

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Fung, Yi Leng

Title:

Volumetric Measures of the Medial Temporal Structures and their Neuropsychological Correlates in Alzheimer's Disease

Date:

2020

Persistent Link:

<https://hdl.handle.net/11343/242431>

Terms and Conditions:

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works

**Volumetric Measures of the Medial Temporal Structures and their Neuropsychological
Correlates in Alzheimer's Disease**

Yi Leng Fung

ORCID: 0000-0002-8914-3969

Submitted in total fulfilment of the requirements for the degree of
Doctor of Philosophy

March 2020

Melbourne School of Psychological Sciences
Faculty of Medicine, Dentistry and Health Sciences
The University of Melbourne

Abstract

Diagnosing Alzheimer’s disease (AD) in its earliest phases is difficult. A combination of standardised cognitive testing with AD-related biomarkers such as medial temporal atrophy observed on neuroimaging scans has been proposed to improve the diagnostic process of AD detection. The availability of automated volumetric segmentation software presents as a promising avenue for the incorporation of quantitative volumetric measures into routine clinical workup. However, the performance of these automated protocols in a clinical context remains to be clarified. This thesis aimed to examine the use automated segmentation in a “real-world” clinical setting against the gold standard of manual segmentation, assess the difficulties surrounding reliable automated measurements and evaluate the relationship between volume, diagnosis and memory function.

Methods utilised to address the research questions included a review of the diagnostic accuracy of various diagnostic tools, in-depth examination of structural variations that could undermine automated segmentation efforts, a series of validation studies and assessment of medial temporal volumes and memory function in samples of clinical and healthy controls.

The results indicate that a multimodal approach to diagnosis that includes combination of cognitive assessment and other biomarkers would likely yield more accurate diagnosis.

Although automated segmentation of the medial temporal structures was promising for the hippocampus, the entorhinal and transentorhinal cortices were inherently more challenging to delineate and produced much poorer reliability measures. Medial temporal structural volumes contributed little beyond what was already provided by comprehensive memory assessment for distinguishing between the clinical and healthy control samples. Taken altogether, these findings highlight the need for additional refinement of automated algorithms and further investigation into the ideal combination of cognitive testing and other biomarkers.

Declarations

This is to certify that this thesis is

- (1) comprised of my original work towards the Doctor of Philosophy, except where indicated in the preface,
- (2) due acknowledgements have been made in the text to all materials used, and
- (3) the thesis is fewer than 100,000 words in length, exclusive of tables, references and appendices.

Yi Leng Fung

16th March 2020

Preface

In accordance to the regulations detailed in the Graduate Research Training Policy (MPF1321) section 4.66, the following details the candidate's contribution to the published material included in Chapter 3 of the thesis, in lieu of an individual chapter.

Chapter 3 consist of the author accepted manuscript of an article in the *Journal of Alzheimer's Disease* that was published in 2019. The candidate, Fung, Y. L. is the primary author and contributed to 60% of this manuscript in terms of conceptualisation, data processing and analysis, manuscript preparation, manuscript submission and manuscript revisions in response to co-authors' and editors' review. Ng contributed towards manuscript preparation and data generation, and Ngo aided with data processing. Bowden, Collins, Vogrin and Meade were involved in all steps of the manuscript preparation. The final publication of this study is available at IOS press through <http://dx.doi.org/10.3233/JAD-181172>.

A letter to the editor published in March 2019 on the Journal of Alzheimer's Disease website is also included as an appendix to the main manuscript in Chapter 3. The candidate, Fung Y. L., was responsible for 80% of the preparation and submission of this letter to the editor. Bowden, Collins, Vogrin and Meade were also involved in the preparation and revision of this letter. The original online publication of this letter is available on the Journal of Alzheimer's Disease website, <https://www.j-alz.com/content/labelling-statistical-output-cytoarchitecture>. Full references for both the final published manuscript and letter to the editor are found in the section "Publications and Presentations".

Funding Acknowledgments

This study was partially supported by the Brain Foundation, awarded in October 2014, towards the recruitment of control participants. Data obtained from these control participants were used in the analyses found in Chapters 4, 5 and 6. The Brain Foundation was not involved in the design, analysis or writing of these studies.

The candidate, Fung Y. L., was supported by the Melbourne International Research Scholarship (2014) and the Melbourne International Fee Remission Scholarship (2014).

Publications and Presentations

Published Manuscript and Letter to the Editor

Fung, Y. L., Ng, K. E. T., Vogrin, S. J., Meade, C., Ngo, M., Collins, S. J., & Bowden, S. C. (2019). Comparative utility of manual versus automated segmentation of hippocampus and entorhinal cortex volumes in a memory clinic sample. *Journal of Alzheimer's Disease*, 68(1), 159-171. doi:10.3233/JAD-181172

Fung, Y. L., Ng, K. E. T., Vogrin, S. J., Meade, C., Ngo, M., Collins, S. J., & Bowden, S. C. (2019, March 14). Labelling of statistical output of cytoarchitecture-defined variables in FreeSurfer version 6. Retrieved from <https://www.j-alz.com/content/labelling-statistical-output-cytoarchitecture>

Published Conference Abstracts

Fung, Y. L., Ng, K. E., Vogrin, S., Meade, C., Collins, S., & Bowden, S. (2015, July).

Validation of automated hippocampal volumetric measures in a clinical population.

Poster session presented at the 5th INS/ASSBI Pacific Rim Conference, Sydney, NSW, Australia.

Fung, Y. L., Ng, K. E., Vogrin, S., Meade, C., Collins, S., & Bowden, S. (2015, November).

Validating the clinical utility of automated hippocampal volumetry. Poster session presented at the 21st Annual Conference APS College of Clinical Neuropsychologists, Shoal Bay, NSW, Australia.

Fung, Y.L., Ng, K.E., Vogrin, S., Meade, C., Fripp, J., Bourgeat, J., Collins, S., & Bowden,

S. (2016, August). *Validating automated hippocampal segmentation in a clinical population.* Poster session presented at the Aikenhead Centre for Medical Discovery Research Week, Fitzroy, VIC, Australia.

- Fung, Y.L., Vogrin, S., Meade, C., Fripp, J., Bourgeat, J., Collins, S., & Bowden, S. (2017, August). *Examining the reliability of automated entorhinal and transentorhinal segmentation*. Poster session presented at the Aikenhead Centre for Medical Discovery Research Week, Fitzroy, VIC, Australia.
- Fung, Y.L., Vogrin, S., Meade, C., Collins, S., & Bowden, S. (2017, August). *Evaluating the association of memory performance with manual and automated volume estimates of mesial temporal lobe structures in a clinical setting*. Poster session presented at the Aikenhead Centre for Medical Discovery Research Week, Fitzroy, VIC, Australia.
- Fung, Y.L., Vogrin, S., Meade, C., Bourgeat, J., Fripp, J., Collins, S., & Bowden, S. (2017, August). *Validating the automated methods of hippocampal segmentations on clinical magnetic resonance images*. Poster session presented at the Aikenhead Centre for Medical Discovery Research Week, Fitzroy, VIC, Australia.
- Fung, Y.L., Vogrin, S., Meade, C., Fripp, J., Bourgeat, J., Collins, S., & Bowden, S. (2017, October). *Examining the validity of automated estimation of entorhinal and perirhinal cortices*. Poster session presented at the 37th Annual Conference of the National Academy of Neuropsychology, Boston, MA.
- Fung, Y.L., Ng, K.E., Vogrin, S., Meade, C., Fripp, J., Bourgeat, P., Collins, S., & Bowden, S. (2017, October). *Validating the utility of automated hippocampal measurement tools in clinical scans*. Poster session presented at the 37th Annual Conference of the National Academy of Neuropsychology, Boston, MA.
- Fung, Y.L., Vogrin, S., Meade, C., Collins, S., & Bowden, S. (2017, November). *Examining the relationship between mesial temporal structures volume estimates and memory function in the clinical population*. Poster presented at the 22nd Annual Conference of the APS College of Clinical Neuropsychologists, Perth, WA, Australia.

Fung, Y.L., Vogrin, S., Meade, C., Fripp, J., Bourgeat, J., Collins, S., & Bowden, S. (2017, November). *Validating automated measures of entorhinal and transentorhinal cortices*. Poster presented at the 22nd Annual Conference of the APS College of Clinical Neuropsychologists, Perth, WA, Australia.

Fung, Y.L., Ng, K.E., Vogrin, S., Meade, C., Fripp, J., Bourgeat, P., Collins, S., & Bowden, S. (2017, November). *Validating automated methods of hippocampal segmentations on clinical magnetic resonance images*. Poster presented at the 22nd Annual Conference of the APS College of Clinical Neuropsychologists, Perth, WA, Australia.

Conference Presentation

College of Clinical Neuropsychology Postgraduate Symposium (2018, October). Symposium presentation titled *Automated versus manual measurements of mesial temporal structures and its neuropsychological correlates*.

Acknowledgments

I am indebted to my incredible supervisory and advisory team. Without them, the completion of this thesis would have been impossible. First and foremost, I would like to express my gratitude to my primary supervisor, Professor Stephen Bowden. With unwavering support and boundless patience, he had gently nudged me in the direction of growth and critical thinking, while giving me the freedom to pursue my own ideas.

To my dream team of co-supervisors who kept me going: I would like to thank Professor Steven Collins, who was instrumental in the development of this thesis from the start to the end and always ready to lend me his clinical expertise at every step of the way. Words fail to express my gratitude for Mr Simon Vogrin, who was extraordinarily generous with his time to share both his technical expertise and infectious passion for pursuing new research ideas. I am also thankful for Dr Catherine Meade, who always kept her door open for discussing clinical opinions and suggestions. I am also grateful to the Chair of my advisory committee, Professor Rob Hester, who with his pragmatic approach and constant encouragement, guided and supported me along the way to completion.

I would also like to extend my gratitude to Sara Vogrin for readily offering her time to clarify my numerous statistical inquiries; to the CSIRO Health and Biosecurity team for providing the AIBL study data; and to Shin Yi Quek and Hana McDonald for contributing considerable time and effort to the data collection. I am also thankful for all the wonderful people who had volunteered their time to participate in the research. Their support reminds me of the immeasurable value of good research, and they pave the way for future generations who will benefit from their involvement.

Finally, I am exceedingly blessed with the support of all my family, but especially my mother. Without her unconditional love and support throughout my life, I would not have been able to begin this amazing journey. Thank you.

Contents

Chapter 1

Review of the Current Diagnostic Tools for Detecting Mild Cognitive Impairment and Alzheimer's Disease	1
Methods.....	12
Results.....	14
Discussion	28

Chapter 2

Morphology of the Collateral Sulcus.....	55
Method	65
Results.....	69
Discussion	76

Chapter 3

Comparative Utility of Manual versus Automated Segmentation of Hippocampus and Entorhinal Cortex Volumes in a Memory Clinic Sample	93
Materials and Methods.....	98
Results.....	104
Discussion	108

Chapter 4

Manual Validation of Entorhinal and Transentorhinal <i>Ex Vivo</i> Protocol in FreeSurfer	126
Method	135
Results.....	145
Discussion	159

Chapter 5

Manual Validation of FreeSurfer Hippocampal Subfields Protocol in a Clinical Population	177
Method	185
Results.....	192

Discussion	204
Chapter 6	
Medial Temporal Volumetric Correlates to Memory Performance and Diagnosis of Mild Cognitive Impairment and Alzheimer’s Disease	223
Method	239
Results.....	248
Discussion	267
Chapter 7	
General Discussion.....	309
Structural Variations of the Collateral Sulcus	311
Utility of FreeSurfer in a Clinical Setting.....	312
Contribution of Volumetric Measures to Diagnosing Alzheimer’s Disease	314
Methodological Considerations	317
Overall Conclusion	320

List of Figures

Figure 2.1	
<i>Graphical Representation of the Sulcal Patterns at the Temporal Lobe Surface</i>	<i>67</i>
Figure 2.2	
<i>Types of Connections Between the Anterior and Posterior Collateral Sulci.....</i>	<i>71</i>
Figure 2.3	
<i>Different Branching Patterns and Connections of the Anterior Collateral Sulcus</i>	<i>72</i>
Figure 2.4	
<i>Types of Interruptions to the Entorhinal and Transentorhinal Cortices</i>	<i>75</i>
Figure 3.1	
<i>Key Landmarks of Hippocampal and Entorhinal Manual Protocols</i>	<i>102</i>
Figure 3.2	
<i>Bland-Altman Plots of Hippocampal and Entorhinal Volumes by Two Independent Raters.</i>	
<i>.....</i>	<i>103</i>
Figure 3.3	
<i>Bland-Altman plot of Hippocampal Volumetric Difference Between Manual and</i>	
<i>FreeSurfer</i>	<i>109</i>
Figure 3.4	
<i>Bland-Altman Plot of Entorhinal Percentage Difference Between Manual and FreeSurfer</i>	<i>111</i>
Figure 4.1	
<i>Example of the Boundaries of Entorhinal and Transentorhinal Cortices Based on Manual</i>	
<i>and Automated Methods</i>	<i>140</i>
Figure 4.2	
<i>Example of the Automated Segmentation Boundaries Using Probabilistic Labels</i>	<i>142</i>

Figure 4.3	
<i>Estimated Volumes of the Entorhinal Cortex Based on Different Methods of Segmentations</i>	<i>147</i>
Figure 4.4	
<i>The Estimated Volumes of the Transentorhinal Cortex Based on Manual and Automated Segmentation.....</i>	<i>149</i>
Figure 4.5	
<i>Bland-Altman Plots for Bias Estimation Between Manual and Automated Methods of Entorhinal Segmentation in the 1.5T Control and Clinical Samples</i>	<i>155</i>
Figure 4.6	
<i>Bland-Altman Plots for Bias Estimation Between Manual and Automated Methods of Entorhinal Segmentation in the 3T Control and Clinical Samples.....</i>	<i>156</i>
Figure 4.7	
<i>Bland-Altman Plots for Bias Estimation Between Manual and Automated Methods of Transentorhinal Segmentation.....</i>	<i>158</i>
Figure 5.1	
<i>Illustration of the Manual and FreeSurfer Automated Segmentation of the Hippocampus ..</i>	<i>190</i>
Figure 5.2	
<i>Bland-Altman Plot of Agreement Between Hippocampal Volumes from Different Scanners in the Control Sample</i>	<i>196</i>
Figure 5.3	
<i>Bland-Altman Plot of Agreement Between Manual and Automated Hippocampal Volumes in the Control Sample</i>	<i>199</i>

Figure 5.4

<i>Bland-Altman Plots of Agreement Between Manual and Automated Hippocampal Volumes in the Clinical Sample</i>	<i>201</i>

Figure 5.5

<i>Bland-Altman Plots of Agreement Between Manual and Automated Hippocampal Volumes in All Samples.....</i>	<i>203</i>

Figure 6.1

<i>Scatterplot and Tukey Post-hoc Comparisons of Medial Temporal Structural Volumes</i>	<i>252</i>

Figure 6.2

<i>Volumetric Differences Between Left and Right Medial Temporal Lobe Volumes</i>	<i>256</i>

Figure 6.3

<i>Boxplots of Lateralised Differences Between Left and Right Manual Volume Estimations..</i>	<i>257</i>

Figure 6.4

<i>Boxplots of Asymmetry Between Left and Right Manual Volume Estimations.....</i>	<i>258</i>

Figure 6.5

<i>Boxplots of Lateralised Differences Between Left and Right FreeSurfer Volume Estimations.....</i>	<i>260</i>

Figure 6.6

<i>Boxplots of Asymmetry Between Left and Right FreeSurfer Volume Estimations.....</i>	<i>261</i>

List of Tables

Table 1.1	
<i>List of Systematic Reviews Included in this Study</i>	<i>16</i>
Table 1.2	
<i>Diagnostic Accuracies of Cognitive Assessment Tools for Alzheimer’s Disease Dementia or All-Cause Dementia</i>	<i>18</i>
Table 1.3	
<i>Diagnostic Accuracies of Cognitive Assessment Instruments and Measures for Cognitive Mild Impairment.....</i>	<i>21</i>
Table 1.4	
<i>Diagnostic Accuracies of CSF, PET and MRI Biomarkers for Detecting AD and MCI</i>	<i>25</i>
Table 2.1	
<i>Types of Connection between Anterior and Posterior Collateral Sulcus, and Presence of Side-branches Along Anterior Collateral Sulcus.....</i>	<i>70</i>
Table 2.2	
<i>Patterns of Interruption to the Entorhinal and Transentorhinal Cortices and Its Relationship to the Collateral Sulcus.....</i>	<i>74</i>
Table 3.1	
<i>Absolute and Percentage Difference of Mean Volumes Between Protocols</i>	<i>107</i>
Table 4.1	
<i>Summary of Demographics Within in Each Sample</i>	<i>138</i>
Table 4.2	
<i>Means and Standard Deviations of the Entorhinal Volumes and Comparisons of these Volumes Across the Different Methods of Volume Estimation</i>	<i>146</i>

Table 4.3	
<i>Means and Standard Deviations of the Transentorhinal Volumes and Comparisons of These Volumes Between the Different Methods of Volume Estimation</i>	<i>148</i>
Table 4.4	
<i>Comparisons of Left and Right Volume Estimates Using Paired-Sample t-test.....</i>	<i>151</i>
Table 4.5	
<i>Comparisons of 1.5T and 3T Manual and Automated Volume Estimates in the Control Sample Using Paired-Sample t-test and Absolute Intraclass Correlation Coefficient (ICC)</i>	<i>152</i>
Table 4.6	
<i>Intraclass Correlation Coefficient (ICC) Between Manual and Automated Estimates of the Entorhinal and Transentorhinal Cortices.....</i>	<i>154</i>
Table 5.1	
<i>Intra- and Inter-rater Reliability Assessed by Intra-class Correlation Coefficient (ICC) Between Manual Hippocampal Volume Estimations.....</i>	<i>191</i>
Table 5.2	
<i>Comparisons of Manual and Automated Hippocampal Volume Estimations (mm³) in the Control and Clinical Samples</i>	<i>193</i>
Table 5.3	
<i>Comparisons of Manual and Automated Hippocampal Volume Estimations Obtained on Different Scanner Magnetic Field Strengths in the Control Sample</i>	<i>195</i>
Table 5.4	
<i>Intra-class Correlation Coefficient Between Manual and Automated Hippocampal Volume Estimations.....</i>	<i>198</i>

Table 6.1	
<i>Demographics and Neuropsychological Characteristics of the Clinical and Control</i>	
<i>Samples</i>	<i>242</i>
Table 6.2	
<i>Correlations Between the Different Structural Volumes Assessed by Manual Delineation and</i>	
<i>FreeSurfer</i>	<i>251</i>
Table 6.3	
<i>Average Raw Volumes of Hippocampus, Entorhinal and Transentorhinal Cortices Between</i>	
<i>the Clinical and Healthy Control Samples</i>	<i>254</i>
Table 6.4	
<i>Stepwise Multiple Regression Analysis Examining the Relationship between Auditory</i>	
<i>Memory and MTL Volumes.....</i>	<i>262</i>
Table 6.5	
<i>Stepwise Multiple Regression Analysis Examining the Relationship between Visual Memory</i>	
<i>and MTL Volumes</i>	<i>263</i>
Table 6.6	
<i>Stepwise Multiple Regression Analysis Examining the Relationship between General Memory</i>	
<i>and MTL Volumes</i>	<i>263</i>
Table 6.7	
<i>Multinomial Logistic Regression Analysis Examining the Relationship Between the General</i>	
<i>Memory, Volumetric Measures and Classification in the Clinical or Healthy Control Sample</i>	
<i>in Final Model.....</i>	<i>266</i>
Table 6.8	
<i>Summary of Coefficient Estimates of the Sum, Difference Volume and Asymmetry Index of the</i>	

<i>Left and Right Hippocampal Volumes for Prediction of Classification in AD, MCI, SMC or</i>	
<i>HC Sample</i>	<i>268</i>

List of Appendices

Appendix A

Letter to the Editor (Chapter 3) 124

Appendix B

Additional Statistical Outputs (Chapter 6) 301

Structure of the Thesis

All the chapters in this thesis have been written to be self-contained. Each chapter, except Chapter 7 that is the final general discussion chapter, consist of an abstract, a detailed review of the literature that is relevant to the specific aims of the chapter, the relevant methodological approach, results and discussion and list of relevant references.

