclusión de valores atípicos basados en números de Euler resulta ser la estrategia automatizada más eficaz, ya que muestra un rendimiento superior en la reducción de varianza inexplicada en comparación con otros enfoques automatizados. Esta estrategia se selecciona como la solución óptima para mantener segmentaciones de neuroimagen de alta calidad cuando la eficiencia es una prioridad. Esta estrategia se aplica entonces a la base de datos completa de *The Maastricht Study*, y se garantiza así la uniformidad y la precisión de los datos de segmentación de neuroimagen utilizados en los capítulos posteriores.

El **Capítulo 3** utiliza la base de datos de *The Maastricht Study* para investigar las asociaciones entre la prediabetes, la DM2, y las medidas continuas de hiperglucemia, con el volumen de las subestructuras del hipocampo. Los hallazgos revelan que la DM2 se asocia con volúmenes hipocampales reducidos, independientemente de los factores demográficos, de estilo de vida y de riesgo cardiovascular. La asociación parece estar muy extendida y afecta a la mayoría de las subestructuras del hipocampo. Sin embargo no se observan asociaciones significativas entre la prediabetes y ninguna subestructura del hipocampo. No obstante, el análisis de medidas continuas de hiperglucemia y el análisis de tendencias indican una asociación dosis-respuesta continua entre el metabolismo de la glucosa y los volúmenes del hipocampo.

En el **Capítulo 4**, el enfoque cambia para examinar los vínculos entre las subestructuras del hipocampo y la prevalencia, el curso y la incidencia de los síntomas depresivos. Los datos longitudinales sobre estos síntomas se obtienen a través del cuestionario PHQ-9. Los resultados muestran que los volúmenes de subestructuras especificas del hipocampo están asociados con la depresión prevalente, específicamente con un curso crónico. Los análisis longitudinales muestran alguna evidencia de que un volumen más pequeño en la estructura de transición entre el hipocampo y la amígdala izquierda está asociado con un riesgo de depresión incidente con un curso crónico. Estos resultados podrían capturar una base biológica para el desarrollo de la depresión crónica, y enfatizar además la necesidad de discriminar entre los subtipos de depresión para desentrañar sus fundamentos biológicos.

Profundizando en el estudio de la depresión, el **Capítulo 5** investiga los vínculos genéticos entre la depresión y la enfermedad de Alzheimer utilizando datos del *UK Biobank*. El análisis muestra 98 variantes genéticas causales superpuestas, e identifica un polimorfismo de un solo nucleótido compartido ubicado en el gen TMEM106B. Posteriormente se evalúan las asociaciones de este gen con la morfometría cerebral. Los hallazgos indican que la presencia de este gen se asocia con una mayor superficie en la circunvolución temporal inferior, mayores volúmenes en el cuerpo calloso y un menor volumen del tercer ventrículo. Estos resultados revelan una superposición genética nunca detectada previamente entre la depresión y la enfermedad de Alzheimer, posiblemente oscurecida anteriormente por direcciones de efecto mixtas.

El **Capítulo 6** proporciona una discusión exhaustiva de los hallazgos clave, las limitaciones de los enfoques estadísticos y la relevancia clínica del contenido presentado en esta tesis.

IMPACT PARAGRAPH

Main findings

The goal of this thesis was to examine quality control in human brain segmentations and investigate the clinical applications of neuroimaging in large population-based cohorts. As a result, several quality control strategies were initially performed on a subsample of The Maastricht Study to compare their performance. Subsequently, high-quality neuroimaging data from The Maastricht Study were utilized to investigate the associations of depression and diabetes with the volume of the hippocampal subfields. Additionally, the genetic overlap between depression and Alzheimer's disease was examined, and the potential morphological correlates of this overlap across brain structures were explored.

It was found that the best results were achieved through manual quality control via visual inspection, followed by the exclusion of outliers based on Euler numbers. Furthermore, an association was discovered between type 2 diabetes mellitus (T2DM) and generalized atrophy of the hippocampus. Additionally, an association was observed between specific subfields of the hippocampus and chronic depressive symptoms. Notably, larger volumes in the molecular layer and smaller volumes in Cornu Ammonis 2/3 were associated with chronic depressive symptoms. Finally, an overlapping locus in the TMEM106B gene between depression and Alzheimer's disease was identified, which was found to be associated with larger volumes of the corpus callosum and a greater surface area in the inferior temporal gyrus.

Relevance

The use of high-quality imaging data is crucial for ensuring the reliability of research findings. Insufficient data quality can significantly impact the trustworthiness of the results obtained. In this thesis, a novel perspective is presented regarding the assessment of MRI segmentation quality, which serves as a fundamental basis for addressing numerous important research inquiries. The conclusions reached through this study are especially relevant in large scale population studies, where the vast amount of available data limits the ability to apply manual quality control strategies.

Furthermore, in recent years, there has been a rise in the prevalence of diabetes, depression, and Alzheimer's disease. This increase can be attributed to various factors associated with the fast-paced and stressful nature of modern life, rapid urbanization, and the aging population. However, the increased numbers of diagnosed cases are also partly a result of the growing awareness surrounding these diseases. This increased awareness enables interventions to be implemented at earlier stages. By identifying the association of hippocampal volume changes and different stages of T2DM pathology, the study emphasizes the importance of recognizing and addressing prediabetes as a potential risk factor for cognitive decline and neurodegenerative processes.

Additionally, while the association between depression and specific hippocampal subfields remains less clear, the results shed light on the existence of distinct depression subtypes with divergent patterns of brain damage. Notably, chronic depression emerges as a condition that appears to exhibit heightened vulnerability to such damage compared to other types of depression. This highlights the need for increased attention and specialized interventions aimed at preserving brain health in individuals with chronic depression, as these findings have significant implications for understanding the long-term consequences of depressive disorders. Moreover, the study reveals a potential genetic underpinning for the frequently observed clinical overlap between depression and Alzheimer's disease, although the available evidence is currently limited and requires replication.

These findings carry significant implications for both clinical practice and future research in the field of neurodegenerative disorders, urging a comprehensive approach towards early intervention and targeted preservation of brain health in at-risk populations. Conducting basic neuroimaging research helps delving into the neurobiological changes associated with these conditions. This insight allows us to comprehend better the underlying pathophysiological processes occurring in the brain during different disease states. Finally, understanding when and where these changes take place provides crucial information about the diseases themselves.

Target group

The target audience for our research findings encompasses neuroscientists, epidemiologists, imaging engineers, and clinicians in internal medicine, psychiatry, neurology and neuropsychology. Specifically, there is a need for additional investigations to enhance neuroimaging quality during acquisition and preprocessing, with a focus on segmentation. This is to avoid acquisition artifacts, improve image quality and ultimately achieving a reliable brain segmentation. Efforts should also be directed towards finding and implementing strategies to identify and improve cases with poor segmentation quality. Furthermore, it is crucial to address the progression of brain damage in longitudinal studies. Understanding how brain damage evolves over time is essential for the development of effective interventions and preventive measures. In addition, researchers should delve into the definition and subtypes of depression, exploring their correlation with brain morphology and its progression over time.

The research conducted in this thesis can contribute to efforts towards early recognition of brain changes associated with metabolic or neurodegenerative disorders. The rising prevalence of T2DM, depression, and Alzheimer's disease have substantial societal and economic impacts. This makes the findings presented in this thesis of significant relevance for the general public. These conditions impose a heavy burden on individuals, families, and healthcare systems worldwide. By shedding light on the intricate relationships these conditions and the brain structure, this research contributes a crucial piece to the puzzle in understanding how brain damage develops and, more importantly, how it can potentially be prevented.

Dissemination activities

Our research findings have received significant recognition and engagement within the academic community. They have been extensively discussed at various research meetings held at Maastricht University, including several years at the annual School for Mental Health and Neuroscience (MHeNs) research day. Additionally, our results on segmentation quality were presented and discussed at the University of Cape Town in South Africa in an invited lecture as a part of a seminar organized between Cape Town University and the Norwegian Center for Mental Disorders Research. Moreover, our findings regarding T2DM were showcased at prominent conferences such as the Research Accelerating Psychosocial Innovations in Diabetes (RAPID) conference, and the Organization for Human Brain Mapping (OHBM) meeting in 2022. It is our aspiration that our dissemination efforts have fostered a broader interest and encouraged researchers worldwide to explore and expand upon the knowledge and insights we have contributed.