Chapter 1 is a review of literature, specifically looking at systematic reviews that addressed the diagnostic test accuracies of various diagnostic tests that are being used to detect Alzheimer's disease (AD) or mild cognitive impairment (MCI). Chapters 2 to 6 are empirical papers that addresses different related research questions. Chapter 2 discusses the collateral sulcus in detail, which is a structure that is highly relevant to volumetric segmentation of the medial temporal structures on structural magnetic resonance imaging scans. Chapters 3 to 5 consist of a series of validation studies on volumetric estimations of the medial temporal structures, specifically comparing automated estimations by FreeSurfer (v.6.0) against the manual segmentation on three structures of interest, namely the hippocampus, the entorhinal cortex and the transentorhinal cortex. Chapter 6 addresses the role of volumetric biomarkers in distinguishing people with AD and MCI from people likely without AD pathology. The chapter also discusses the relationship between these volumes and memory function. Lastly, a general discussion of all findings from the preceding chapters is contained within Chapter 7.

Chapter 1

Review of the Current Diagnostic Tools for Detecting Mild Cognitive Impairment and Alzheimer's Disease

This chapter presents a review of the diagnostic test accuracy of various cognitive assessment tools, and fluid and neuroimaging biomarkers, in the detection of Alzheimer's disease and mild cognitive impairment. The findings from this review highlights the strengths and limitations of the different diagnostic tools, including cognitive screening tools, longer assessment tools of various cognitive domains, and five of the main fluid and neuroimaging biomarkers. The findings from this review also highlights how the addition of the use of fluid and neuroimaging biomarkers can serve as a complementary tool above objective cognitive assessment, to improve diagnostic accuracy in Alzheimer's disease and mild cognitive impairment.

Abstract

Alzheimer's disease (AD) is one of the most common forms of neurodegenerative disorder in the older population. In addition to the presence of objective cognitive impairment, fluid and neuroimaging biomarkers have been included as supportive evidence for a clinical diagnosis of AD or mild cognitive impairment (MCI). To evaluate the diagnostic test accuracy of current diagnostic tools, a specific search strategy was employed to search for relevant systematic reviews on the diagnostic test accuracy of the available cognitive assessment tools and five main fluid and neuroimaging biomarkers, namely CSF A β ₄₂, CSF tau, amyloid PET, FDG-PET and MRI, in the detection of AD and MCI. Systematic reviews reported adequate diagnostic accuracy of current cognitive tools for AD detection, but cognitive screening tools such as the MMSE had very poor diagnostic accuracy statistics for MCI detection. Reported sensitivity of memory measures for MCI was variable and tended to be lower than .80, but specificity of memory measures was generally greater than .80. Conversely, fluid and neuroimaging biomarkers tended to have high sensitivity (>.80) but poor specificity (<.80) in both AD and MCI groups. Given the limitations of the diagnostic accuracy of each diagnostic tool, particularly for the detection of MCI, a multimodal system would be more likely to improve the ability to detect AD-related cognitive decline in its earliest stages.

Keywords: Alzheimer's disease, mild cognitive impairment, cognitive assessment, memory, biomarkers, diagnostic test accuracy, review, systematic review

Review of the Current Diagnostic Tools for Detecting Mild Cognitive Impairment and Alzheimer's Disease

Dementia is a clinical syndrome that comprises deficits in different areas of cognition, including areas such as memory, language and problem-solving skills, which could impact a person's ability to carry out activities of daily living (Alzheimer's Association, 2018).

Alzheimer's disease (AD) is one of the most common form of dementia in the older adult population and features memory impairments as one of its earliest symptoms (Alzheimer's Association, 2018).

The clinical criteria for probable AD established by the National Institute of Neurological and Communicative Disorder and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) was updated in 2011 from the 1984 criteria, but the clinical criteria remain largely similar (Jack, Albert, et al., 2011; McKhann et al., 1984; McKhann et al., 2011). In brief, the 2011 clinical criteria for AD includes (1) interference with daily work or usual function, (2) an insidious onset and decline from previous levels of functioning, (3) cognitive decline is not explained by major psychiatric disorders, delirium or other neurological or non-neurological medical conditions, (4) cognitive impairment is established through history taking from informant and patient, and an objective cognitive assessment that could be either a cognitive screen or more comprehensive neuropsychological testing, and (5) cognitive impairment in two or more domains that could be either an amnestic or non-amnestic presentation (McKhann et al., 2011).

Advances in the understanding of the pathology and course of AD has led to the conceptual shift of understanding the disease as a continuum, wherein AD may have a long asymptomatic phase in which neuropathological changes occur prior to any detectable cognitive decline and before the onset of functional impairment (Dubois, Hampel, et al., 2016; Sperling et al., 2011). The transition from asymptomatic to symptomatic pre-dementia

phase, and from the symptomatic pre-dementia phase to the onset of dementia is difficult to identify (McKhann et al., 2011; Sperling et al., 2011).

In order to capture this continuum of impairment in AD, the National Institute on Ageing and the Alzheimer's Association (NIA-AA) established a working group to outline the core clinical criteria for identifying the individuals with symptomatic pre-dementia AD, which is referred to as mild cognitive impairment due to AD (MCI; Albert et al., 2011). The recommendations for a clinical diagnosis of MCI highlighted that although impairment in episodic memory is most commonly seen in this population, other cognitive domains in addition to memory function should be assessed. However, the main distinguishing clinical criteria for people with MCI from people with dementia is the preservation of independence in daily life (Albert et al., 2011; McKhann et al., 2011).

The current clinical criteria for dementia and pre-dementia phases of AD both describe a syndromic presentation of the disease based on clinical judgment and cognitive functioning without neuropathological verification. However, the development of *in vivo* AD pathologic biomarkers has spurred an evolution in the thinking about the definition of AD, away from the traditional syndromal classification of AD towards a biological definition of AD, in which cognitive features are simply an expression of disease neuropathological evolution with biomarkers providing status (Jack et al., 2018). Biomarker characterisation of AD is crucial to defining the AD continuum, notably for the earlier pre-symptomatic and pre-dementia stages, but the relationship between progression of neuropathology and evincing of cognitive impairment is not clear-cut (Jack et al., 2018). Many uncertainties, such as the appropriate cut-offs to distinguish between normal and abnormal biomarkers, rates of cognitive decline in relationship to the various biomarkers, protective factors against and risk factors for cognitive decline, all continues to be investigated for further clarification (Brookmeyer, Abdalla, Kawas, & Corrada, 2018; Ebenau et al., 2020; Jack et al., 2018; Jack,

Wiste et al., 2019; Vemuri et al., 2017). As such, high quality cognitive assessment will continue to constitute an important component of the evaluation of dementia and pre-dementia symptomatic stages of AD, a stance that is highlighted in the current clinical diagnostic criteria (Albert et al., 2011; McKhann et al., 2011).

Timely detection of AD, particularly the pre-dementia MCI population, is important as it means that patients have earlier opportunities to make healthy lifestyle changes, advanced care planning or end-of-life decisions (Dubois, Padovani, Scheltens, Rossi, & Dell'Agnello, 2016). Another important benefit of early detection is allowing people diagnosed with MCI to undergo further investigation to determine whether AD may be the basis for their cognitive symptoms and seek other medical care if other causes are more likely (Alzheimer's Association, 2018). Furthermore, although there are no therapeutic treatments that have been shown to halt the progressive nature of AD at the moment, when pharmaceutical interventions do become available in the future, the treatments are potentially going to be most effective when applied during the earliest stages of AD (Sperling et al., 2011).

Cognitive Measures

As noted in the clinical criteria for both probable AD and MCI due to AD detailed above, it is important to determine objective evidence of cognitive decline. One of the proposed methods of cognitive assessment is the use of brief “bedside” cognitive screening tools (McKhann et al., 2011). Brief cognitive screening tools are cheap, fast and non-invasive instruments that can be used to identify people with symptomatic AD.

Traditionally, the Mini Mental state Examination (MMSE) was recommended as the primary choice for a brief cognitive screening and was one of the most commonly used screening instruments in clinical practice (Folstein, Folstein, & McHugh, 1975; Shulman et al., 2006). An early meta-analysis examined the accuracy of MMSE in detecting MCI and

AD in both clinical and non-clinical settings (Mitchell, 2009). The MMSE was observed to be modestly effective at ruling out dementia in both specialist and community settings. However, the authors found a paucity of research on the accuracy of MMSE in the MCI group and noted that based on the five studies selected, the ability of MMSE to detect MCI was limited in a clinical setting (Mitchell, 2009). Based on different cut-off points in the studies, the reported sensitivities ranged from .18 – .86 and specificities ranged from .48 – 1.00 (Mitchell, 2009).

The copyright restrictions with the MMSE that begun in 2000 has limited the use of this instrument in clinical practice (Feldman & Newman, 2013). However, there is now a wide variety of cognitive screening instruments available, such as a similarly brief cognitive screen (administration time $\leq$ 10 mins) like the Montreal Cognitive Assessment (MoCA: Nasreddine et al., 2005), or longer cognitive screening battery (administration time $\leq$ 20 minutes) such as the Addenbrooke's Cognitive Examination-Revised (ACE-R; Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006).

The clinical recommendations for diagnosing MCI includes cognitive testing that is used to identify individuals performing 1 to 1.5 standard deviations below the mean based on appropriate normative data (Albert et al., 2011). Two main clinical phenotypes of MCI exist, namely amnestic MCI and non-amnestic MCI, and either of these clinical phenotypes can present with impairment in a single cognitive domain or in multiple cognitive domains (Petersen et al., 2009). However, the importance of the distinction between an amnestic and non-amnestic presentation in MCI, regardless of whether the impairment occurs in a single or several domains, is the differentiation between underlying disease aetiologies and consequent disease progression (Petersen et al., 2009). A non-amnestic presentation has largely been associated with a greater likelihood of progression into other non-AD dementias (Ferman et al., 2013; Sarazin et al., 2007). On the other hand, an amnestic presentation has generally

been associated with a greater likelihood of progression into AD (Bradfield et al., 2018; Sarazin et al., 2007). These findings point towards memory dysfunction as being one of the most discriminant symptom for the identification of MCI that is due to an underlying AD aetiology.

Even though there is no reference to the various subtypes of MCI in the most recent core clinical diagnostic criteria, the importance of memory testing continues to be emphasised as memory dysfunction is the most common early symptom in MCI (Albert et al., 2011). The working group for the clinical criteria for MCI highlighted the need for memory tests that examined both immediate and delayed recall, such as word-list learning tasks like the Rey Auditory Verbal Learning Test, California Verbal Learning Test, or other memory measures found in the Wechsler Memory Scale, such as Logical Memory and Visual Reproduction (Albert et al., 2011). Although there is no explicit mention of the efficacy of cognitive screening tools in detecting MCI, Albert et al. (2011) emphasised that physicians need to be cognisant of tests that might not be sensitive to subtle cognitive dysfunction that might be present in the early stages of MCI.

Fluid and Neuroimaging Biomarkers

Fluid and neuroimaging biomarkers are physiological, biochemical and anatomical parameters that can be measured *in vivo* to indicate specific pathophysiological features that are related to the disease (Jack, Albert, et al., 2011). The hallmark pathologies of AD are the progressive build-up of β -amyloid outside the neurons in the brain as fibrillary plaques, and the increase of misfolded tau proteins within the brain neuronal cell bodies. Therefore, an AT(N) system has been proposed to represent the three main physiological features of AD neuropathology, namely the deposition of β -amyloid (A), presence of pathologic tau proteins (T) and evidence of neurodegeneration or neuronal injury (N; Jack et al., 2018; Jack et al., 2016).

The modalities to capture these pathophysiological changes *in vivo* include using cerebrospinal fluid (CSF), positron emission tomography (PET) and magnetic resonance imaging (MRI). In line with AT(N) system, CSF A β_{42} and amyloid PET are measures of β -amyloid, CSF phosphorylated tau (p-tau) is a measure of tau, and CSF total tau (t-tau), fluorodeoxyglucose (FDG) PET and MRI are measures of neurodegeneration or neuronal injury. These five biomarkers, CSF A β_{42} , CSF tau, amyloid PET, FDG-PET and MRI, have been formally integrated into the research diagnostic criteria in both the existing research diagnostic classification systems, namely the International Working Group (IWG: Dubois et al., 2014), and the NIA-AA. Although there is disagreement between these two groups on the interpretation of these biomarkers with regards to classification along the AD spectrum, the value of these biomarkers lies not in the diagnostic classification ability but the use of these biomarkers as a complementary system with existing clinical diagnostic criteria (Jack et al., 2016).

The main modality for measuring fluid biomarkers is via cerebrospinal fluid (CSF), which is thought to serve as a substrate for neuropathological-related biochemical changes as CSF is in direct contact with the cerebrum (Counts, Ikonomic, Mercado, Vega, & Mufson, 2017). Biomarkers for AD within the CSF have been related primarily to the proteolytic fragment of the amyloid precursor protein (β -amyloid), total tau (t-tau) and hyperphosphorylated forms of tau (p-tau) levels (X. Q. Chen & Mobley, 2019; Chi, Chang, & Sang, 2018). In particular, a pattern of reduced A β_{42} and increased t-tau and p-tau levels is closely associated with AD (Sperling & Johnson, 2013). A decrease in A β_{42} protein reflects increased deposition of β -amyloid in amyloid plaques in the brain, while increased levels of t-tau and p-tau points to the increased synthesis and release of these abnormal tau proteins in the brain (Chen & Mobley, 2019). P-tau has been suggested to increase specificity for AD in

the CSF biomarker profile as it is more closely related to the neurofibrillary tangle formation in AD (Sperling & Johnson, 2013).

One of the main targets for molecular neuroimaging of AD is amyloid plaques in the brain. The first radioligand that binds to the insoluble forms of amyloid that was developed for PET imaging is ^{11}C -labelled Pittsburgh Compound B (^{11}C -PIB; Klunk et al. 2004). The examination of patients with post-mortem confirmed AD demonstrated high correlation between the brain regions that retained the PIB with quantitatively analysed amyloid plaques, therefore supporting the validity of using PIB as a method of amyloid plaque burden analysis (Ikonomovic et al., 2008). However, one of the shortcomings of using PIB labelled with ^{11}C is the short half-life of 20 minutes, which limits the use of this radiotracer to centres with expertise in ^{11}C -radiochemistry (Duara et al., 2015). To overcome the limitation that the short half-life of ^{11}C -PIB imposes on the clinical use, several ^{18}F -labelled radioligands with longer half-life has been developed for the imaging of amyloid deposits. The radioligands include the ^{18}F -florbetapir, ^{18}F -florbetaben and the more recently developed ^{18}F -flutemetamol (Clark et al., 2011; Filippi, Chiaravalloti, Bagni, & Schillaci, 2018; Sabri, Seibyl, Rowe, & Barthel, 2015). These ^{18}F -labelled radioligands have been shown to have a high affinity to β -amyloid fibrils, therefore allowing *in vivo* localisation of amyloid plaques within the brain (Filippi et al., 2018).

Fluorodeoxyglucose (FDG) is a radionuclide used in PET that is combined with glucose to provide focal measures of cerebral metabolic activity. Therefore, FDG-PET serves as a topographical marker of reduced glucose metabolism that indicates areas of neuronal loss or synaptic dysfunction, in which regions of hypometabolism is typically observed in the temporoparietal cortex in people with AD (Dubois et al., 2014; McKhann et al., 2011). Although a similar pattern of functional abnormality had been shown through the use of single-photon emission computed tomography (SPECT), the use of SPECT has fallen out of

favour because of the availability and spatial resolution superiority of PET imaging (Sperling & Johnson, 2013).

Structural MRI allows the *in vivo* high-resolution examination of atrophy within the hippocampus or medial temporal cortex, using visual ratings or quantitative measures. There have been several advancements in neuroimage processing techniques, including the development of automated neuroimaging software such as FreeSurfer (<http://surfer.nmr.mgh.harvard.edu>), which allows the quantification of cortical changes in MCI and AD (Dill, Franco, & Pinho, 2015). Even though automated segmentation is not yet established for routine medical diagnosis, several lines of research have demonstrated promising discriminatory ability of automated quantification to discriminate between people with AD and MCI from healthy controls (Dill et al., 2015; Lotjonen et al., 2011). Furthermore, since MRI is widely available and frequently used in clinical setting, should the diagnostic accuracy of structural MRI achieve the threshold of clinical utility, it could be easily incorporated into any updated clinical diagnostic models.

Since the development and publication of the AT(N) system in 2018, remarkable progress has been made in other fluid and neuroimaging biomarkers that complement this original framework (Jack et al., 2018). In terms of additional fluid AD biomarkers, there has been significant additions in blood plasma measures as markers for amyloid, tau and neurodegeneration using $A\beta_{42/40}$, p-tau, and neurofilament light respectively (Janelidze et al., 2020; Karikari et al., 2020; Preische et al., 2019; Schindler et al., 2019; Thijssen et al., 2020). There has also been increasing competence in CSF measurements of neurodegeneration using neurofilament light and neurogranin measures, as well as continual development of more effective radioligands for tau PET neuroimaging (Andersson et al., 2020; Dhiman et al., 2020; Leuzy et al., 2020; Liu et al., 2020). Besides the fluid and neuroimaging modalities, there has also been advances in other technologies for assessing AD hallmark pathologies,

such as emerging techniques in detecting amyloid via retinal imaging (Koronyo-Hamaoui, Doustar, Oviatt, Black, & Koronyo, 2020).

However, one of the main limitations of using fluid and neuroimaging biomarkers in the detection of AD is the lack of consensus on the interpretation. Imaging data can be evaluated through qualitative assessment via visual evaluation of structural atrophy in MRI, hypometabolism in FDG-PET and amyloid uptake in the cerebrum, and through quantitative evaluation via a wide range of non-standardised computer-based methods. However, either method may introduce bias into imaging data interpretation (Sheikh-Bahaei, Sajjadi, Manavaki, & Gillard, 2017). Although there are fewer concerns about biased interpretation in CSF biomarkers, there is a large variation in the cut-off criteria that would also introduce significant heterogeneity across research studies (Molinuevo et al., 2014). In general, the IWG and NIA-AA agree that much work is still required in order to standardise the analysis and interpretation of these biomarkers before these biomarkers can be applied routinely in clinical practice (Dubois, Hampel, et al., 2016; Jack et al., 2018; Jack et al., 2016; Sperling & Johnson, 2013). As alluded to earlier, the diagnostic framework of AD should not be construed as divorcing the clinical and cognitive aspects from the underlying neuropathology. Rather, the former can be viewed as an expression of progression in evolution to levels exceeding symptomatic thresholds, with biomarkers representing an independent source of information about the underlying neuropathological process that can increase the certainty of an underlying AD pathophysiology to explain the clinical syndrome (Dubois, Hampel, et al., 2016; Jack et al., 2018). At this juncture, fluid and neuroimaging biomarker findings are still best considered in a complete clinical context as both amyloid-beta and tau can be found in other neurodegenerative disorders and are therefore not an entirely reliable discriminator of clinical diagnosis on their own (Sperling & Johnson, 2013).

Aims of Current Study

This study aims to review the current systematic reviews that evaluate the diagnostic accuracy of the cognitive measures that have been used for assessing AD and MCI, which include screening measures and longer diagnostic cognitive tests. The diagnostic accuracy of the five main fluid and neuroimaging biomarkers that are included in both the research diagnostic criteria of the IWG and NIA-AA will be reviewed, which includes CSF A β ₄₂, CSF tau, amyloid PET, FDG-PET and MRI. One of the recommended ranges for sensitivity and specificity that has been put forward by the Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the NIA, is a value more than .80 (The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group, 1998). This review will therefore discuss the findings with reference to this proposed guideline for diagnostic test accuracy.

Methods

Sourcing Strategy

Publication of systematic reviews from January 2014 to July 2019 were identified by searching through the electronic databases, EMBASE, MEDLINE, PsycINFO and the Cochrane library via OvidSP. A specific search strategy was applied, firstly using combinations of the terms “systematic review”, “Alzheimer*”, “AD”, “dementia”, “cognitive impairment”, “mild cognitive impairment” and “MCI” to identify the systematic reviews that were relevant to AD and MCI. Next, the terms used to identify diagnostic reviews were “sensitiv*”, “detect*”, “accurac*”, “specific*” and reliab*”, which were search terms that have been shown to have a sensitivity of more than 90.2% and precision of 10.4 for identifying relevant diagnostic test accuracy studies in MEDLINE via OvidSP (Beynon et al., 2013).

Additional searches were conducted for specific subsections. To identify the systematic reviews that addressed cognitive tests, a combination of the search terms, “neuropsycholog*”, “cogniti*”, “screen*”, “test*” and “assess*” were used. To identify the systematic reviews that addressed the different fluid and neuroimaging biomarkers, a combination of the search terms, “biomarker*”, “cerebrospinal fluid”, “positron emission tomography”, “fluorodeoxyglucose”, “magnetic resonance imaging”, “tau”, “amyloid”, “CSF”, “PET”, “FDG”, “MRI” were used. An additional search was later done to include relevant articles between August 2019 to September 2019. Comments, letters and conference abstracts were all excluded. Only articles available in English from peer-reviewed journals were included in this review.

Study Selection

Initial inclusion criteria included systematic reviews that used the following (1) study population - adults with suspected MCI, early stages of late-onset AD or early stages of all-cause dementia, (2) setting - any case-controlled, primary, secondary or tertiary care settings, memory clinics or community-based settings, (3) reference standard - any published clinical criteria such as the third and fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III, DSM-IV; American Psychiatric Association, 2000; American Psychiatric Association, 2013), National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA; McKhann et al., 2011), Petersen’s criteria for MCI (Petersen et al., 1999) or a diagnosis based on autopsy findings, (4) index test - any cognitive diagnostic tool, or biomarkers using CSF measures of tau or amyloid-beta, PET radiotracers that targeted amyloid plaques, FDG-PET or MRI, (5) diagnostic accuracy statistics - pooled estimates of sensitivities and specificities calculated from at least three independent studies were reported.

Exclusion criteria were journal articles that were not systematic reviews, or systematic reviews that addressed diagnostic test accuracy for severe dementia, or for mild cognitive impairment and dementia that were specifically related to other aetiologies, such as those related to but not limited to vascular factors, Lewy bodies, alcohol, or human immunodeficiency virus. Other exclusions include systematic reviews that had reference standards that were not clearly operationalised or used diagnostic reference standards that were based only on a cognitive screen such as the MMSE. Index tests for cognition that were based on self-administered cognitive tests or informant reports were excluded. Biomarkers index tests that were not based on assessing amyloid, tau or neurodegeneration using CSF, PET or MRIs were excluded. The resulting database of potentially relevant studies were managed via EndNote X8.2 (Thomson Reuters, CA, USA).

Results

Search Results

The search for systematic reviews of diagnostic test accuracy of cognitive tests returned a total of 779 studies, of which 149 were conference abstracts and were removed. A total of 226 duplicate journal articles were removed via the deduplicate function in EndNote X8.2 and via manual identification. A total of 406 titles and abstracts were screened for relevance of which 56 journal articles were shortlisted for full-text assessment. Unspecified, all-cause dementia was included, as the most prevalent form of dementia is AD and it was likely that all-cause dementia would capture a large proportion of the AD diagnosis unless otherwise specified. A total of 12 systematic reviews were included in this review and the list of studies included is found in Table 1.1.

The search for systematic reviews of diagnostic test accuracy of fluid and neuroimaging biomarkers identified a total of 134 studies. After deduplication via EndNote

X8.2 and removal of conference abstracts, a total of 70 titles and abstracts were screened for relevance, of which 37 publications were retained for full text assessment. However, several systematic reviews, particularly those studies that investigated the diagnostic accuracy of amyloid PET, were unable to conduct meaningful meta-analysis due to the small number of available studies. In order to gain an overview of the current diagnostic utilities of the proposed biomarkers, the systematic reviews that were without statistically pooled results but provided the sensitivities and specificities of the evaluated studies were also included. A total of 14 systematic reviews were included in this review. The studies included are listed in Table 1.1.

Majority of the systematic reviews included in this review presented data using the recommendation provided in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Liberati et al., 2009) or another system of reporting that is similar to the PRISMA criteria. Majority of the systematic reviews also reported having used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool, or a similar tool, as a guideline to evaluating study quality (Whiting et al., 2011).

Performance of Cognitive Measures in Alzheimer's Disease and All-cause Dementia

A total of 12 individual global cognitive screening instruments that assessed diagnostic accuracy in detecting AD or all-cause dementia from six systematic reviews were identified (Table 1.2). These included the Modified Mini-Mental State (3MS; $k = 9$), 7-Minute Screen (7MS; $k = 4$), Addenbrooke's Cognitive Examination (ACE; $k = 5$), ACE-R ($k = 21$), Abbreviated Mental Test (AMT; $k = 14$), DemTect ($k = 3$), General Practitioner Assessment of Cognition (GPCOG; $k = 5$), Mini-Cog ($k = 9$), Memory Impairment Screen (MIS; $k = 6$), MMSE ($k = 157$), MoCA ($k = 35$), and Rowland Universal Dementia Assessment Scale (RUDAS; $k = 8$). Of these, only ACE-R, MMSE and MoCA were studied more than once. Reported point estimates of sensitivity and specificity ranged from .92 - .96

and .83 - .89 respectively for the ACE-R, .81 – .97 and .70 - .90 respectively for the MMSE, and .91 - .93 and .81 - .91 respectively for the MoCA. For the global cognitive screening tools that were only reported in a single systematic review, ACE had the highest point estimate sensitivity of .97 and correspondingly the lowest specificity of .77 (Table 1.2).

Table 1.1

List of Systematic Reviews Included in this Study

Study Source Listed in Alphabetical Order	
Systematic Reviews on Cognitive Measures	
1	Arevalo-Rodriguez, Smailagic, Roquei Figuls, Ciapponi, Sanchez-Perez, . . . Cullum, 2015
2	Belleville, Fouquet, Hudon, Zomahoun, & Croteau, 2017
3	Breton, Casey, & Arnaoutoglou, 2019
4	Creavin, Wisniewski, Noel-Storr, Trevelyan, Hampton, Rayment, . . . Cullum, 2016
5	De Roeck, De Deyn, Dierckx, & Engelborghs, 2019
6	Larner & Mitchell, 2014
7	Naqvi, Haider, Tomlinson, & Alibhai, 2015
8	Paddick, Gray, McGuire, Richardson, Dotchin, & Walker, 2017
9	Park, Jeong, & Seomun, 2018
10	Tsoi, Chan, Hirai, Wong, Mok, Lam, . . . Wong, 2017
11	Tsoi, Chan, Hirai, Wong, & Kwok, 2015
12	Weissberger, Strong, Stefanidis, Summers, Bondi, & Stricker, 2017
Systematic Reviews on Fluid and Neuroimaging Biomarkers	
1	Cure, Abrams, Belger, Dell'agnello, & Happich, 2014
2	Ferreira, Perestelo-Perez, Westman, Wahlund, Sarria, & Serrano-Aguilar, 2014
3	Martinez, Vernooij, Fuentes Padilla, Zamora, Bonfill Cosp, & Flicker, 2017
4	Martinez, Vernooij, Fuentes Padilla, Zamora, Flicker, & Bonfill Cosp, 2017a
5	Martinez, Vernooij, Fuentes Padilla, Zamora, Flicker, & Bonfill Cosp, 2017b
6	Mo, Lim, Sul, Lee, Youn, & Kim, 2015
7	Morris, Chalkidou, Hammers, Peacock, Summers, & Keevil, 2016
8	Rice & Bisdas, 2017
9	Ritchie, Smailagic, Noel-Storr, Takwoingi, Flicker, Mason, & McShane, 2014
10	Ritchie, Smailagic, Noel-Storr, Ukoumunne, Ladds, & Martin, 2017
11	Rivero-Santana, Ferreira, Perestelo-Perez, Westman, Wahlund, Sarria, & Serrano-Aguilar, 2017
12	Smailagic, Lafortune, Kelly, Hyde, & Brayne, 2018
13	Smailagic, Vacante, Hyde, Martin, Ukoumunne, & Sachpekidis, 2015
14	Zhang, Smailagic, Noel-Storr, Takwoingi, McShane, & Feng, 2014

Also shown in Table 1.2, three systematic reviews assessed the diagnostic accuracy of other cognitive domain-specific tests. Only one systematic review evaluated the diagnostic accuracy of longer memory measures in the AD population (Weissberger et al., 2017). As there was a large variety of memory measures, the memory measures were organised into different categories representing the types of measure used, such as immediate and delayed memory, word list, story or visual recall. All reported sensitivities and specificities were above .80 for the memory measures, with sensitivity point estimates ranging from .84 - .92 and specificity point estimates ranging from .81 - .93 (Table 1.2). The only exception was immediate story free recall ($k = 3$) that had a pooled sensitivity point estimate of .71 and pooled specificity point estimate of .75. One study reported pooled point estimates of sensitivity and specificity of verbal fluency test ($k = 7$) to be .80 and .82 respectively. Pooled sensitivity point estimates of the clock drawing test ($k = 38$) that was scored based on several scoring system was highly variable, ranging from .57 - .83, while reported point estimates of specificity ranged from .76 - .88 (Table 1.2).

Performance of Cognitive Measures in Mild Cognitive Impairment

A total of five studies reported the sensitivity and specificity point estimates of individual global screening tests for detecting mild cognitive impairment (Table 1.3). Eight cognitive screenings tests were evaluated. These cognitive screening tests were ACE-R ($k = 13$), CAMCOG ($k = 4$), Consortium to Establish a Registry for Alzheimer's Disease neuropsychological battery (CERAD; $k = 5$), DemTect ($k = 3$), MMSE ($k = 137$), MoCA ($k = 95$), Quick Mild Cognitive Impairment screen (Qmci; $k = 8$), and Memory Alteration Test (M@T; $k = 5$).

As shown in Table 1.3, the MMSE was the most well-studied cognitive screen in this diagnostic group and diagnostic test accuracy was reported in four different systematic reviews. In the groups ($k = 125$) that investigated the diagnostic accuracy of MMSE in

Table 1.2*Diagnostic Accuracies of Cognitive Assessment Tools for Alzheimer's Disease Dementia or All-Cause Dementia*

Source	Instrument/Measure	Target condition	Number of groups	Sensitivity (95% CI)	Specificity (95% CI)
Global Screen					
Tsoi et al. (2015)	3MS	dementia	9	.86 (.83 - .89)	.85 (.74 - .92)
De Roeck, De Deyn, Dierckx, and Engelborghs (2019)	7MS	AD	4	.94 (n.r.)	.93 (n.r.)
Larner and Mitchell (2014)	ACE	dementia	5	.97 (.93 - .99)	.77 (.58 - .92)
	ACE-R	dementia	4	.96 (.92 - .98)	.88 (.64 - .99)
De Roeck et al. (2019)	ACE-R	AD	4	.96 (n.r.)	.83 (n.r.)
Tsoi et al. (2015)	ACE-R	dementia	13	.92 (.90 - .94)	.89 (.84 - .93)
	AMT	dementia	14	.88 (.82 - .92)	.85 (.81 - .89)
De Roeck et al. (2019)	DemTect	AD	3	.96 (n.r.)	.86 (n.r.)
Tsoi et al. (2015)	GPCOG	dementia	5	.92 (.81 - .97)	.87 (.83 - .90)
	Mini-Cog	dementia	9	.91 (.80 - .96)	.86 (.74 - .93)
	MIS	dementia	6	.80 (.68 - .86)	.91 (.84 - .96)
Creavin et al. (2016)	MMSE (cut-off < 24)	dementia	15	.85 (.74 - .92)	.90 (.82 - .95)
	MMSE (cut-off < 25)	dementia	10	.87 (.78 - .93)	.82 (.65 - .92)
	MMSE (cut-off education adjusted)	dementia	7	.97 (.83 - 1.00)	.70 (.50 - .85)
Larner and Mitchell (2014)	MMSE	dementia	9	.92 (.85 - .97)	.87 (.81 - .92)
Paddick et al. (2017)	MMSE (in low literacy population)	dementia	8	.83 (.79 - .86)	.82 (.72 - .89)
Tsoi et al. (2015)	MMSE	dementia	108	.81 (.78 - .84)	.89 (.87 - .91)
	MoCA	dementia	20	.91 (.84 - .95)	.81 (.71 - .88)
De Roeck et al. (2019)	MoCA	AD	15	.93 (n.r.)	.91 (n.r.)
Naqvi, Haider, Tomlinson, and Alibhai (2015)	RUDAS	MCI/dementia	8	.77 (.67 - .85)	.86 (.75 - .93)

Table 1.2 continued

Source	Instrument/Measure	Target condition	Number of groups	Sensitivity (95% CI)	Specificity (95% CI)
	Memory (administration time $\geq$ 20 mins)				
Weissberger et al. (2017)	Immediate recall	AD	26	.87 (.83 - .90)	.88 (.85 - .90)
	Immediate word list free recall	AD	17	.87 (.83 - .90)	.88 (.85 - .91)
	Immediate story free recall	AD	3	.71 (.61 - .78)	.75 (.58 - .86)
	Immediate word list cued recall	AD	7	.87 (.78 - .93)	.93 (.87 - .97)
	Immediate visual free recall	AD	4	.92 (.86 - .96)	.90 (.78 - .95)
	Delayed recall	AD	27	.89 (.87 - .91)	.89 (.87 - .91)
	Delayed word list free recall	AD	16	.90 (.86 - .92)	.87 (.84 - .89)
	Delayed story free recall	AD	4	.93 (.84 - .98)	.89 (.79 - .94)
	Delayed word list cued recall	AD	10	.91 (.87 - .94)	.92 (.88 - .95)
	Delayed word list retention	AD	5	.84 (.73 - .91)	.81 (.77 - .84)
	Delayed visual free recall	AD	6	.86 (.82 - .89)	.88 (.85 - .91)
	Language				
Tsoi et al. (2015)	Verbal fluency	dementia	7	.80 (.73 - .86)	.82 (.73 - .88)
	Visuospatial/visuoconstructive				
Park, Jeong, and Seomun (2018)	Clock Drawing Test (Shulman scoring system)	AD/dementia	7	.82 (.73 - .88)	.76 (.54 - .90)
	Clock Drawing Test (Sunderland scoring system)	AD/dementia	5	.73 (.63 - .88)	.88 (.74 - .95)
	Clock Drawing Test (Rouleau scoring system)	AD/dementia	5	.61 (.52 - .68)	.80 (.68 - .88)
	Clock Drawing Test (Watson scoring system)	AD/dementia	3	.57 (.37 - .75)	.83 (.44 - .97)
Tsoi et al. (2015)	Clock Drawing Test (Shulman scoring system)	dementia	9	.83 (.75 - .89)	.84 (.69 - .92)
	Clock Drawing Test (Sunderland scoring system)	dementia	9	.76 (.69 - .83)	.85 (.76 - .91)

Diagnostic accuracy was based on the ability of the tests to discriminate the target condition from people without AD or dementia. AD: Alzheimer's disease dementia, MCI: mild cognitive impairment; n.r.: not reported. 3MS: Modified Mini-Mental State, 7MS: 7 Minute Screen, ACE: Addenbrooke's Cognitive Examination, ACE-R: Addenbrooke's Cognitive Examination Revised, AMT: Abbreviated Mental Test, GPCOG: General Practitioner Assessment of Cognition, MIS: Memory Impairment Screen, MMSE: Mini-Mental State Examination, MoCA: Montreal Cognitive Assessment, RUDAS: Rowland Universal Dementia Assessment Scale

MCI, pooled sensitivity point estimates ranged from .62 - .71 and pooled specificity point estimates ranged from .74 - .87. One systematic review reported the diagnostic accuracy for MMSE in people with progressive MCI, which consisted of patients who were diagnosed with MCI at baseline and later progressed into all-cause dementia and AD at follow-up (Arevalo-Rodriguez et al., 2015). Sensitivity point estimate reported in the MCI to all-cause dementia diagnostic group was .40 with corresponding specificity point estimate of .88, and sensitivity point estimate in MCI to AD diagnostic group was .54 with corresponding specificity point estimate of .80.

The next most studied cognitive screen was MoCA, with reported pooled point estimates of sensitivity and specificity ranging from .81 - .89 and .74 - .86 respectively (Table 1.3). As shown in Table 1.3, two other cognitive screens were reported in more than one study, namely the ACE-R (sensitivity point estimates = .82 and .83, corresponding specificity point estimates = .78 and .77) and the Qmci (sensitivity point estimates = .77 and .86, corresponding specificity point estimates = .79 and .86).

Two systematic reviews analysed the diagnostic accuracy of memory measure in MCI. The pooled point estimates of sensitivity and specificity reported in Weissberger et al. (2017) ranged from .69 - .75 and .74 - .85 respectively, while pooled point estimates of sensitivity and specificity reported in Tsoi et al. (2017) ranged from .71 - .91 and .61 - .86 (Table 1.3). One systematic review by Belleville, Fouquet, Hudon, Zomahoun, and Croteau (2017) analysed the diagnostic accuracy of memory measure in people with MCI who later developed AD and the authors reported sensitivity point estimates of .55 - .78 and specificity point estimates of .79 - .90 (Table 1.3).