Conclusion

In conclusion, this thesis aimed to examine quality control in brain segmentations and investigate the clinical applications of neuroimaging in large population-based cohorts. The study involved comparing different quality control strategies and analyzing high-quality neuroimaging data to explore the associations between depression, diabetes, and the volume of hippocampal subfields. Additionally, the genetic overlap between depression and Alzheimer's disease was examined. The findings revealed the effectiveness of manual quality control and identified associations between T2DM and hippocampal atrophy, specific hippocampal subfields and chronic depressive symptoms, and a genetic locus shared between depression and Alzheimer's disease. These findings have significant implications for improving diagnosis precision, personalized treatment, and understanding the underlying mechanisms of brain damage.

ACKNOWLEDGMENTS

This thesis would never have been possible without the help and support of many people. I would like to thank everyone who helped in any way.

I would like to express my sincere gratitude to my promotion team. David Linden, Miranda Schram, and Jacobus Jansen: I am truly thankful for your invaluable guidance and support throughout this journey. David, your instructive input has been instrumental in shaping my research, and I am very grateful for the freedom and trust you placed in me to pursue my desired projects. Miranda, I am immensely grateful for your kindheartedness and support, which have been a source of strength throughout this process. I would also like to thank you for your invaluable assistance in writing, and in the fields of statistics and epidemiology. Lastly, I would like to express my deepest appreciation to Jaap for his expertise in neuroimaging, resourcefulness in problem-solving, and willingness to assist: Your daily support and guidance have been an unwavering source of motivation. I am genuinely grateful for your presence and availability whenever I needed any form of assistance.

Next, I extend my gratitude to the esteemed members of the assessment committee. Prof. dr. Thérèse van Amelsvoort, Prof. dr. Pieter Visser, Dr. Liesbeth Goosens, Prof. dr. Meike Vernooij, and Dr. Thomas van Sloten: thank you for your valuable time and effort dedicated to reviewing this thesis.

I would also like to express my gratitude to all the co-authors for their invaluable contributions and insightful input. Joost de Jong, Gerald Drenthen, Magdalena Beran, Walter Backes, Coen Stehouwer, Sebastian Köhler, Martin van Boxtel, Abraham Kroon, Jeroen Kooman, Casper Schalkwijk, Oleksandr Frei, Kevin O'Connell, Alexey Shadrin, Olav Smeland, Lars Westlye, Ole Andreassen, Tobias Kaufmann, and Dennis van der Meer: I sincerely appreciate your collaboration and constructive feedback to the scientific publications within this thesis.

I would like to extend my gratitude to my colleagues, despite the shortened time we had together due to COVID. Gerald, Lisanne, Harmen, Inge, Marieke, Elles, May, Merel, Paulien, Magdalena, and Darius: The moments we spent in the office were enjoyable, and your scientific input was truly enriching. Thank you for your company and for making our interactions both productive and fulfilling. A special thanks to my colleagues Joost, Jos, and Jan for their invaluable camaraderie and assistance in all the technical aspects of image processing. I deeply appreciate their expertise and support, which have been instrumental in overcoming challenges and achieving the desired outcomes: Thank you for being reliable partners and contributing to the success of this research. A special thanks to Dennis, with whom my love for science blossomed. I am immensely grateful for the evenings spent in the garden, indulging in passionate discussions about science over a glass of wine. Your unwavering enthusiasm and encouragement have pushed me to dream big and pursue my scientific aspirations with greater zeal. Thank you for being an inspiring companion on this journey of intellectual growth and discovery.

I would also like to express my gratitude to all the participants and employees of The Maastricht Study: Without your invaluable cooperation, this thesis would not have been feasible. Thank you for your willingness to contribute, you are an integral part of this research.