Two systematic reviews, by Tsoi et al. (2017) and Belleville et al. (2017), reported the pooled point estimates of sensitivity and specificity of other measures of cognitive function, including language, executive functioning and visuospatial skills (Table 1.3). Sensitivity and

Table 1.3*Diagnostic Accuracies of Cognitive Assessment Instruments and Measures for Cognitive Mild Impairment*

Source	Instrument/Cognitive Measure	Target condition	Number of groups	Sensitivity (95% CI)	Specificity (95% CI)
Global Screen					
Breton et al. (2019)	ACE-R	MCI	6	.82 (.77 - .87)	.78 (.67 - .86)
De Roeck et al. (2019)	ACE-R	MCI	7	.83 (n.r.)	.77 (n.r.)
	CAMCOG	MCI	4	.77 (.66 - .85)	.78 (.70 - .84)
Breton et al. (2019)	CERAD	MCI	5	.82 (.78 - .86)	.76 (.67 - .83)
De Roeck et al. (2019)	DemTect	MCI	3	.84 (n.r.)	.92 (n.r.)
Arevalo-Rodriguez et al. (2015)	MMSE	pMCI (all-cause dementia)	4	.40 (.23 - .76)	.88 (.40 - .94)
	MMSE	pMCI	8	.54 (.27 - .89)	.80 (.33 - .95)
Breton et al. (2019)	MMSE	MCI	46	.66 (.61 - .72)	.74 (.69 - .78)
Tsoi et al. (2015)	MMSE	MCI	21	.62 (.52 - .71)	.87 (.80 - .92)
Tsoi et al. (2017)	MMSE	MCI	58	.71 (.66 - .75)	.74 (.70 - .78)
De Roeck et al. (2019)	MoCA ^a	MCI	20	.84 (n.r.)	.75 (n.r.)
	MoCA ^b	MCI	7	.83 (n.r.)	.86 (n.r.)
Breton et al. (2019)	MoCA	MCI	24	.81 (.77 - .85)	.74 (.68 - .80)
Tsoi et al. (2015)	MoCA	MCI	9	.89 (.84-.92)	.75 (.62 - .85)
Tsoi et al. (2017)	MoCA	MCI	35	.83 (.8 - .92)	.75 (.69 - .80)
Breton et al. (2019)	Qmci	MCI	5	.77 (.71 - .82)	.79 (.71 - .85)
De Roeck et al. (2019)	Qmci	MCI	3	.86 (n.r.)	.86 (n.r.)
Breton et al. (2019)	M@T	MCI	5	.95 (.89 - .98)	.84 (.67 - .93)

Table 1.3 continued

Source	Instrument/Cognitive Measure	Target condition	Number of groups	Sensitivity (95% CI)	Specificity (95% CI)
Memory (administration > 10 - 20 mins)					
Weissberger et al. (2017)	Immediate recall (verbal or visual)	MCI	17	.72 (.63 - .79)	.81 (.75 - .85)
	Immediate word list free recall	MCI	13	.72 (.62 - .81)	.81 (.75 - .86)
	Immediate story free recall	MCI	6	.74 (.50 - .89)	.74 (.60 - .84)
	Immediate list cued recall	MCI	3	.74 (.54 - .87)	.84 (.73 - .90)
	Delayed recall (verbal or visual)	MCI	22	.75 (.69 - .81)	.81 (.77 - .84)
	Delayed list free recall	MCI	15	.73 (.64 - .81)	.83 (.77 - .88)
	Delayed story free recall	MCI	8	.74 (.56 - .86)	.79 (.70 - .85)
	Delayed list cued recall	MCI	5	.72 (.58 - .86)	.85 (.76 - .91)
	Delayed visual free recall	MCI	3	.69 (.33 - .91)	.82 (.64 - .92)
Tsoi et al. (2017)	Immediate word list free recall	MCI	3	.71 (.65 - .77)	.86 (.80 - .91)
	Immediate story free recall	MCI	4	.82 (.70 - .90)	.61 (.39 - .79)
	Short-delay word list recall (2 - 3 mins delay)	MCI	5	.92 (.87 - .95)	.79 (.74 - .83)
	Long-delay word list recall	MCI	9	.88 (.83 - .92)	.84 (.77 - .89)
	Long-delayed story recall	MCI	6	.91 (.82 - .96)	.83 (.74 - .89)
	Overall recall measures	MCI	15	.89 (.86 - .92)	.84 (.79 - .89)
Belleville et al. (2017)	Immediate verbal recall	pMCI	5	.75 (.51 - .91)	.77 (.58 - .90)
	Delayed story recall	pMCI	5	.78 (.54 - .93)	.79 (.68 - .91)
	Delayed word-list free delayed recall with non-oriented encoding ^c	pMCI	4	.74 (.63 - .87)	.81 (.52 - .96)
	Delayed word-list free delayed recall with oriented encoding ^c	pMCI	5	.78 (.70 - .87)	.80 (.65 - .89)
	Delayed word list cued recall with oriented encoding ^c	pMCI	3	.68 (.53 - .80)	.90 (.82 - .95)
	Delayed word list recognition	pMCI	3	.55 (.43 - .67)	.79 (.60 - .93)
	Delayed visual free recall	pMCI	4	.68 (.41 - .89)	.85 (.72 - .91)

Table 1.3 continued

Source	Instrument/Cognitive Measure	Target condition	Number of groups	Sensitivity (95% CI)	Specificity (95% CI)
Language					
Tsoi et al. (2017)	Verbal fluency	MCI	9	.57 (.45 - .68)	.74 (.61 - .84)
Belleville et al. (2017)	Naming	pMCI	6	.70 (.51 - .87)	.71 (.62 - .81)
	Semantic knowledge	pMCI	3	.70 (.44 - .91)	.81 (.63 - .98)
	Semantic fluency	pMCI	6	.71 (.47 - .88)	.70 (.54 - .84)
Executive Functioning					
Belleville et al. (2017)	Switching tasks	pMCI	3	.54 (.31 - .70)	.68 (.54 - .80)
	Working memory	pMCI	3	.60 (.46 - .72)	.67 (.57 - .75)
Visuospatial/visuoconstructive					
Tsoi et al. (2017)	Clock Drawing Test	MCI	16	.58 (.46 - .69)	.73 (.61 - .83)
Belleville et al. (2017)	Visuospatial tasks	pMCI	4	.68 (.37 - .91)	.75 (.61 - .87)
	Visuoconstructive tasks	pMCI	4	.64 (.47 - .78)	.64 (.50 - .78)

^a sensitivity and specificity from memory clinic-based studies

^b sensitivity and specificity from population-based studies

^c oriented encoding referred to wordlists that had groups of words that were semantically-related, while non-oriented encoding referred to wordlists that consisted of unrelated words.

Diagnostic accuracies were the ability of the tests to discriminate MCI from healthy controls, or people with pMCI from people with MCI at baseline who did not develop Alzheimer's disease type dementia at follow-up. MCI: mild cognitive impairment; pMCI: progressive mild cognitive impairment, diagnosed with Alzheimer's disease at follow-up, unless otherwise stated; n.r.: not reported

ACE-R: Addenbrooke's Cognitive Examination - Revised ACE-R, CERAD: Consortium to Establish a Registry for Alzheimer's Disease neuropsychological battery, MMSE: Mini-Mental State Examination, MoCA: Montreal Cognitive Assessment, Qmci: Quick Mild Cognitive Impairment screen, M@T: Memory Alteration Test.

specificity point estimates obtained from these non-memory measures were generally lower than that seen in the memory measures, with sensitivity point estimates ranging from .54 - .71 and specificity point estimates ranging from .64 - .81 (Table 1.3).

Performance of Fluid and Neuroimaging Biomarkers in Mild Cognitive Impairment and Alzheimer's Disease

Four systematic reviews investigated the diagnostic accuracy of CSF A β ₄₂ for detecting AD, amnesic MCI and progressive MCI, in which progressive MCI referred to individuals who were diagnosed with MCI at baseline and was later diagnosed to have AD at follow-up (Ferreira et al., 2014; Mo et al., 2015; Ritchie et al., 2014; Rivero-Santana et al., 2017). As shown in Table 1.4, sensitivity point estimates of CSF A β ₄₂ ($k = 91$) ranged from .77 - .85 and corresponding reported specificity point estimates ranged from .61 - .84. Three systematic reviews reported the diagnostic accuracy of CSF t-tau and p-tau (Ferreira et al., 2014; Ritchie et al., 2014; Rivero-Santana et al., 2017). Sensitivity and specificity point estimates of CSF t-tau measures ($k = 54$) ranged from .74 - .78 and .58 - .82 respectively, while sensitivity and specificity point estimates of CSF p-tau measures ($k = 45$) ranged from .78 - .84 and .48 - .78 respectively (Table 1.4). Systematic reviews that evaluated different combinations of A β ₄₂, t-tau, p-tau or the ratios of t-tau/A β ₄₂ and p-tau/A β ₄₂ reported sensitivity point estimates of .84 - .89 and specificity point estimates of .63 - .84 ($k = 98$; Table 1.4).

There were fewer systematic reviews that examined the diagnostic test accuracy of amyloid PET imaging. One systematic review by Zhang et al. (2014) examined the diagnostic accuracy of ¹¹C-PIB in detecting progressive MCI ($k = 9$) and reported an estimated sensitivity of .96 based on a median specificity of .58 (Table 1.4). There were limited systematic reviews on the diagnostic accuracy of ¹⁸F-labelled tracers. Only one systematic review did a meta-analysis of studies on ¹⁸F-labelled tracer using AD as a target condition

Table 1.4*Diagnostic Accuracies of CSF, PET and MRI Biomarkers for Detecting AD and MCI*

Biomarker and Study Source	Target condition	Number of groups	Range of study sensitivities and specificities reported in the systematic review				Pooled single sensitivity and specificity reported in the systematic review			
			Sensitivity	(95% CI)	Specificity	(95% CI)	Sensitivity	(95% CI)	Specificity	(95% CI)
CSF										
<i>Aβ₄₂</i>										
Ferreira et al. (2014)	AD	11	.64 - .97	(n.r.)	.68 - .92	(n.r.)	.83	(SD = .09)	.80	(SD = .08)
Mo et al. (2015)	AD	12	.70 - 1.00	(.51 - 1.00)	.73 - .95	(.52 - 1.00)	.84	(.82 - .86)	.84	(.82 - .87)
Ferreira et al. (2014)	AD vs NADD	5	.62 - .95	(n.r.)	.22 - .80	(n.r.)	.85	(SD = .05)	.61	(SD = .24)
Mo et al. (2015)	AD vs NADD	16	.71 - .91	(.62 - .97)	.44 - .95	(.22 - .98)	.80	(.78 - .82)	.76	(.74 - .78)
Rivero-Santana et al. (2017)	AD vs FTD	16	.50 - .98	(n.r.)	.21 - 1.00	(n.r.)	.82 ^d	(.72 - .88)	.70	n/a
Mo et al. (2015)	aMCI vs naMCI	8	.54 - .85	(.25 - .97)	.50 - .85	(.01 - .99)	.77	(.73 - .81)	.65	(.61 - .69)
Ferreira et al. (2014)	pMCI	9	.55 - .91	(n.r.)	.36 - .96	(n.r.)	.79	(SD = .14)	.63	(SD = .20)
Ritchie et al. (2014)	pMCI	14	.36 - 1.00	(.11 - 1.00)	.29 - .91	(.36 - .99)	.81 ^d	(.72 - .87)	.64	n/a
<i>T-tau</i>										
Ferreira et al. (2014)	AD	12	.61 - .91	(n.r.)	.53 - .97	(n.r.)	.78	(SD = .09)	.82	(SD = .14)
Ferreira et al. (2014)	AD vs NADD	4	.62 - .92	(n.r.)	.40 - .93	(n.r.)	.75	(SD = .14)	.71	(SD = .22)
Rivero-Santana et al. (2017)	AD vs FTD	22	.25 - .89	(n.r)	.54 - .94	(n.r.)	.74 ^d	(.68 - .79)	.75	n/a
Ferreira et al. (2014)	pMCI	9	.61 - .88	(n.r.)	.39 - .88	(n.r.)	.76	(SD = .12)	.58	(SD = .17)
Ritchie et al. (2017)	pMCI	7	.51 - .90	(.21 - .98)	.48 - .88	(.27 - .96)	.75 ^d	(.67 - .85)	.72	n/a
<i>P-tau</i>										
Ferreira et al. (2014)	AD	12	.61 - .87	(n.r.)	.37 - .92	(n.r.)	.78	(SD = .10)	.77	(SD = .18)
Ferreira et al. (2014)	AD vs NADD	4	.77 - .88	(n.r.)	.56 - .88	(n.r.)	.80	(SD = .08)	.78	(SD = .15)
Rivero-Santana et al. (2017)	AD vs FTD	16	.44 - .97	(n.r.)	.45 - 1.00	(n.r.)	.84 ^d	(.77 - .89)	.78	n/a
Ferreira et al. (2014)	pMCI	7	.64 - .85	(n.r.)	.30 - .90	(n.r.)	.78	(SD = .09)	.56	(SD = .18)
Ritchie et al. (2017)	pMCI	6	.40 - .10	(.05 - .93)	.22 - .86	(.03 - .93)	.81 ^d	(.64 - .91)	.48	n/a

Table 1.4 continued

Biomarker and Study Source	Target condition	Number of groups	Range of study sensitivities and specificities reported in the systematic review				Pooled single sensitivity and specificity reported in the systematic review			
			Sensitivity	(95% CI)	Specificity	(95% CI)	Sensitivity	(95% CI)	Specificity	(95% CI)
CSF										
<i>T-tau/Aβ₄₂ ratio</i>										
Rivero-Santana et al. (2017)	AD vs FTD	9	.50 - .97	(n.r.)	.40 - 1.00	(n.r.)	.87 ^d	(.81 - .90)	.80	n/a
Ritchie et al. (2017)	pMCI (all cause dementia)	4	.42 - .79	(n.r.)	.64 - .95	(n.r.)	-	-	-	-
<i>P-tau/Aβ₄₂ ratio</i>										
Rivero-Santana et al. (2017)	AD vs FTD	10	.77 - .93	(n.r.)	.63 - 1.00	(n.r.)	.89 ^d	(.68 - .97)	.79	n/a
Ritchie et al. (2017)	pMCI	5	.50 - .51	(n.r.)	.91 - .96	(n.r.)	-	-	-	-
<i>Combination of Aβ₄₂ and/or p-tau and/or t-tau</i>										
Ferreira et al. (2014)	AD	25	.70 - .98	(n.r.)	.53 - .97	(n.r.)	.87	(SD = .06)	.84	(SD = .09)
Ferreira et al. (2014)	AD vs NADD	19	.67 - 1.00	(n.r.)	.36 - .97	(n.r.)	.86	(SD = .10)	.78	(SD = .14)
Cure et al. (2014)	AD vs HC/NADD	7	.68 - .96	(.54 - 1.00)	.60 - .92	(.35 - .98)	.89 ^c	(.84 - .95)	.71 ^c	(.60 - .81)
Ferreira et al. (2014)	pMCI	19	.57 - .98	(n.r.)	.38 - .95	(n.r.)	.84	(SD = .10)	.63	(SD = .19)
Amyloid PET										
<i>¹¹C-PIB</i>										
Zhang et al. (2014)	pMCI ^{a, b}	9	.83 - 1.00	(.40 - 1.00)	.46 - .88	(.28 - 1.00)	.96 ^d	(.87 - .99)	.58	n/a
<i>¹⁸F-Florbetaben</i>										
Rice and Bisdas (2017)	AD	1	.80	(.71 - .89)	.91	(.84 - .98)	-	-	-	-
Rice and Bisdas (2017)	AD vs NADD	1	.97	(n.r.)	.88	(n.r.)	-	-	-	-
Martinez, Vernooij, et al. (2017a)	pMCI ^{a, b}	2	1.00 - 1.00	(.84 - 1.00)	.83 - .88	(.63 - .97)	-	-	-	-
<i>¹⁸F-Florbetapir</i>										
Rice and Bisdas (2017)	AD	1	.85	(.55 - .98)	.38	(.18 - .62)	-	-	-	-
Cure et al. (2014)	AD vs NADD ^{a, b}	1	.92	(.79 - .98)	1.00	(.83 - 1.00)	.92 ^c	(.74 - 1.00)	1.00 ^c	(1.00 - 1.00)
Martinez et al. (2017)	pMCI ^{a, b}	3	.67 - .89	(.30 - .94)	.51 - .71	(.45 - .85)	-	-	-	-

Table 1.4 continued

Biomarker and Study Source	Target condition	Number of groups	Range of study sensitivities and specificities reported in the systematic review				Pooled single sensitivity and specificity reported in the systematic review			
			Sensitivity	(95% CI)	Specificity	(95% CI)	Sensitivity	(95% CI)	Specificity	(95% CI)
Amyloid PET										
<i>¹⁸F -Flutemetamol</i>										
Rice and Bisdas (2017)	AD/MCI	1	.93	(n.r.)	.93	(n.r.)	-	-	-	-
Martinez, Vernooij, et al. (2017b)	pMCI ^{a, b}	2	.64 - .89	(.52 - 1.00)	.69 - .80	(.44 - .97)	-	-	-	-
<i>¹⁸F-labelled tracer (Florbetaben/Florbetapir/Flutemetamol)</i>										
Morris et al. (2016)	AD ^a	7	.89 - .95	(.55 - .98)	.81 - .89	(.24 - .98)	.90	(.82-.95)	.85	(.68 - .94)
Morris et al. (2016)	AD ^b	5	.91 - .94	(.59 - .99)	.63 - .87	(.42 - .94)	.90	(.83 .94)	.84	(.71 - .91)
FDG-PET										
Rice and Bisdas (2017)	AD	2	.78 - .83	(n.r.)	.78 - .94	(n.r.)	-	-	-	-
Rice and Bisdas (2017)	AD/MCI	1	.91	(n.r.)	.87	(n.r.)	-	-	-	-
Rice and Bisdas (2017)	AD vs NADD	1	.85	(n.r.)	.90	(n.r.)	-	-	-	-
Cure et al. (2014)	AD vs HC/NADD	5	.84 - .94	(.64 - 1.00)	.57 - .84	(.24 - .95)	.93 ^c	(.89 - .97)	.73 ^c	(.63 - .83)
Rice and Bisdas (2017)	AD vs HC/NADD	1	.98	(n.r.)	.99	(n.r.)	-	-	-	-
Smailagic et al. (2015)	pMCI	14	.25 - 1.00	(.03 - 1.00)	.15 - 1.00	(.02 - 1.00)	.76 ^d	(.54 - .90)	.82	n/a
Smailagic, Lafortune, Kelly, Hyde, and Brayne (2018)	pMCI	11	.63 - 1.00	(.21 - 1.00)	.24 - 1.00	(.11 - 1.00)	-	-	-	-
MRI										
Cure et al. (2014)	AD vs HC/NADD ^{a, b}	5	.43 - .92	(.23 - 1.00)	.10 - .94	(.00 - 1.00)	.75 ^c	(.67 - .84)	.80 ^c	(.73 - .87)

^a Visual analysis^b Quantitative analysis^c Derived from a bivariate random-effect model, with imaging or non-imaging techniques as covariates, to calculate the sensitivity and specificity for each imaging modality^d Meta-analytical estimate of the sensitivity derived from the hierarchical summary receiver operating characteristic (HSROC) model at the median specificity

Diagnostic accuracies were based on the ability of the test to discriminate the target condition from the healthy controls unless otherwise specified. AD: Alzheimer's disease dementia, HC: healthy controls MCI: mild cognitive impairment, pMCI: progressive MCI with conversion to Alzheimer's disease at follow-up, aMCI: amnesic mild cognitive impairment, naMCI: non-amnesic mild cognitive impairment, NADD: non-Alzheimer's disease dementia

(Morris et al., 2016). The authors of that systematic review reported a pooled sensitivity of .90 and pooled specificity of .85 on studies that used visual interpretation ($k = 7$), and a pooled sensitivity of .90 and pooled specificity of .84 based on studies that used quantitative interpretation ($k = 5$; Table 1.4). Other systematic reviews that consisted of three or fewer studies investigating the diagnostic accuracy of various ^{18}F -labelled tracers reported sensitivity point estimates of .64 – 1.00 and specificity point estimates of .51 – 1.00 (Table 1.4).

Discussion

The aim of this review was to evaluate the diagnostic accuracy of available instruments for cognitive assessment and other fluid and neuroimaging biomarkers in reliably identifying MCI and AD. The cognitive tests selected included screening tools as well as longer cognitive assessments. Although there has been an increasing number of available fluid and neuroimaging biomarkers under evaluation, only the clinically well-validated biomarkers identified in both the IWG and NIA-AA research criteria were reviewed. As recommended in the paper by the Ronald and Nancy Regan Research Institute and the NIA (1998), sensitivity and specificity of more than .80 were considered to be good.

Cognitive Measures

The review of the systematic reviews that examined cognitive screening tools showed that although similar reasonable specificity point estimates that were between .70 to .90 were reported for the MMSE in the detection of people with MCI and AD (Table 1.2; Table 1.3), the reported sensitivities were quite disparate. While the MMSE had good sensitivity in detecting people with AD (Table 1.2), this cognitive screen was particularly poor at detecting people with MCI, with all reported sensitivity point estimates falling well below the recommended range of .80 (Table 1.3). This finding suggests that while the MMSE can reasonably distinguish people with AD from people without AD, this screening test was poor

at providing information on whether someone does or does not have MCI. On the other hand, the MoCA showed similar ability to MMSE in detecting AD but was superior to the MMSE in distinguishing people with MCI from the people without MCI (Table 1.2; Table 1.3). Another screening tool that presented with promising results was the ACE-R. Although the findings from the ACE-R were based on a slightly smaller number of studies, all available systematic reviews reported sensitivity point estimates of more than .80 and specificity point estimates of more than .75 in identifying people with MCI from the healthy controls (Table 1.3), and sensitivity point estimates of more than .90 and specificity point estimates of more than .90 in discriminating AD from healthy controls (Table 1.2). Most of the other cognitive screening tests reported sensitivity point estimates of above .80 in the ability to detect MCI and AD, but these numbers were based on a small number of studies.

The MMSE is one of the mostly widely utilised cognitive screening tools, but it was originally designed for the detection of dementia (Folstein et al., 1975). The current results from the various systematic reviews do suggest that the MMSE performs reasonably well in the original context that it was designed for but falls short in the ability to accurately detect the more subtle cognitive changes seen in people with MCI. Given the findings from the systematic reviews, the MoCA or the ACE-R would be the better alternative to MMSE. Of note, ACE-R has been updated with the ACE-III, which was designed to replace the MMSE items on the ACE-R and to address some of the weaknesses in the ACE-R (Hodges & Larner, 2017; Hsieh, Schubert, Hoon, Mioshi, & Hodges, 2013). Although no systematic reviews of the ACE-III were found in the current review, diagnostic test accuracy of the ACE-III for detecting dementia and MCI has been reported to be similar to the ACE-R (Hodges & Larner, 2017; Stott, Scior, Mandy, & Charlesworth, 2017).

On longer memory measures, diagnostic accuracy statistics comparing AD against healthy controls yielded excellent sensitivity and specificity, with reported point estimates

that were well above the recommended .80 (Table 1.2). On the other hand, diagnostic test accuracy of longer memory measures in the MCI groups were weaker and varied between the studies. The sensitivity point estimates of the various memory measures reported in two of the systematic reviews mostly fell in the .70 - .80 range, while one systematic review reported sensitivities that were mostly above .80 and up to .92 (Table 1.3). Despite this variability, the specificity point estimates of the different memory measures reported in all three systematic reviews were very similar, generally above .80 (Table 1.3). Specificity values reported for immediate and delayed recall were similar between the cross-sectional and longitudinal studies comparing MCI with healthy controls.

Diagnostic performance of the cognitive measures for language, executive functioning and visuospatial skills were generally unsatisfactory, particularly in the MCI group comparisons. Most of the sensitivity and specificity point estimates reported in the systematic reviews of the diagnostic test accuracy of these measures for diagnosing MCI fell well below the recommended .80 value (Table 1.3). Although diagnostic test accuracy statistics for verbal fluency in the AD group comparisons were satisfactory, the sensitivity point estimates of the clock drawing tests were highly variable and majority of the values fell below .80 (Table 1.2). However, the interpretability of the diagnostic performance of these cognitive measures is limited by the small number of studies that were included in the systematic reviews and further research is necessary to validate the current findings.

Fluid and Neuroimaging Biomarkers

The most well-studied AD biomarkers found in this review were the CSF biomarkers. By and large, the combinations of CSF biomarkers which included combinations of $A\beta_{42}$, t-tau and p-tau, and the $A\beta_{42}$ to tau ratios, produced the best sensitivity point estimates of .84 - .89 for both AD and MCI comparisons to healthy controls or non-Alzheimer's disease dementia (Table 1.4). Interpretation of individual CSF biomarkers produced slightly lower

sensitivity point estimates, .74 - .85, but the values were largely similar between both AD and MCI group comparisons (Table 1.4). The similarity between the reported sensitivity values for both AD and MCI groups is consistent with the concept of AD being a protracted continuum of neuropathologic changes, with MCI stages representing the early symptomatic expression of the disease (Dubois, Hampel, et al., 2016; Sperling et al., 2011).

However, although CSF A β ₄₂ had good diagnostic accuracy at distinguishing AD from healthy controls, specificity was generally weaker for discriminating AD from other kinds of dementia, and for discriminating people with progressive MCI to those who remained stable, resulting in an inflated probability of false positives in people without AD pathology. In fact, majority of the specificity values reported for CSF biomarkers in the various systematic reviews did not meet the proposed .80 specificity value, and specificity values were much lower in the MCI comparison groups.

An interesting point highlighted in the biomarker literature is that phosphorylated tau protein is believed to be one of the key contributors to AD-related neurofibrillary tangle formation in the brain and therefore could add more specificity to the diagnosis of AD-related disorders (Sperling & Johnson, 2013). However, the question of whether or not p-tau measures corresponded directly to pathological tau burden in the brain remains unclear (Sperling & Johnson, 2013). The current reported sensitivity point estimates of CSF p-tau was only slightly better than CSF t-tau, and both biomarkers had similar specificity point estimates that were mostly below .80. Specificity point estimates for progressive MCI was particularly poor for both p-tau and t-tau. These findings are consistent with the limitation of CSF tau measures raised by Sperling and Johnson (2013) – that the correlation between the different tau measures and tau burden were not necessarily direct.

Among the five main biomarkers reviewed, as seen in Table 1.4, ¹¹C-PIB-PET showed one of the highest sensitivity estimates but also one of the lowest specificity

estimates in predicting the conversion of people with MCI to dementia of the Alzheimer's type (Zhang et al., 2014). One of the limitations of using ^{11}C -PIB is that this radioligand binds not only to cortical amyloid, but it also has an affinity to amyloid in blood vessels such as those seen in cerebral amyloid angiopathy, which could result in a high rate of false positive (Farid, Charidimou, & Baron, 2017; Sperling et al., 2011; Sperling & Johnson, 2013).

In the systematic reviews in AD that investigated the newer amyloid PET radioligands that were labelled with ^{18}F isotopes, the reported diagnostic accuracies were more well-balanced, with one study by Morris et al. (2016) reporting single sensitivity and specificity values that were both above .80 in the AD patient group. The reported sensitivities and specificities of other imaging modalities, namely FDG-PET and MRI, also appeared to be well-balanced. However, all these imaging systematic reviews were plagued by the same issues. Besides the small number of studies available for inclusion in the systematic reviews, there was a wide range of reported values and confidence intervals, variable thresholds for diagnostic interpretation, and consequently a lack of information to pursue meaningful meta-analysis particularly in the MCI patient group. Therefore, although diagnostic accuracies of the various imaging modalities appeared promising, there was very limited quality evidence of the suitability of using these biomarkers as a reliable method for detecting and distinguishing people with MCI from those without true AD-related MCI.

Summary of Findings

Overall, the current diagnostic accuracy findings from the cognitive tests and the various biomarkers highlights the challenging nature of diagnosing AD in the earlier symptomatic stages, or in other words, AD-related MCI. The aetiology of impairment underlying people with MCI is heterogenous and it is difficult to determine whether people with MCI would later progress to develop AD or other dementias (Albert et al., 2011;

Dubois, Hampel, et al., 2016). Some studies have found that patients diagnosed with MCI do not always transition to a dementia of any sort, while other patients were later spontaneously found to not have any cognitive impairment, which indicates that MCI is not a stable longitudinal diagnosis (Klekociuk, Saunders, & Summers, 2016; Klekociuk, Summers, Vickers, & Summers, 2014; Summers & Saunders, 2012).

In general, the available cognitive screening tools, memory measures and some measures of cognitive functioning showed very good sensitivities and specificities in the detection of AD. On the other hand, the diagnostic test accuracies of the cognitive tools and measures were more variable in the detection of MCI. The majority of the cognitive screening tests for MCI that met the recommended sensitivity of .80 were associated with specificities below .80. However, memory measures showed the reverse pattern, in which even though most reported sensitivities were below .80 the majority of the corresponding specificities were above .80. All other cognitive domains assessed reported poor diagnostic test accuracy in MCI detection. Although the core clinical features of MCI include an insidious decline in any number of cognitive domains without loss in daily functional abilities, an early amnesic presentation in MCI has been shown to be one of the distinguishing precursor features of AD (Bradfield et al., 2018; Petersen et al., 2009; Sarazin et al., 2007). The findings in this review emphasise the importance of assessing memory function and that the presence of memory dysfunction will play a core role in separating the worried-well from the individuals who are experiencing the early cognitive signs of AD. Although ideally, diagnostic cognitive assessments should have both high sensitivity and high specificity, the findings from this review suggest that such accuracy is difficult to achieve particularly in the detection of people with MCI. Therefore, there may be greater value in the use of biomarkers as a complementary system to support diagnostic clarification (Jack et al., 2016). In the MCI group comparisons, the various biomarkers showed a pattern of high

sensitivity but poor specificity, which suggests that biomarkers would play a more useful role in ruling out, rather than ruling in MCI in a clinical context.

The assumption that biomarkers may be more valuable in ruling out rather than ruling in a disease is consistent with the view that was presented in a consensus paper on the use of CSF biomarkers from the Alzheimer's Biomarkers Standardization Initiative (Molinuevo et al., 2014). The same paper emphasised that the interpretation of results should be made using all the three CSF biomarkers, A β ₄₂, t-tau and p-tau (Molinuevo et al., 2014). Investigations using CSF biomarkers should also be done in tandem with other biomarkers such as amyloid PET, FDG-PET, and MRI, in addition to careful neuropsychological evaluation and the overall clinical picture (Molinuevo et al., 2014; Trzepacz et al., 2014). Therefore, for a patient who tests negative on both memory measures and biomarkers, it would be clinically reassuring that the person in question likely do not have MCI. However, should a patient test negative on memory measures but positive on biomarkers, in order to mitigate the possibility of a missed diagnosis from memory testing and the high rates of false positives observed in fluid and neuroimaging biomarker investigations, it would be advisable to provide further clinical investigation or keep the patient on follow-up for clinical diagnostic clarification.

Future Directions, Limitations and Strengths

Most of the systematic reviews found in this review focused on cognitive screening tests for AD and MCI. There is a need for more studies and systematic reviews on comprehensive neuropsychological assessment particularly for MCI, as some studies have shown that the diagnosis of MCI based on comprehensive testing yielded a more accurate diagnosis of progressive MCI and had a greater association with AD-related physiological biomarkers (Bondi et al., 2014; Klekociuk et al., 2014). Related to the need for comprehensive neuropsychological assessment, there was an absence of systematic reviews of the diagnostic test accuracies of test batteries that used composite scores. Composite

scores, such as memory index scores found in test batteries such as WMS-III/IV and RBANS, could be important for improving accuracy in identifying MCI patients. Index scores could improve detection due to increased precision since the scores are typically more reliable and strongly correlated to the operationalised theoretical construct that it is assessing (Nunnally, 1978).

One of the difficulties in analysing and interpreting diagnostic accuracy tests lies in the variable cut-offs used. Despite the recommended cut-offs for various cognitive tests, the reported cut-off values for common cognitive screens such as the MMSE ranged from 13 to 29 (Arevalo-Rodriguez et al., 2015; Breton, Casey, & Arnaoutoglou, 2019; Creavin et al., 2016; Larner & Mitchell, 2014; Paddick et al., 2017; Tsoi et al., 2017; Tsoi, Chan, Hirai, Wong, & Kwok, 2015). In the 2011 consensus paper, members of the IWG and NIA-AA agreed that it would have been premature to include specific biomarker cut-offs (Jack, Albert, et al., 2011). However, it is well-recognised that evaluations based on qualitative evaluations based on visual assessments of structural atrophy in MRI, of hypometabolism in FDG-PET and of amyloid uptake in the cerebrum, and quantitative evaluations based on variable cut-offs, would introduce inconsistencies in interpretation, and standardisation efforts for the various biomarkers are needed (Blennow & Zetterberg, 2018; Jack, Barkhof, et al., 2011; Molinuevo et al., 2014; Sheikh-Bahaei et al., 2017). Diagnostic accuracy can be biased by data-driven selection of optimal cut-off values rather than the recommended clinical cut-offs, which makes the generalisation of the findings to clinical settings more difficult (Leeflang, Deeks, Gatsonis, & Bossuyt, 2008). Given that cut-offs variability affects sensitivity and specificity estimates, Leeflang et al. (2008) recommended the bivariate model for data analysis or the construction of a hierarchical summary receiver-operating characteristic, to mitigate the effects of variable thresholds and therefore facilitate interpretability of diagnostic accuracy statistics.

There are several modalities currently being used in the detection of AD and MCI, namely cognitive testing, CSF biomarkers, amyloid PET, FDG-PET and MRI. However, an important observation from this review is that there is insufficient evidence to show that any one modality is superior to the others. Instead, it is more conceivable that a combination of these methods would provide superior discrimination than any individual test on its own as each of these modalities contribute unique information for diagnostic clarification. Most of the research on AD and MCI have placed focus only on cognitive or fluid and neuroimaging biomarker measures. However, several studies have already shown that a multimodal approach, which includes combination of different biomarkers or combining biomarkers with cognitive measures, improves differentiation between AD, MCI and healthy controls (Bauer, Cabral, & Killiany, 2018; Gross et al., 2012). In this review, no systematic reviews on a multimodal approach to identifying AD and MCI was found, which highlights that need for future studies to identify combinations of cognitive and biomarker measures that could best distinguish MCI and predict progression.

An important consideration for future studies is the need for longitudinal studies to validate the reliable identification of people with MCI due to AD, especially in the elderly when only one cognitive domain on screening assessment lies around 1.5 standard deviations below the mean. Cognitive assessment is one of the core diagnostic components for MCI, but MCI remains heterogeneous and potentially imprecise as a diagnostic construct (Albert et al., 2011; Petersen, 2004; Petersen et al., 1999). There is a recognised risk of false positives when diagnosis are based on an abnormal score on any single test, with a reported range of 26 – 56% of individuals potentially being misclassified when based on this conventional criterion (Binder, Iverson, & Brooks, 2009; Brooks, Iverson, Holdnack, & Feldman, 2008; Brooks, Iverson, & White, 2007; Edmonds et al., 2015; Edmonds et al., 2019). Biologically, biomarker findings have been similarly heterogeneous, with different AT(N) biomarker profiles

sometimes following patterns of longitudinal cognitive change not always consistent with the presumed sequence of biomarkers associated with development of AD dementia (Jack et al., 2019; Nettiksimmons, DeCarli, Landau, Beckett, & Alzheimer's Disease Neuroimaging, 2014). These findings reiterate the need to evaluate the contribution of both comprehensive cognitive and biomarker testing in diagnosing MCI using longitudinal data. Longitudinal evidence of progressive cognitive decline supported by relevant biomarker profiles will provide additional evidence of an accurate cross-sectional diagnosis of MCI that is due to AD. Such longitudinal data will help identify those cognitive and biological features that most accurately depict the high likelihood of progression of MCI hopefully parse those factors determining conversion from MCI to AD-related and non-AD-related dementia.

This current review has been limited to only systematic reviews, which makes it probable that there may be overlaps in individual studies between different systematic reviews and therefore the diagnostic accuracy statistics reported may not be entirely independent of each other. Furthermore, the newer diagnostic test accuracy studies that were published between the time of the publication of the systematic reviews and this review would also inevitably be missed. However, only the most relevant and recent systematic reviews that had assessed the quality of studies using the QUADAS-2 or similar standards were included in this review, which provides confidence in the quality of the findings while providing an excellent overview of the work that has been done in recent years.

A strength of this review is that a systematic approach was taken to review the available literature. The search terms used for this review were chosen to provide the highest precision for retrieving relevant publication on diagnostic test accuracy on MEDLINE via OvidSP (Beynon et al., 2013). Although the search period of this review was limited to the past five years, an earlier review of systematic reviews on cognitive screening tools had already been completed by Lin, O'Connor, Rossom, Perdue, and Eckstrom (2013).

Furthermore, according to Weissberger et al. (2017), there were no systematic reviews on longer memory measures prior to their paper. In addition, in order to maintain the relevance of findings to clinical applicability, only the well-validated biomarkers that were first introduced and detailed in the 2011 consensus papers were selected for review (Albert et al., 2011; Jack, Albert, et al., 2011). Therefore, this current search strategy was chosen as relevant systematic reviews were unlikely to be missed and substantial overlap with previous reviews was avoided. More importantly, the results from this search strategy likely captured a greater portion of studies that have used the newer diagnostic criteria and therefore would be more representative of current diagnostic methods and findings.

Conclusion

Current cognitive assessment tools showed adequate diagnostic accuracy in the detection of AD but had poorer diagnostic accuracy in detecting MCI. Of note, brief cognitive tools such as the MMSE had particularly poor sensitivity and other screening tools with greater sensitivity and better balance between sensitivity and specificity values, like the MoCA and the ACE-R should be considered instead. Overall, memory measures met the suggested guideline of having greater than .80 sensitivity and specificity for the differentiation of AD from controls. However, sensitivity of the memory measures for detecting MCI fell short of the recommended guideline, but specificity values of these memory measures were generally above .80. Among the main fluid and neuroimaging biomarkers recommended in the 2011 diagnostic guidelines, CSF measures were the best studied, while there were limited comprehensive systematic reviews on the diagnostic test accuracy of biomarkers using neuroimaging modalities. Overall, fluid and neuroimaging biomarkers showed a pattern of high sensitivity but low specificity, particularly in the MCI group comparisons. These findings suggest that there is no one ideal diagnostic test suitable

for the early detection of symptomatic AD. Future studies should therefore focus on longitudinal investigations that implements a multimodal approach to an early diagnosis AD.