Finally, I would like to thank my family and friends:

Mamá, ninguno de mis logros, incluyendo esta tesis, habría sido posible sin tu incansable esfuerzo y persistencia. Agradezco profundamente que hayas sido un modelo a seguir, por tu amor ilimitado, tu respaldo constante y por nunca permitirme perderme por el camino. Xavi, quiero agradecerte por tantos años de apoyo y ayuda incondicional, por animarme a crecer siempre, y por darme fortaleza. Todo esto no habría sido posible si tu no estuvieras en mi vida. Papá, gracias por la música, por impulsarme a conocerme y a ser una mejor persona. Tus enseñanzas han sido invaluables. Carla, gracias por ser la mejor hermana, por las risas, los viajes, el apoyo y los ratos que te has pasado leyendo mi tesis, aunque te pareciera un tostón. Abuela Ceci, gracias por tu cariño y paciencia inagotables. Yaya Flori y yayo Ramón, gracias por quererme como propia, por vuestro apoyo y vuestro amor durante tantos años. Estoy inmensamente agradecida. Dani, gracias por enseñarme que la paz y la magia son posibles. Gracias por acompañarme con tanta compresión en un camino que ha sido tan difícil. Yoel, Danna, Ana, Anael, y Fredi gracias por vuestra paciencia, por vuestro apoyo y vuestro amor. Magdalena, Sam y Quini, gracias por ser mi luz en Maastricht. Danna, Josep y Marc, gracias por inspirarme a soñar un poquito más allá. A mis primos, primas, tíos y tías, gracias por vuestro cariño. A todos, os estoy enormemente agradecida.

CURRICULUM VITAE

Jennifer Monereo Sánchez, born on July 26th, 1988, in Barcelona, Spain, holds an associate degree in Radiology obtained in 2010, followed by a Bachelor's degree in Psychology in 2014. She later pursued a Master of Science in Neuroscience at the Norwegian University of Science and Technology. As a part of the master's program, she enrolled in an internship at the Norwegian Centre for Mental Disorders Research. Jennifer successfully graduated from the master's program in 2017 and subsequently worked as a research technician at the same center, specializing in MRI data acquisition, preprocessing, quality control, and storage for various research projects. In 2019, she embarked on her Ph.D. project at the Department of Radiology and Nuclear Medicine at the Maastricht University Medical Center in the Netherlands. Her PhD project focused on the methodological considerations and clinical applications of large population-based neuroimaging cohorts. The research presented in the current thesis was conducted under the supervision of Prof. Dr. David E.J. Linden, Prof. Dr. Miranda T. Schram, and Dr. Jacobus F.A. Jansen.

PUBLICATIONS

Journal Contributions

This thesis

Monereo-Sánchez, J., de Jong, J. J., Drenthen, G. S., Beran, M., Backes, W. H., Stehouwer, C. D., . . . Jansen, J. F. (2021). Quality control strategies for brain MRI segmentation and parcellation: Practical approaches and recommendations-insights from the Maastricht study. *Neuroimage*, 237, 118174.

Monereo-Sánchez, J., Schram, M. T., Frei, O., O'Connell, K., Shadrin, A. A., Smeland, O. B., . . . van der Meer, D. (2021). Genetic overlap between Alzheimer's disease and depression mapped onto the brain. *Frontiers in Neuroscience*, 15, 653130.

Monereo-Sánchez, J., Jansen, J. F., Köhler, S., van Boxtel, M.P.J., Backes, W. H., Stehouwer, C. D., Kroon, A.A., Kooman, J.P., Schalkwijk, C.G., Linden, D.E.J., Schram, M.T. (2023). The association of prediabetes and type 2 diabetes with hippocampal subfields volume: The Maastricht Study. NeuroImage: Clinical, 103455.

Monereo-Sánchez, J., Jansen, J. F., van Boxtel, M.P.J., Backes, W. H., Köhler, S., Stehouwer, C. D., Linden, D.E.J., Schram, M.T. (2023). Association of hippocampal subfield volumes with prevalence, course and incidence of depressive symptoms - The Maastricht Study. The British Journal of Psychiatry 2023, 1–8.

Other

Richard, G., Kolskår, K., Sanders, A.-M., Kaufmann, T., Petersen, A., Doan, N. T., […], **Monereo-Sánchez, J.,** […], Dørum, E. S. (2018). Assessing distinct patterns of cognitive aging using tissue-specific brain age prediction based on diffusion tensor imaging and brain morphometry. *PeerJ*, 6, e5908.

Richard, G., Kolskår, K., Ulrichsen, K. M., Kaufmann, T., Alnæs, D., Sanders, A.-M., […], **Monereo-Sánchez, J.,** […], Ihle-Hansen, H. (2020). Brain age prediction in stroke patients: Highly reliable but limited sensitivity to cognitive performance and response to cognitive training. *NeuroImage: Clinical*, 25, 102159.

van der Meer, D., Frei, O., Kaufmann, T., Chen, C.-H., Thompson, W. K., O'Connell, K. S., […], **Monereo-Sánchez, J.,** […], Dale, A. M. (2020). Quantifying the polygenic architecture of the human cerebral cortex: Extensive genetic overlap between cortical thickness and surface area. *Cerebral Cortex*, 30(10), 5597-5603.

Van der Meer, D., Rokicki, J., Kaufmann, T., Córdova-Palomera, A., Moberget, T., Alnæs, D., […], **Monereo-Sánchez, J.,** […], Sønderby, I. E. (2020). Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. *Molecular psychiatry*, 25(11), 3053-3065.

Monereo-Sánchez, J., Luykx, J. J., Pinzon-Espinosa, J., Richard, G., Motazedi, E., Westlye, L. T., . . . van der Meer, D. (2021). Diphtheria and tetanus vaccination history is associated with lower odds of COVID-19 hospitalization. *Frontiers in immunology*, 12, 749264.

Sønderby, I. E., Van der Meer, D., Moreau, C., Kaufmann, T., Walters, G. B., Ellegaard, M., […], **Monereo-Sánchez, J.,** […], Andersson, M. (2021). 1q21. 1 distal copy number variants are associated with cerebral and cognitive alterations in humans. *Translational psychiatry*, 11(1), 182.

Van Der Meer, D., Kaufmann, T., Shadrin, A. A., Makowski, C., Frei, O., Roelfs, D., […], **Monereo-Sánchez, J.,** […], Alnæs, D. (2021). The genetic architecture of human cortical folding. *Science advances*, 7(51), eabj9446.

Ulrichsen, K. M., Kolskår, K. K., Richard, G., Alnæs, D., Dørum, E. S., Sanders, A.-M., […], **Monereo-Sánchez, J.,** […], Ihle-Hansen, H. (2021). Structural brain disconnectivity mapping of post-stroke fatigue. *NeuroImage: Clinical*, 30, 102635.

Siafarikas, N., Alnæs, D., **Monereo-Sánchez, J.**, Lund, M. J., Selbaek, G., Stylianou-Korsnes, M., . . . Fladby, T. (2021). Neuropsychiatric symptoms and brain morphology in patients with mild cognitive impairment and Alzheimer's disease with dementia. *International Psychogeriatrics*, 33(11), 1217-1228.

Kolskår, K. K., Richard, G., Alnæs, D., Dørum, E. S., Sanders, A. M., Ulrichsen, K. M., […], **Monereo-Sánchez, J.,** […], Westlye, L. T. (2021). Reliability, sensitivity, and predictive value of fMRI during multiple object tracking as a marker of cognitive training gain in combination with tDCS in stroke survivors. *Human brain mapping*, 42(4), 1167-1181.

Kolskår, K. K., Ulrichsen, K. M., Richard, G., Dørum, E. S., de Schotten, M. T., Rokicki, J., […], **Monereo-Sánchez, J.,** […], Nordvik, J. E. (2022). Structural disconnectome mapping of cognitive function in poststroke patients. *Brain and Behavior*, 12(8), e2707.

Sønderby, I. E., Ching, C. R., Thomopoulos, S. I., van Der Meer, D., Sun, D., Villalon-Reina, J. E., […], **Monereo-Sánchez, J.,** […], Armstrong, N. J. (2022). Effects of copy number variations on brain structure and risk for psychiatric illness: Large-scale studies from the ENIGMA working groups on CNVs. *Human brain mapping*, 43(1), 300-328.