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to 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), 270-279. doi:10.1016/j.jalz.2011.03.008
- Alzheimer's Association. (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 14(3), 367-425. doi:10.1016/j.jalz.2018.02.001
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text Revision). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Andersson, E., Janelidze, S., Lampinen, B., Nilsson, M., Leuzy, A., Stomrud, E., . . . Hansson, O. (2020). Blood and cerebrospinal fluid neurofilament light differentially detect neurodegeneration in early Alzheimer's disease. *Neurobiology of Aging*, 95, 143-153. doi:10.1016/j.neurobiolaging.2020.07.018
- Arevalo-Rodriguez, I., Smailagic, N., Roquei Figuls, M., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., . . . Cullum, S. (2015). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*(3), CD010783. doi:10.1002/14651858.CD010783.pub2
- Bauer, C. M., Cabral, H. J., & Killiany, R. J. (2018). Multimodal discrimination between normal aging, mild cognitive impairment and Alzheimer's disease and prediction of cognitive decline. *Diagnostics*, 8(1), 20. doi:10.3390/diagnostics8010014

- Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. T. V., & Croteau, J. (2017). Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. *Neuropsychology Review*. doi:10.1007/s11065-017-9361-5
- Beynon, R., Leeflang, M. M., McDonald, S., Eisinga, A., Mitchell, R. L., Whiting, P., & Glanville, J. M. (2013). Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database of Systematic Reviews*(9), MR000022. doi:10.1002/14651858.MR000022.pub3
- Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, 24(1), 31-46. doi:10.1093/arclin/acn001
- Blennow, K., & Zetterberg, H. (2018). The past and the future of Alzheimer's disease fluid biomarkers. *Journal of Alzheimer's Disease*, 62(3), 1125-1140. doi:10.3233/JAD-170773
- Bondi, M. W., Edmonds, E. C., Jak, A. J., Clark, L. R., Delano-Wood, L., McDonald, C. R., . . . Alzheimer's Dis Neuroimaging, I. (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *Journal of Alzheimer's Disease*, 42(1), 275-289. doi:10.3233/jad-140276
- Bradfield, N. I., Ellis, K. A., Savage, G., Maruff, P., Burnham, S., Darby, D., . . . Ames, D. (2018). Baseline amnesic severity predicts progression from amnesic mild cognitive impairment to Alzheimer disease dementia at 3 years. *Alzheimer Disease & Associated Disorders*, 32(3), 190-196. doi:10.1097/WAD.0000000000000252
- Breton, A., Casey, D., & Arnaoutoglou, N. A. (2019). Cognitive tests for the detection of mild cognitive impairment (MCI), the prodromal stage of dementia: Meta-analysis of

- diagnostic accuracy studies. *International Journal of Geriatric Psychiatry*, 34(2), 233-242. doi:10.1002/gps.5016
- Brookmeyer, R., Abdalla, N., Kawas, C. H., & Corrada, M. M. (2018). Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States. *Alzheimer's & Dementia*, 14(2), 121-129. doi:10.1016/j.jalz.2017.10.009
- Brooks, B. L., Iverson, G. L., Holdnack, J. A., & Feldman, H. H. (2008). Potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International Neuropsychological Society*, 14(3), 463-478. doi:10.1017/S1355617708080521
- Brooks, B. L., Iverson, G. L., & White, T. (2007). Substantial risk of "Accidental MCI" in healthy older adults: base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 13(3), 490-500. doi:10.1017/S1355617707070531
- Chen, X. Q., & Mobley, W. C. (2019). Alzheimer disease pathogenesis: Insights from molecular and cellular biology studies of oligomeric Abeta and tau species. *Frontiers in Neuroscience*, 13, 659. doi:10.3389/fnins.2019.00659
- Chi, H., Chang, H.-Y., & Sang, T.-K. (2018). Neuronal cell death mechanisms in major neurodegenerative diseases. *International Journal of Molecular Sciences*, 19(10), 3082. doi:10.3390/ijms19103082
- Clark, C. M., Schneider, J. A., Bedell, B. J., Beach, T. G., Bilker, W. B., Mintun, M. A., . . . AV45-A07 Study Group, f. t. (2011). Use of florbetapir-PET for imaging β -amyloid pathology. *JAMA - Journal of the American Medical Association*, 305(3), 275-283. doi:10.1001/jama.2010.2008

- Counts, S. E., Ikonomic, M. D., Mercado, N., Vega, I. E., & Mufson, E. J. (2017). Biomarkers for the early detection and progression of Alzheimer's disease. *Neurotherapeutics*, 14(1), 35-53. doi:10.1007/s13311-016-0481-z
- Creavin, S. T., Wisniewski, S., Noel-Storr, A. H., Trevelyan, C. M., Hampton, T., Rayment, D., . . . Cullum, S. (2016). Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database of Systematic Reviews*(1), CD011145. doi:10.1002/14651858.CD011145.pub2
- Cure, S., Abrams, K., Belger, M., dell'agnello, G., & Happich, M. (2014). Systematic literature review and meta-analysis of diagnostic test accuracy in Alzheimer's disease and other dementia using autopsy as standard of truth. *Journal of Alzheimer's Disease*, 42(1), 169-182. doi:10.3233/JAD-131559
- Dhiman, K., Gupta, V. B., Villemagne, V. L., Eratne, D., Graham, P. L., Fowler, C., . . . Martins, R. N. (2020). Cerebrospinal fluid neurofilament light concentration predicts brain atrophy and cognition in Alzheimer's disease. *Alzheimer's & Dementia (Amst)*, 12(1), e12005. doi:10.1002/dad2.12005
- Dill, V., Franco, A. R., & Pinho, M. S. (2015). Automated methods for hippocampus segmentation: The evolution and a review of the state of the art. *Neuroinformatics*, 13(2), 133-150. doi:10.1007/s12021-014-9243-4
- Duara, R., Barker, W., Loewenstein, D., Greig, M. T., Rodriguez, R., Goryawala, M., . . . Adjouadi, M. (2015). Insights into cognitive aging and Alzheimer's disease using amyloid PET and structural MRI scans. *Clinical and Translational Imaging*, 3(1), 65-74. doi:10.1007/s40336-015-0110-6
- Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., . . . Cummings, J. L. (2014). Advancing research diagnostic criteria for Alzheimer's

- disease: the IWG-2 criteria. *The Lancet Neurology*, 13(6), 614-629.
doi:10.1016/S1474-4422(14)70090-0
- Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., . . . Washington Dc, U. S. A. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 12(3), 292-323.
doi:10.1016/j.jalz.2016.02.002
- Dubois, B., Padovani, A., Scheltens, P., Rossi, A., & Dell'Agnello, G. (2016). Timely diagnosis for Alzheimer's disease: A literature review on benefits and challenges. *Journal of Alzheimer's Disease*, 49(3), 617-631. doi:10.3233/jad-150692
- Ebenau, J. L., Timmers, T., Wesselman, L. M. P., Verberk, I. M. W., Verfaillie, S. C. J., Slot, R. E. R., . . . van der Flier, W. M. (2020). ATN classification and clinical progression in subjective cognitive decline: The SCIENCe project. *Neurology*, 95(1), e46-e58.
doi: 10.1212/WNL.00000000000009724
- Edmonds, E. C., Delano-Wood, L., Clark, L. R., Jak, A. J., Nation, D. A., McDonald, C. R., . . . Alzheimer's Dis Neuroimaging, I. (2015). Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. *Alzheimers & Dementia*, 11(4), 415-424. doi:10.1016/j.jalz.2014.03.005
- Edmonds, E. C., McDonald, C. R., Marshall, A., Thomas, K. R., Eppig, J., Weigand, A. J., . . . Alzheimer's Disease Neuroimaging, I. (2019). Early versus late MCI: Improved MCI staging using a neuropsychological approach. *Alzheimers & Dementia*, 15(5), 699-708. doi:10.1016/j.jalz.2018.12.009
- Farid, K., Charidimou, A., & Baron, J. C. (2017). Amyloid positron emission tomography in sporadic cerebral amyloid angiopathy: A systematic critical update. *NeuroImage: Clinical*, 15, 247-263. doi:10.1016/j.nicl.2017.05.002

- Feldman, R., & Newman, J. (2013). Copyright at the bedside: Should we stop the spread? *Stanford Technology Law Review* 16(3), 623-655.
- Ferman, T. J., Smith, G. E., Kantarci, K., Boeve, B. F., Pankratz, V. S., Dickson, D. W., . . . Petersen, R. C. (2013). Nonamnesic mild cognitive impairment progresses to dementia with Lewy bodies. *Neurology*, 81(23), 2032-2038.
doi:10.1212/01.wnl.0000436942.55281.47
- Ferreira, D., Perestelo-Perez, L., Westman, E., Wahlund, L. O., Sarria, A., & Serrano-Aguilar, P. (2014). Meta-review of CSF core biomarkers in Alzheimer's disease: The state-of-the-art after the new revised diagnostic criteria. *Frontiers in Aging Neuroscience*, 6, 47. doi:10.3389/fnagi.2014.00047
- Filippi, L., Chiaravalloti, A., Bagni, O., & Schillaci, O. (2018). 18F-labeled radiopharmaceuticals for the molecular neuroimaging of amyloid plaques in Alzheimer's disease. *American journal of nuclear medicine and molecular imaging*, 8(4), 268-281.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189-198. doi:10.1016/0022-3956(75)90026-6
- Gross, A. L., Manly, J. J., Pa, J., Johnson, J. K., Park, L. Q., Mitchell, M. B., . . . McLaren, D. G. (2012). Cortical signatures of cognition and their relationship to Alzheimer's disease. *Brain Imaging and Behavior*, 6(4), 584-598. doi:10.1007/s11682-012-9180-5
- Hodges, J. R., & Larner, A. J. (2017). Addenbrooke's Cognitive Examinations: ACE, ACE-R, ACE-III, ACEapp, and M-ACE. In A. J. Larner (Ed.), *Cognitive Screening Instruments: A Practical Approach* (pp. 109-137). Cham: Springer International Publishing.

- Hsieh, S., Schubert, S., Hoon, C., Mioshi, E., & Hodges, J. R. (2013). Validation of the Addenbrooke's Cognitive Examination III in Frontotemporal dementia and Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, 36(3-4), 242-250. doi:10.1159/000351671
- Ikonomic, M. D., Klunk, W. E., Abrahamson, E. E., Mathis, C. A., Price, J. C., Tsopelas, N. D., . . . DeKosky, S. T. (2008). Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. *Brain*, 131(Pt 6), 1630-1645. doi:10.1093/brain/awn016
- Jack, C. R., Jr., Albert, M. S., Knopman, D. S., McKhann, G. M., Sperling, R. A., Carrillo, M. C., . . . Phelps, C. H. (2011). Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 257-262. doi:10.1016/j.jalz.2011.03.004
- Jack, C. R., Jr., Barkhof, F., Bernstein, M. A., Cantillon, M., Cole, P. E., Decarli, C., . . . Foster, N. L. (2011). Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. *Alzheimer's & Dementia*, 7(4), 474-485 e474. doi:10.1016/j.jalz.2011.04.007
- Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., . . . Sperling, R. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), 535-562. doi:10.1016/j.jalz.2018.02.018
- Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., . . . Dubois, B. (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, 87(5), 539-547. doi:10.1212/WNL.0000000000002923

- Jack, C. R., Jr., Wiste, H. J., Therneau, T. M., Weigand, S. D., Knopman, D. S., Mielke, M. M.,... Petersen, R. C. (2019). Associations of amyloid, tau, and neurodegeneration biomarker profiles with rates of memory decline among individuals without dementia. *JAMA*, 321(23), 2316-2325. doi: 10.1001/jama.2019.7437
- Janelidze, S., Mattsson, N., Palmqvist, S., Smith, R., Beach, T. G., Serrano, G. E., . . . Hansson, O. (2020). Plasma P-tau181 in Alzheimer's disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer's dementia. *Nature Medicine*, 26(3), 379-386. doi:10.1038/s41591-020-0755-1
- Karikari, T. K., Pascoal, T. A., Ashton, N. J., Janelidze, S., Benedet, A. L., Rodriguez, J. L., . . . Blennow, K. (2020). Blood phosphorylated tau 181 as a biomarker for Alzheimer's disease: a diagnostic performance and prediction modelling study using data from four prospective cohorts. *The Lancet Neurology*, 19(5), 422-433. doi:10.1016/s1474-4422(20)30071-5
- Klekociuk, S. Z., Saunders, N. L., & Summers, M. J. (2016). Diagnosing mild cognitive impairment as a precursor to dementia: Fact or fallacy? *Australian Psychologist*, 51(5), 366-373. doi:10.1111/ap.12178
- Klekociuk, S. Z., Summers, J. J., Vickers, J. C., & Summers, M. J. (2014). Reducing false positive diagnoses in mild cognitive impairment: The importance of comprehensive neuropsychological assessment. *European Journal of Neurology*, 21(10), 1330-+. doi:10.1111/ene.12488
- Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., . . . Langstrom, B. (2004). Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound-B. *Annals of Neurology*, 55(3), 306-319. doi:10.1002/ana.20009

- Koronyo-Hamaoui, M., Doustar, J., Oviatt, M., Black, K. L., & Koronyo, Y. (2020). Advances in Retinal Imaging: Retinal Amyloid Imaging. In *OCT and Imaging in Central Nervous System Diseases* (pp. 83-122).
- Larner, A. J., & Mitchell, A. J. (2014). A meta-analysis of the accuracy of the Addenbrooke's Cognitive Examination (ACE) and the Addenbrooke's Cognitive Examination-Revised (ACE-R) in the detection of dementia. *International Psychogeriatrics*, 26(4), 555-563. doi:10.1017/S1041610213002329
- Leeftang, M. M. G., Deeks, J. J., Gatsonis, C., & Bossuyt, P. M. M. (2008). Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine*, 149(12), 889. doi:10.1097/XEB.0000000000000061
- Leuzy, A., Smith, R., Ossenkoppele, R., Santillo, A., Borroni, E., Klein, G., . . . Hansson, O. (2020). Diagnostic Performance of RO948 F 18 Tau Positron Emission Tomography in the Differentiation of Alzheimer Disease From Other Neurodegenerative Disorders. *JAMA Neurology*. doi:10.1001/jamaneurol.2020.0989
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *British Medical Journal*, 339, b2700. doi:10.1136/bmj.b2700
- Lin, J. S., O'Connor, E., Rossom, R. C., Perdue, L. A., & Eckstrom, E. (2013). Screening for cognitive impairment in older adults: A systematic review for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, 159(9), 601-612. doi:10.7326/0003-4819-159-9-201311050-00730
- Liu, W., Lin, H., He, X., Chen, L., Dai, Y., Jia, W., . . . Chen, L. (2020). Neurogranin as a cognitive biomarker in cerebrospinal fluid and blood exosomes for Alzheimer's

- disease and mild cognitive impairment. *Translational Psychiatry*, 10(1), 125.
doi:10.1038/s41398-020-0801-2
- Lotjonen, J., Wolz, R., Koikkalainen, J., Julkunen, V., Thurfjell, L., Lundqvist, R., . . . Rueckert, D. (2011). Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease. *Neuroimage*, 56(1), 185-196.
doi:10.1016/j.neuroimage.2011.01.062
- McKhann, G. M., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7), 939-944. doi:10.1212/wnl.34.7.939
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to 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), 263-269. doi:10.1016/j.jalz.2011.03.005
- Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, 21(11), 1078-1085.
doi:10.1002/gps.1610
- Mitchell, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of Psychiatric Research*, 43(4), 411-431. doi:10.1016/j.jpsychires.2008.04.014
- Mo, J. A., Lim, J. H., Sul, A. R., Lee, M., Youn, Y. C., & Kim, H. J. (2015). Cerebrospinal fluid beta-amyloid 1-42 levels in the differential diagnosis of Alzheimer's disease -

- Systematic review and meta-analysis. *PLoS One*, 10(2).
doi:10.1371/journal.pone.0116802
- Molinuevo, J. L., Blennow, K., Dubois, B., Engelborghs, S., Lewczuk, P., Perret-Liaudet, A., . . . Parnetti, L. (2014). The clinical use of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis: A consensus paper from the Alzheimer's Biomarkers Standardization Initiative. *Alzheimer's & Dementia*, 10(6), 808-817.
doi:10.1016/j.jalz.2014.03.003
- Morris, E., Chalkidou, A., Hammers, A., Peacock, J., Summers, J., & Keevil, S. (2016). Diagnostic accuracy of 18F amyloid PET tracers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *European Journal of Nuclear Medicine and Molecular Imaging*, 43(2), 374-385. doi:10.1371/journal.pone.0116802
- Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695-699. doi:10.1111/j.1532-5415.2005.53221.x
- Nettiksimmons, J., DeCarli, C., Landau, S., Beckett, L., & Alzheimer's Disease Neuroimaging, I. (2014). Biological heterogeneity in ADNI amnesic mild cognitive impairment. *Alzheimers & Dementia*, 10(5), 511-521 e511.
doi:10.1016/j.jalz.2013.09.003
- Nunnally, J. C. (1978). *Psychometric theory* (2d ed. ed.): McGraw-Hill.
- Paddick, S. M., Gray, W. K., McGuire, J., Richardson, J., Dotchin, C., & Walker, R. W. (2017). Cognitive screening tools for identification of dementia in illiterate and low-educated older adults, a systematic review and meta-analysis. *International Psychogeriatrics*, 29(6), 897-929. doi:10.1017/S1041610216001976

- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3), 183-194. doi:10.1111/j.1365-2796.2004.01388.x
- Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., . . . Jack, C. R., Jr. (2009). Mild cognitive impairment: ten years later. *Archives of Neurology*, 66(12), 1447-1455. doi:10.1001/archneurol.2009.266
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment - Clinical characterization and outcome. *Archives of Neurology*, 56(3), 303-308. doi:10.1001/archneur.56.3.303
- Preische, O., Schultz, S. A., Apel, A., Kuhle, J., Kaeser, S. A., Barro, C., . . . Dominantly Inherited Alzheimer, N. (2019). Serum neurofilament dynamics predicts neurodegeneration and clinical progression in presymptomatic Alzheimer's disease. *Nature Medicine*, 25(2), 277-283. doi:10.1038/s41591-018-0304-3
- Ritchie, C., Smailagic, N., Noel-Storr, A. H., Takwoingi, Y., Flicker, L., Mason, S. E., & McShane, R. (2014). Plasma and cerebrospinal fluid amyloid beta for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*(6), CD008782. doi:10.1002/14651858.CD008782.pub4
- Rivero-Santana, A., Ferreira, D., Perestelo-Perez, L., Westman, E., Wahlund, L. O., Sarria, A., & Serrano-Aguilar, P. (2017). Cerebrospinal fluid biomarkers for the differential diagnosis between Alzheimer's disease and frontotemporal lobar degeneration: Systematic review, HSROC analysis, and confounding factors. *Journal of Alzheimer's Disease*, 55(2), 625-644. doi:10.3233/JAD-160366
- Sabri, O., Seibyl, J., Rowe, C., & Barthel, H. (2015). Beta-amyloid imaging with florbetaben. *Clinical and Translational Imaging*, 3(1), 13-26. doi:10.1007/s40336-015-0102-6

- Sarazin, M., Berr, C., De Rotrou, J., Fabrigoule, C., Pasquier, F., Legrain, S., . . . Dubois, B. (2007). Amnestic syndrome of the medial temporal type identifies prodromal AD: a longitudinal study. *Neurology*, 69(19), 1859-1867.
doi:10.1212/01.wnl.0000279336.36610.f7
- Schindler, S. E., Bollinger, J. G., Ovod, V., Mawuenyega, K. G., Li, Y., Gordon, B. A., . . . Bateman, R. J. (2019). High-precision plasma beta-amyloid 42/40 predicts current and future brain amyloidosis. *Neurology*, 93(17), e1647-e1659.
doi:10.1212/WNL.00000000000008081
- Sheikh-Bahaei, N., Sajjadi, S. A., Manavaki, R., & Gillard, J. H. (2017). Imaging biomarkers in Alzheimer's disease: A practical guide for clinicians. *Journal of Alzheimer's disease reports*, 1(1), 71-88. doi:10.3233/ADR-170013
- Shulman, K. I., Herrmann, N., Brodaty, H., Chiu, H., Lawlor, B., Ritchie, K., & Scanlan, J. M. (2006). IPA survey of brief cognitive screening instruments. *International Psychogeriatrics*, 18(2), 281-294. doi:10.1017/s1041610205002693
- Smailagic, N., Vacante, M., Hyde, C., Martin, S., Ukoumunne, O., & Sachpekidis, C. (2015). 18F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, 1(1), CD010632. doi:10.1002/14651858.CD010632.pub2
- Sperling, R., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., . . . Phelps, C. H. (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. doi:10.1016/j.jalz.2011.03.003

- Sperling, R., & Johnson, K. (2013). Biomarkers of Alzheimer disease: Current and future applications to diagnostic criteria. *Continuum*, 19(2 Dementia), 325-338.
doi:10.1212/01.CON.0000429181.60095.99
- Stott, J., Scior, K., Mandy, W., & Charlesworth, G. (2017). Dementia screening accuracy is robust to premorbid IQ variation: Evidence from the Addenbrooke's Cognitive Examination-III and the Test of Premorbid Function. *Journal of Alzheimer's Disease*, 57(4), 1293-1302. doi:10.3233/jad-161218
- Summers, M. J., & Saunders, N. L. (2012). Neuropsychological measures predict decline to Alzheimer's dementia from mild cognitive impairment. *Neuropsychology*, 26(4), 498-508. doi:10.1037/a0028576
- The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group. (1998). Consensus report of the Working Group on: "Molecular and biochemical markers of Alzheimer's disease". *Neurobiology of Aging*, 19(2), 109-116.
- Thijssen, E. H., La Joie, R., Wolf, A., Strom, A., Wang, P., Iaccarino, L., . . . Treatment for Frontotemporal Lobar Degeneration, i. (2020). Diagnostic value of plasma phosphorylated tau181 in Alzheimer's disease and frontotemporal lobar degeneration. *Nature Medicine*, 26(3), 387-397. doi:10.1038/s41591-020-0762-2
- Trzepacz, P. T., Yu, P., Sun, J., Schuh, K., Case, M., Witte, M. M., . . . Alzheimer's Disease Neuroimaging, I. (2014). Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to Alzheimer's dementia. *Neurobiology of Aging*, 35(1), 143-151. doi:10.1016/j.neurobiolaging.2013.06.018
- Tsoi, K. K. F., Chan, J. Y. C., Hirai, H. W., Wong, A., Mok, V. C. T., Lam, L. C. W., . . . Wong, S. Y. S. (2017). Recall tests are effective to detect mild cognitive impairment: A systematic review and meta-analysis of 108 diagnostic studies. *Journal of the*

American Medical Directors Association, 18(9), 807 e817-807 e829.

doi:10.1016/j.jamda.2017.05.016

Tsoi, K. K. F., Chan, J. Y. C., Hirai, H. W., Wong, S. Y. S., & Kwok, T. C. Y. (2015).