Van der Meer, D., Shadrin, A. A., O'Connell, K., Bettella, F., Djurovic, S., Wolfers, T., […], **Monereo-Sánchez, J.,** […], Melle, I. (2022). Boosting schizophrenia genetics by utilizing genetic overlap with brain morphology. *Biological psychiatry*, 92(4), 291-298.

van Kruining, D., Losen, M., Crivelli, S. M., de Jong, J. J., Jansen, J. F., Backes, W. H., **Monereo-Sánchez, J.,** …, & Martinez-Martinez, P. (2023). Plasma ceramides relate to mild cognitive impairment in middle-aged men: The Maastricht Study. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 15(3), e12459.

Research Dissemination

Monereo-Sánchez, J., de Jong, J. J., Drenthen, G. S., Beran, M., Backes, W. H., Stehouwer, C. D., . . . Jansen, J. F. Quality control of structural MR brain images in the Maastricht Study. 12th Annual School for Mental Health and Neuroscience Research Day: New technologies in Neuroscience and Mental Health (2019). Maastricht, the Netherlands (E-pitch)

Monereo-Sánchez, J., de Jong, J. J., Drenthen, G. S., Beran, M., Backes, W. H., Stehouwer, C. D., . . . Jansen, J. F. Quality Control (QC) of structural brain MRI in the Maastricht Study: Can automated QC replace manual QC?. 12th ISMRM Benelux chapter Meeting (2020). Arnhem, the Netherlands (Poster)

Monereo-Sánchez, J., de Jong, J. J., Drenthen, G. S., Beran, M., Backes, W. H., Stehouwer, C. D., . . . Jansen, J. F. Quality control of brain segmentations in large cohorts. Cape Town University neuroimaging symposium (2022), Cape Town, South Africa (Invited speaker)

Monereo-Sánchez, J., Jansen, J. F., Köhler, S., van Boxtel, M.P.J., Backes, W. H., Stehouwer, C. D., Kroon, A.A., Kooman, J.P., Schalkwijk, C.G., Linden, D.E.J., Schram, M.T. Association of prediabetes and T2DM with hippocampus subfields atrophy: The Maastricht Study. 14th Annual School for Mental Health and Neuroscience Research Day: Resilience: Genes to Brain to Behavior (G-B-B) (2022). Maastricht, the Netherlands (Poster and e-pitch) [Best e-pitch award]

Monereo-Sánchez, J., Jansen, J. F., Köhler, S., van Boxtel, M.P.J., Backes, W. H., Stehouwer, C. D., Kroon, A.A., Kooman, J.P., Schalkwijk, C.G., Linden, D.E.J., Schram, M.T. Hippocampal subfields atrophy patterns in prediabetes and type 2 diabetes: Preliminary results from The Maastricht Study. Annual Dutch Diabetes Research Meeting (ADDRM) (2022). Wageningen, the Netherlands (Oral presentation)

Monereo-Sánchez, J., Jansen, J. F., Köhler, S., van Boxtel, M.P.J., Backes, W. H., Stehouwer, C. D., Kroon, A.A., Kooman, J.P., Schalkwijk, C.G., Linden, D.E.J., Schram, M.T. Prediabetes and type 2 diabetes associations with hippocampal subfields atrophy: The Maastricht Study. Organization for Human Brain Mapping (OHBM) Annual Meeting (2022). Glasgow, United Kingdom (Poster)

Monereo-Sánchez, J., Jansen, J. F., Köhler, S., van Boxtel, M.P.J., Backes, W. H., Stehouwer, C. D., Kroon, A.A., Kooman, J.P., Schalkwijk, C.G., Linden, D.E.J., Schram, M.T. Hippocampal subfields atrophy patterns in prediabetes and type 2 diabetes: The Maastricht Study. Research Accelerating Psychosocial Innovations in Diabetes (RAPID) Conference (2022). Copenhagen, Denmark. (Oral presentation)