Cognitive tests to detect dementia a systematic review and meta-analysis. *JAMA*

Internal Medicine, 175(9), 1450-1458. doi:10.1001/jamainternmed.2015.2152

Vemuri, P., Knopman, D. S., Lesnick, T. G., Przybelski, S. A., Mielke, M. M., Graff-

Radford, J.,... Jack, C. R., Jr. (2017). Evaluation of amyloid protective factors and

Alzheimer disease neurodegeneration protective factors in elderly individuals. *JAMA*

Neurology, 74(6), 718-726. doi:10.1001/jamaneurol.2017.0244

Weissberger, G. H., Strong, J. V., Stefanidis, K. B., Summers, M. J., Bondi, M. W., &

Stricker, N. H. (2017). Diagnostic accuracy of memory measures in Alzheimer's

dementia and mild cognitive impairment: a systematic review and meta-analysis.

Neuropsychology Review, 27(4), 354-388. doi:10.1007/s11065-017-9360-6

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., . . .

Bossuyt, P. M. (2011). QUADAS-2: a revised tool for the quality assessment of

diagnostic accuracy studies. *Annals of Internal Medicine*, 155(8), 529-536.

doi:10.7326/0003-4819-155-8-201110180-00009

Zhang, S., Smailagic, N., Hyde, C., Noel-Storr, A. H., Takwoingi, Y., McShane, R., & Feng,

J. (2014). 11C-PIB-PET for the early diagnosis of Alzheimer's disease dementia and

other dementias in people with mild cognitive impairment (MCI). *Cochrane Database*

of Systematic Reviews. doi:10.1002/14651858.CD010386.pub2

Chapter 2

Morphology of the Collateral Sulcus

This chapter describes and quantifies the different variations of collateral sulcal patterns that could be observed along the base of the temporal lobe. This chapter arose during the process of manual delineation of the entorhinal and the transentorhinal cortices (to be covered in Chapters 3 and 4), during which it was observed that there were several patterns of sulcal folding of the collateral sulcus. The collateral sulcus is a major anatomical landmark as it delineates various regions of the medial temporal structures and variations complicated the segmentation process, but the complexity is not clearly addressed in the segmentation protocols. This exercise in identifying the patterns of collateral sulcal variability provided some elucidation to the validation results of the automated entorhinal and transentorhinal volumetric estimations in Chapter 3 and 4, and could hopefully pave the way for future refinement of manual and automated segmentation protocols

Abstract

The collateral sulcus (CS) is an important landmark for locating the entorhinal and transentorhinal cortices (EC-TEC), which are two of the earliest medial temporal structures affected in Alzheimer's disease. The morphology of the CS is known to be highly variable and this variability makes the neuroanatomical localisation of the EC-TEC a challenging task. This study investigated the rates and patterns of variability of the CS, with focus on the length of CS contained within the limits of the EC-TEC. The examination of 70 T1-weighted MRI scans acquired from healthy older adults demonstrated that the localisation of the EC-TEC could be disrupted in several ways. Although a discontinuous CS was a common form of disruption (26.9%), other common forms of disruption included a merger between two segments of the CS (26.9%) and the presence of side-branches (38.5%). These various patterns of disruption emphasise the need to consider the different rates and patterns of CS variability arising not only from CS continuity, but also from other sulcal mergers and bifurcations, during the development of manual and automated segmentation protocols in future studies.

Keywords: collateral sulcus, morphology, magnetic resonance imaging, sulcal pattern, sulcal variability

Morphology of the Collateral Sulcus

Structures in the medial temporal lobe (MTL) are essential for memory function and accurate localisation of these structures is critical for its potential use as a reliable biomarker in the clinical context. Alzheimer's disease is one of the most prevalent neurodegenerative diseases with detectable neuropathology in the MTL as one of its early hallmarks.

In recent years, coupled with advances in the understanding of the pathophysiological process of AD, there is broad consensus of the numerous advantages of incorporating biomarkers for the detection of incipient AD and the monitoring of its progression (Marilyn S. Albert et al., 2011; Bayram, Caldwell, & Banks, 2018; Dubois et al., 2010; McKhann et al., 2011; Sperling & Johnson, 2013). A major focus in AD research have been on mild cognitive impairment (MCI) due to AD. MCI is a condition in which there is evidence of cognitive decline with preservation of functional independence and may mark a transitional stage towards AD dementia (Marilyn S. Albert et al., 2011; Petersen et al., 1999). Several lines of research have investigated the use of MTL abnormalities in the detection of MCI and the prediction of conversion to AD dementia (Kulason et al., 2018; Meyer et al., 2013; Pennanen et al., 2004; Tapiola et al., 2008; Tward et al., 2017; Velayudhan et al., 2013). However, precise identification of some of these MTL structures have continued to be challenging for reasons that will be later discussed.

Involvement of the Medial Temporal Lobe in Alzheimer's Disease

One of the earliest biomarker features implicated in AD is atrophy in the MTL. Prominent regions in the MTL implicated early in AD include the hippocampus, entorhinal and transentorhinal cortex, structures that are recognised to be critical regions for memory functions (Braak, Braak, & Bohl, 1993). Volumetric changes in the hippocampus can be observed readily on structural magnetic resonance images (MRI) and a multitude of studies have consistently demonstrated the association of hippocampal volumetric reductions with

inefficiencies in memory function (e.g. Anstey et al., 2013; Bonner-Jackson, Mahmoud, Miller, & Banks, 2015; Chételat et al., 2003; Deweer et al., 1995; Mueller et al., 2010; Peter et al., 2016; Vyhnalek et al., 2014).

However, the hallmark of intracellular neurofibrillary tangles (NFT) that is associated with AD pathology does not start in the hippocampus. The distribution of NFT shows a very characteristic pattern of development in the cortex. The NFT first develops in the transentorhinal cortex, followed by the entorhinal cortex, the hippocampus and then the rest of the neocortex (Braak & Braak, 1991, 1995). This distribution of NFT suggests that volumetric quantification of the entorhinal and transentorhinal cortices could potentially play a critical role in the early detection of AD.

Several studies have examined the role of entorhinal cortex in AD. Autopsy studies have demonstrated that there were significant neuronal losses in the entorhinal cortex of individuals with mild AD (Kordower et al., 2001; J. L. Price et al., 2001). There is also an increasing number of research studies based on MRI-derived volumetric measurements that have suggested that entorhinal volume loss precedes hippocampal volume loss in mild cognitive impairment (e.g. Pennanen et al., 2004; Tapiola et al., 2008). Others have also investigated and found a positive relationship between reduced entorhinal cortex volumes and memory deficits in people with or without AD (e.g. Li et al., 2012; Maass et al., 2018; Olsen et al., 2017; C. C. Price et al., 2010).

In a review of the research that used data from the Alzheimer's Disease Neuroimaging Initiative, the involvement of the hippocampus and entorhinal cortex in MCI and AD was discussed, but not the transentorhinal cortex (Weiner et al., 2015). Therefore, against the background of increasing consensus regarding the role of anatomical changes in the hippocampus and entorhinal cortex in detecting AD, the role of the other critical MTL structure - the transentorhinal cortex – is less clear.

Specific research on the transentorhinal cortex is complicated by its nomenclature. The term “transentorhinal” cortex was first coined by Braak and Braak (1985) and referred to the cortical region adjoining the entorhinal cortex. The transentorhinal cortex, which correspond to Brodmann’s area 35, serves as a complex zone of transition between the neocortex and the entorhinal cortex (Brodmann & Garey, 2006). Initially, Brodmann area 35 was synonymous with the perirhinal cortex (Brodmann & Garey, 2006). However, many studies have combined Brodmann area 35 with the neighbouring Brodmann area 36 and referred to these two structures collectively as the perirhinal cortex (Ding & Van Hoesen, 2010; Insausti et al., 1998). Others have combined the entorhinal cortex with the transentorhinal cortex and referred to them collectively as the rhinal cortex (Meunier, Hadfield, Bachevalier, & Murray, 1996).

Despite the inconsistent terminology, recent studies have turned their attention to changes in transentorhinal cortex for detecting AD. Cross-sectional studies have shown that the pattern of MTL atrophy corresponded to the pattern of NFT deposition described by Braak and Braak (1991) and the volumes of the transentorhinal cortex in people with preclinical or very early AD were significantly smaller than control participants (Wolk et al., 2017; Yushkevich, Pluta, et al., 2015). However, these studies have acknowledged that a limitation of their segmentation methods was the variability of the collateral sulcus. Longitudinally, two studies for the Alzheimer’s Disease Neuroimaging Initiative detected significant atrophy in the transentorhinal cortex in people with MCI, however the authors had only examined the left hemispheres and excluded subjects with variants in their collateral sulcus (Kulason et al., 2018; Tward et al., 2017).

Variability of the Collateral Sulcus

The nomenclature “rhinal sulcus” and “collateral sulcus” have been used ambiguously in the literature. Some studies defined the entire length of the collateral sulcus as running

from the anterior portion of the medial temporal lobe into the occipital lobe (Duvernoy, 1999; Insausti & Amaral, 2012), while others referred to the anterior portion of the collateral sulcus that binds the uncus as the rhinal sulcus (Gloor, 1997; Huntgeburth & Petrides, 2012; Ono, Kubik, & Abernathey, 1990; Ovalioglu, Ovalioglu, Canaz, & Emel, 2018). Although some might favour the latter definition, as it delimits the entorhinal cortex and transentorhinal cortex, this term originated from primate studies in which the location of the entorhinal and transentorhinal cortices coincided with the rhinal sulcus (Amaral, Insausti, & Cowan, 1987; Insausti & Amaral, 2012). However, in the human brain, there is no sulcus that easily identifies the limits of the entorhinal and transentorhinal cortices. In the human brain, the term “rhinal sulcus” is limited to the incipient indentation located at the ventromedial surface of the temporal pole, a sulcus that typically lies anterior to the entorhinal and transentorhinal cortices (Duvernoy, 1999; Insausti & Amaral, 2012; Insausti, Tunon, Sobreviela, Insausti, & Gonzalo, 1995).

For the most part, studies that focused on the localisation of the entorhinal and transentorhinal cortices have strongly favoured the use of the term “collateral sulcus” to refer to the main medial sulcus that runs on the inferior surface of the temporal and occipital lobes (Feczko, Augustinack, Fischl, & Dickerson, 2009; Franko, Insausti, Artacho-Perula, Insausti, & Chavoix, 2014; Insausti et al., 1998; Kivisaari, Probst, & Taylor, 2013; Pruessner et al., 2002). The collateral sulcus, in this definition, is a major anatomical landmark for defining the borders of the entorhinal and transentorhinal cortices. However, the collateral sulcus is a highly variable structure, which makes it a challenge to reliably determine the precise anatomical localisation of these two structures. Often there are interruptions and different cortical folding along its length, occurring between individuals as well as between hemispheres within the same individual (Ovalioglu et al., 2018; Pruessner et al., 2002; Taylor & Probst, 2008).

Several studies have examined the variability of the collateral sulcus along its entire length. In one of the earliest publications, Ono et al. (1990) found that in 72% of the 50 post-mortem hemispheres examined, the collateral sulcus was discontinuous, such that the anterior segment of the collateral sulcus not being connected to the posterior segment of the collateral sulcus. The authors also observed that in some of the cases with a discontinuous collateral sulcus, the two sections of the collateral sulcus would overlap as the anterior collateral sulcus terminated after the anterior portion of the posterior collateral sulcus (Ono et al., 1990). However, in other cases, the anterior collateral sulcus would terminate prior to the start of the posterior collateral sulcus (Ono et al., 1990). Other studies have reported similar observations of overlaps and non-overlaps between the anterior collateral sulcus and posterior collateral sulcus, and frequencies of these observations ranged from 31% - 46% and 17% - 36% respectively (Chau, Stewart, & Gragnaniello, 2014; Ding & Van Hoesen, 2010; Feczko et al., 2009; Novak et al., 2002; Ono et al., 1990).

Ono and his colleagues (1990) also observed that the anterior collateral sulcus connected to the occipitotemporal cortex in some rare instances. This connection between the anterior collateral sulcus and the occipitotemporal cortex have been observed and reported with similarly low frequencies in other studies (Chau et al., 2014; Novak et al., 2002; Ovalioglu et al., 2018). In addition to connections to the occipitotemporal sulcus, Ono et al. (1990) also identified that in 44% of the cases, side-branches, namely, sulci that enter the inner surface of the main gyri to form V-shaped convolutions, could be observed to extend from the anterior collateral sulcus. The presence of side-branches from the anterior collateral sulcus have also been observed by Cikla et al. (2016), but compared to Ono et al. (1990), they reported a lower frequency of 27.1%.

In the observations for a continuous collateral sulcus by Ono et al. (1990), they report a similar distribution across both hemispheres, 24% for right and 28% for the left. In contrast

to these observations, more recent studies reported higher rates of continuity, ranging from 34% to 50% in brains without significant pathology (Chau et al., 2014; Cikla et al., 2016; Ding & Van Hoesen, 2010; Feczko et al., 2009; Huntgeburth & Petrides, 2012; Kim, Bernasconi, Bernhardt, Colliot, & Bernasconi, 2008). Other studies have also broken down the continuity of the collateral sulcus more finely, noting that the anterior and posterior segments of the collateral sulcus are connected, but in a contiguous manner. Unlike the end-to-end connection seen in a continuous collateral sulcus, a contiguous collateral sulcus was observed and characterised by the posterior segment of the collateral sulcus ?emerging from one of the banks of the anterior collateral sulcus (Cikla et al., 2016; Huntgeburth & Petrides, 2012).

Some studies have also indicated that frequencies of a continuous collateral sulcus differed between subjects with and without neuropathology. For example, one study reported that a continuous collateral sulcus was observed in 30% of the left hemispheres of healthy control cases compared to 57% of the left hemispheres of people with AD (Feczko et al., 2009). In another study, continuity was observed in 44% of healthy control cases versus 75% in people with temporal lobe epilepsy (Kim et al., 2008).

Localisation of the Entorhinal and Transentorhinal Cortices

Medially, the entorhinal cortex occupies the gyrus ambiens or the dorsomedial bank of the parahippocampal gyrus (Insausti et al., 1998; Insausti et al., 1995; Kivisaari et al., 2013). Laterally, the entorhinal cortex shares a border with the transentorhinal cortex. The location of this border is dependent on the depth of the collateral sulcus and varies from the medial edge to the fundus of the collateral sulcus (Insausti & Amaral, 2012; Krimer, Hyde, Herman, & Saunders, 1997). The lateral border of the transentorhinal cortex also changes according to the shape and depth of the collateral sulcus (Taylor & Probst, 2008). The transentorhinal cortex typically lies on the medial bank of the collateral sulcus; however,

when the collateral sulcus is shallow, the transentorhinal cortex is found on both the medial and the lateral banks (Augustinack et al., 2013; Ding & Van Hoesen, 2010; Taylor & Probst, 2008; Van Hoesen, Augustinack, Dierking, Redman, & Thangavel, 2000). Typically, this change of collateral sulcus depth occurs towards the end of a discontinuous collateral sulcus or with sulcal bifurcation (Augustinack et al., 2013; Taylor & Probst, 2008).

A common method of quantifying MTL neuronal loss associated with AD is using *in vivo* MRI-based volumetry. Numerous manual delineation protocols have been developed to quantify the entorhinal cortical volumes using structural MRI scans (Goncharova, Dickerson, Stoub, & deToledo-Morrell, 2001; Honeycutt et al., 1998; Insausti et al., 1998; Meade, Bowden, Whelan, & Cook, 2008; Pruessner et al., 2002). A smaller number of manual segmentation protocols for the transentorhinal cortex have also been developed (Berron et al., 2017; Kivisaari et al., 2013). With advancements in high-resolution imaging and refinements in automated segmentation techniques and capabilities, a number of semi-automated and automated segmentation protocols for the entorhinal and transentorhinal cortices are available (Augustinack et al., 2013; Feczko et al., 2009; Xie et al., 2017; Yushkevich, Pluta, et al., 2015).

However, due to the highly variable nature of the collateral sulcus, the definitions of the borders of the entorhinal and transentorhinal cortices varies substantially between the different segmentation protocols. The impact of the heterogeneity of segmentation protocols could be seen in the wide range of entorhinal volumes that have been reported in previous studies. One study by C. C. Price et al. (2010) compared the entorhinal volumes of 39 healthy individuals (mean age = 71), estimated using three different published entorhinal manual segmentation protocols. The authors reported the mean left and right entorhinal volumes in mm³ to be 1350 and 1351 respectively using the protocol by Honeycutt et al. (1998), 971 and 1021 respectively using the protocol by Insausti et al. (1998), and 760 and 804 respectively

using the protocol by Goncharova et al. (2001). The absence of convergence in volumes suggested that these protocols were not interchangeable, which the authors suggested was partly an effect of the lack of agreement between protocols in addressing the variability of collateral sulcal depth (C. C. Price et al., 2010).

Aims of Current Study

Despite the importance of identifying the collateral sulcus and accounting for its variability, there are no widely accepted guidelines to account for the neuroanatomical variations of the collateral sulcus. The protocol for localising the entorhinal and transentorhinal cortices is dependent on macroscopic anatomical landmarks, and any form of discontinuity or bifurcation in the collateral sulcus, particularly occurring on its medial bank, is likely to impede reliable localisation of the MTL structures.

Several studies have discussed the variability of the collateral sulcus in its entirety. However, it is important to note that at the time of this writing, there is no literature found that specifically discussed the patterns of variability in relation to the location of the entorhinal and transentorhinal cortices. Therefore, it remains unclear whether the patterns and rates of variability would be similar between the entire length of the collateral sulcus and along the length of the entorhinal and transentorhinal cortices.

To clarify the rates and patterns of variability, this study aims to identify and classify the anatomical variants of the collateral sulcus, with specific focus on its relationship to the nominated locations of the entorhinal and transentorhinal cortices. Through building a coherent picture of variability and interruptions within this specific segment of collateral sulcus, such study may help to guide the definition of these structures in future segmentation protocols.

Method

Subjects

The data used in this study were retrieved from the Australian Imaging Biomarkers and Lifestyle study of aging database (AIBL; aibl.csiro.au). The AIBL study, launched in November 2006 in Australia, was designed to identify biomarkers, cognitive features and health and lifestyle aspects that may be related to the development of symptomatic AD.

For this study, 70 3T structural MRI scans (voxel size 1 x 1 x 1 mm³) were retrieved from the AIBL database. The age and gender of the participants from whom the scans were acquired from were not provided. These scans were part of a “super-normal” control cohort, defined as participants who was assessed to have low retention of amyloid beta on positron emission tomography and presented with no cognitive deficits on neuropsychological assessment. The structural MRI scans, biomarker tests and neuropsychological assessments were undertaken as part of the AIBL study, and its study methodology have been described in a previous paper (Ellis et al., 2009).

Image Processing

All scans were processed through the openly available FreeSurfer software (version 6.0, Martinos Center for Biomedical Imaging, Harvard-MIT, Boston, MA). The fully automated cortical reconstruction, parcellation and labelling process was performed using the standard batched script, “recon-all” (Fischl et al., 2002; Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, & Dale, 1999). No post-processing of output was required.

Following completed processing, each scan was independently examined through the view of the three-dimensional pial surface generated by FreeSurfer, along with its matched two-dimensional coronal slices. This examination enabled the simultaneous visualisation of the relationship between the anatomical features visible on the reconstructed cortical surface and the native scans.

Locating the Anterior and Posterior Segment of the Collateral Sulcus

The three main sulci on the inferior surface (1) anterior collateral sulcus (ACS), (2) posterior collateral sulcus (PCS) and (3) occipitotemporal sulcus (OTS), were identified and labelled with different colours for each sulcus onto a graphical outline of the different sulcal connection (for an example, see Figure 2.1). The variation of the OTS and its relationship to the anterior and posterior collateral sulcus was not fully analysed in this study but its relationship with the ACS will be discussed briefly in the results section.

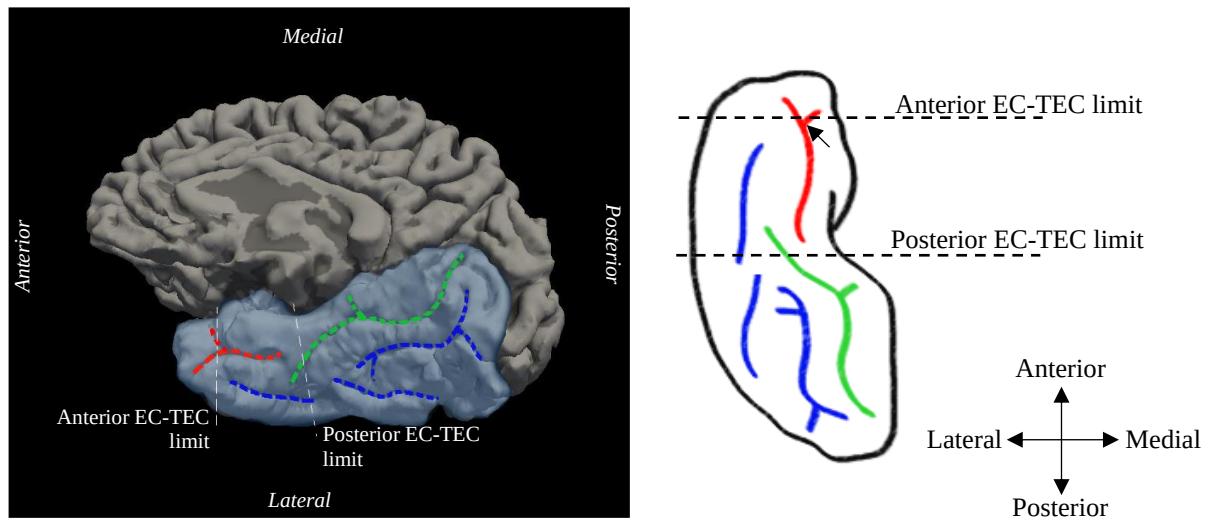
The specific section of collateral sulcus (CS) that was of interest was determined by the anterior and posterior limit of the entorhinal and transentorhinal cortices (EC-TEC) and the limits were indicated accordingly on the graphical representation of the sulcal patterns (Figure 2.1). Anatomical landmarks identified in cytoarchitectural studies that are useful for defining the rostral and caudal limit of the EC-TEC are the limen insulae and intralimbic gyrus (Insausti et al., 1998; Insausti et al., 1995). The most rostral section of the EC-TEC lies approximately 2 mm posterior to appearance of the limen insulae, while the most caudal section is usually located 1.1 mm posterior to the intralimbic gyrus (Insausti et al., 1998; Insausti et al., 1995). As seen on the coronal plane on a $1 \times 1 \times 1$ mm isotropic MRI scan, aligned perpendicular to the hippocampal axis, the anterior limit of the collateral sulcus of interest was therefore defined as two slices (2 mm) posterior to the limen insulae and the posterior limit was defined as one slice (1 mm) posterior to the level of the last visible slice of intra-limbic gyrus on the coronal plane. The sulcus is determined to be present when a sulcal depth of more than 1 mm is observed on the coronal slice.

Classification of the Patterns of Collateral Sulcus

The first main classification was the continuity of the entire length of CS based on the sulcal patterns observed on the reconstructed inferior cortical surface of the temporal lobe. In a discontinuous CS, the ACS is separated from the PCS by a small bridge of unconnected

Figure 2.1

Graphical Representation of the Sulcal Patterns at the Temporal Lobe Surface



Note. The left image is an inferomedial view of the right temporal cortex. The blue shaded area indicates the temporal cortex with the sulci of interests marked out. The image to the right shows how the sulcal patterns observed at the base of this temporal lobe surface, as seen on the left, would be represented graphically. For both images, the red line represents the anterior collateral sulcus (ACS), green line represents the posterior collateral sulcus (PCS) and blue line represents the occipitotemporal sulcus (OTS). The anterior and posterior limit of the entorhinal and transentorhinal cortices (EC-TEC) are marked and labelled by dotted lines. Any double sulcus seen on the coronal slices that was observed to disrupt the EC-TEC was marked with an arrow and the source of that disruption was later categorised.

cortex on the inferior cortical surface of the temporal lobe. In some discontinuous cases, the ACS terminated before it arrived at the coronal level where the PCS was present, as observed by the absence of a CS on the coronal slices. When the ACS was clearly separate from the PCS on the cortical surface, it was noted whether the ACS terminated before or after the posterior limit of the EC-TEC. There were two patterns in which the CS would not be categorised as discontinuous. In one pattern, the ACS was joined to the PCS via an end-to-end connection to form a continuous CS. In this pattern, the ACS was identified as the length of CS that was at and rostral to the posterior limits of the EC-TEC. In another pattern, the medial or lateral bank of the ACS would be joined to the end of the PCS, and this pattern of connection was categorised as a side-to-end contiguous connection. In addition to the continuity of the CS, the sulci that extended from the main ACS but did not form part of the PCS or OTS were noted and defined as side-branches.

The second main group of categorisations was based on EC-TEC interruptions. The EC-TEC was evaluated using both the reconstructed cortical surface and MRI coronal native scans, and then classified as uninterrupted or interrupted. An interruption would occur when there was a distinct change in the shared border of EC-TEC along the shoulder or fundus of the medial bank of the CS from one coronal slice to the next. The source of interruption was then categorised to originate from sulcal mergers between the ACS and PCS, discontinuous ACS and PCS, presence of side-branches or connection to the OTS.

Statistical Analyses

Statistical analysis was conducted using SPSS (version 25). Chi-square tests were used to examine whether there is presence of any hemispheric differences in the variations in sulcal patterns.

Results

Sulcal Patterns of the Collateral Sulcus

The entire length of the CS was examined regardless of the location of the EC-TEC. As shown in Table 2.1, a common pattern that was observed was a clear separation between the ACS and PCS, that is, a discontinuous collateral sulcus. When both hemisphere were examined, in 29.3% of cases ACS terminate after the anterior segment of PCS with a slight overlap (Pattern 1a: see Figure 2.2), and 20.7% of cases ACS would terminate before the anterior segment of PCS with no overlap between the two sulci (Pattern 1b: see Figure 2.2).

The second most common observation, Pattern 2, was a continuous CS, in which ACS appeared to be connected to PCS end-to-end (32.1%). Finally, Pattern 3 was the occurrence of a contiguous ACS and PCS via side-to-end connections (17.9%). A chi-square test showed that there were no hemispheric differences in the types of connections between the ACS and PCS, $\chi^2(2, N = 140) = 1.26, p = .53$. The types of connections between the anterior and posterior segments are illustrated in Figure 2.2. All observed connections between the ACS and PCS in these scans could be categorised under Pattern 1, 2 or 3. No other patterns were noted and there was no incidence in which the ACS or the PCS was absent.

Figure 2.3 shows the types of side-branches that were seen along the ACS. As shown in Table 2.1, observations from both hemispheres showed that most commonly, no side-branches were present along the ACS (67.9%). However, when side-branches were present, they most frequently appeared either towards the anterior portion of the ACS (20.0%) or the posterior end of the ACS (10.0%). Less frequently, the side-branches appeared on both the anterior and posterior end of the ACS (2.1%; Table 2.1). Occasionally, ACS made a connection with the OTS (2.9%; Figure 2.3). Chi-square tests did not demonstrate any hemispheric differences in the types of branching patterns in the ACS, $\chi^2(4, N = 140) = 1.50, p = .83$.

Table 2.1

Types of Connection between Anterior and Posterior Collateral Sulcus, and Presence of Side-branches Along Anterior Collateral Sulcus

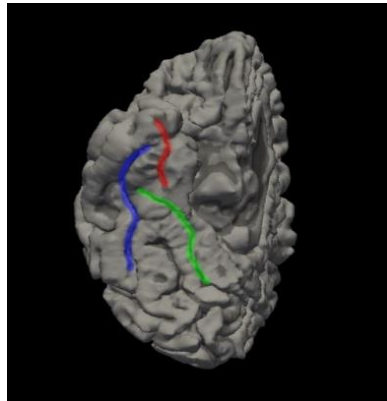
	<u>Hemisphere</u>		
	<u>Left</u> <i>n</i> (%)	<u>Right</u> <i>n</i> (%)	<u>Total</u> <i>n</i> (%)
Connection between ACS and PCS			
Pattern 1a: No connection with overlap	17 (24.3)	24 (34.3)	41 (29.3)
Pattern 1b: No connection without overlap	17 (24.3)	12 (17.1)	29 (20.7)
Pattern 2: End-to-end connection	21 (30.0)	24 (34.3)	45 (32.1)
Pattern 3: Side-to-end connection	15 (21.4)	10 (14.3)	25 (17.9)
Side-branches in ACS			
No side-branches	48 (68.6)	47 (67.1)	95 (67.9)
Anterior side branch	14 (20.0)	14 (20.0)	28 (20.0)
Posterior side branch	6 (8.6)	8 (11.4)	14 (10.0)
Anterior and posterior side-branches	2 (2.9)	1 (1.4)	3 (2.1)

Note. A total of 140 hemispheres were examined, left = 70, right = 70. ACS: anterior collateral sulcus, PCS: posterior collateral sulcus

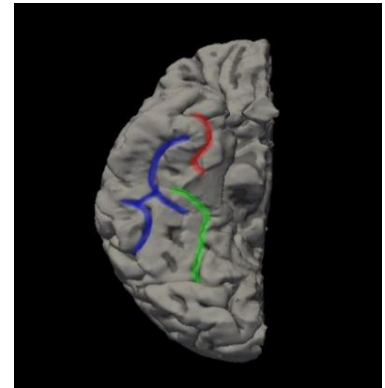
Figure 2.2

Types of Connections Between the Anterior and Posterior Collateral Sulci

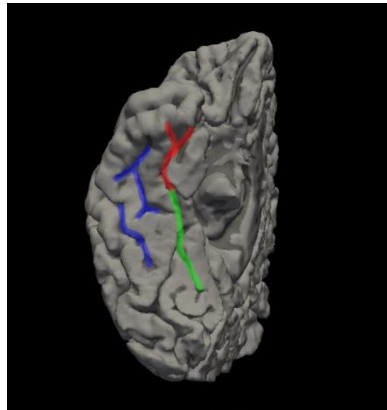
Pattern 1a



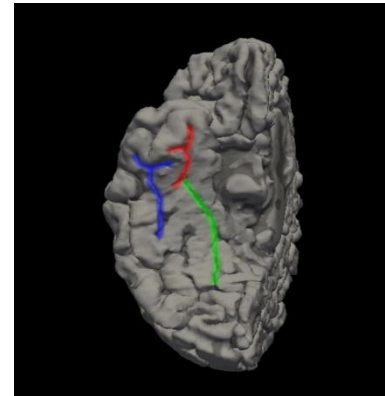
Pattern 1b



Pattern 2



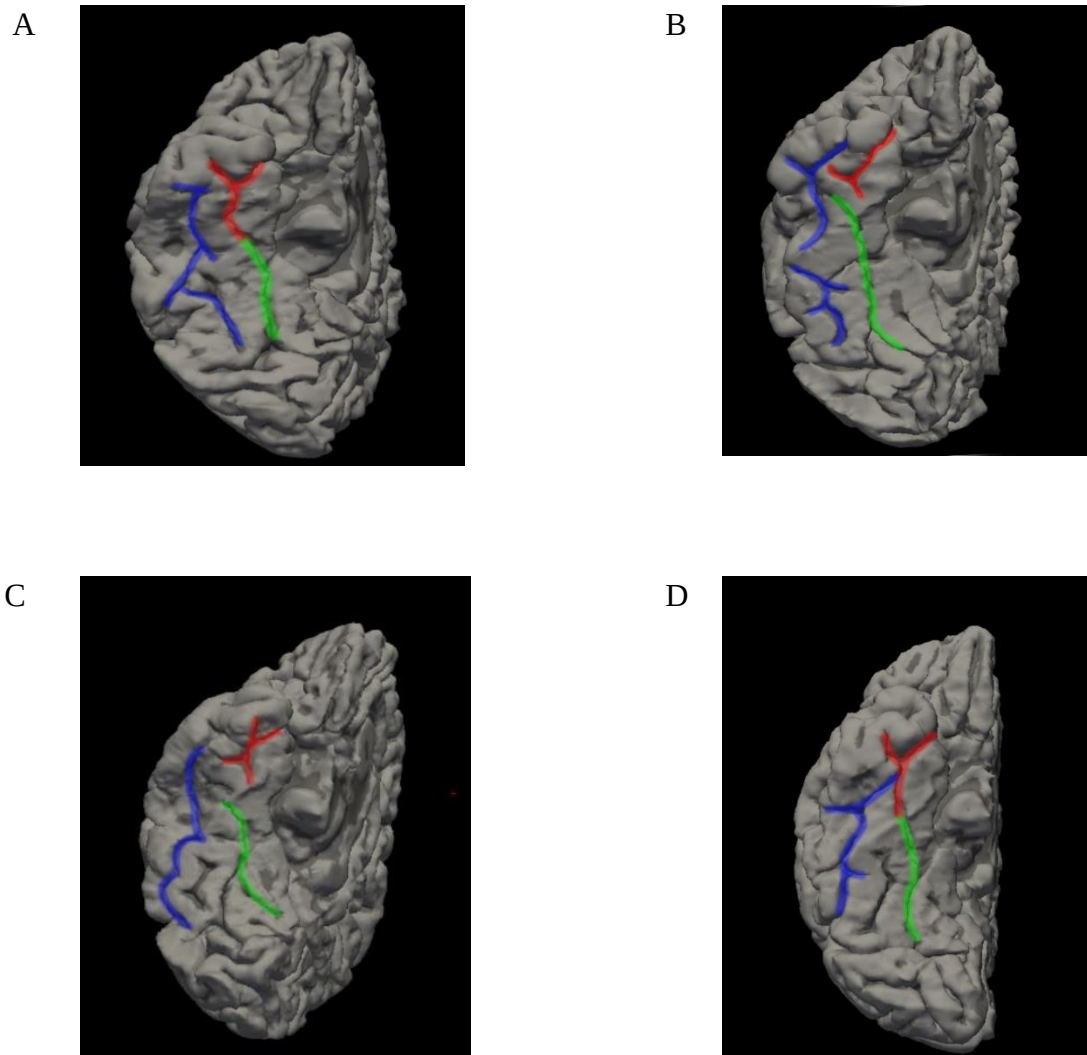
Pattern 3



Note. These four figures represent the different ways that the anterior and posterior collateral sulcus may be connected to each other. Red line represents the anterior collateral sulcus (ACS), green line represents the posterior collateral sulcus (PCS), blue line represents the occipitotemporal sulcus. Pattern 1a: no connection, overlap between ACS and PCS, Pattern 1b: no connection, no overlap between ACS and PCS, Pattern 2: end-to-end connections between ACS and PCS, Pattern 3: side-to-end connection between ACS and PCS.

Figure 2.3

Different Branching Patterns and Connections of the Anterior Collateral Sulcus



Note. These four figures show the common observations on the anterior collateral sulcus. Red line represents the anterior collateral sulcus (ACS), green line represents the posterior collateral sulcus (PCS), blue line represents the occipitotemporal sulcus (OTS). Besides having no side-branches along the ACS (not shown), the most common observations are (A) side-branches on the anterior end of the ACS, (B) side-branches on the posterior end of the ACS, or (C) side-branches on both the anterior and posterior end of the ACS. Occasionally, ACS and OTS makes a connection (D).

Interruptions of the Entorhinal and Transentorhinal Cortices

In 62.9% of the cases, EC-TEC was defined as uninterrupted (Type 1). In some cases, the ACS terminated before the posterior limit of the EC-TEC, but the PCS was absent on the coronal slices. The frequencies of such occurrences would be discussed later in this results section. However, in these cases, the EC-TEC was categorised as uninterrupted Type 1 as there was no neuroanatomical indication for a change in the location of the EC-TEC.

As shown in Figure 2.4, The EC-TEC could be interrupted in three ways due to the different folding patterns of the CS, in which the folding patterns led to the presence of double sulci on the coronal scans. Detailed frequencies of the types of interruption to the EC-TEC can be found in Table 2.2. Examination of both hemispheres from the anterior to the posterior direction revealed that 14.3% of the cases, the medial CS conflated with a lateral sulcus to eventually form a single CS (Type 2a). Another type of interruption that appeared in 20.7% of the cases was when the medial sulcus became progressively shallower and eventually disappeared but did not conflate with a lateral sulcus (Type 2b). In Type 2b, at the point of the disappearance of the original medial sulcus, the lateral sulcus running alongside it became the most medial sulcus where the nominated location of the EC-TEC. Finally, less frequently seen was the appearance of an additional sulcus developing from the medial bank of the ACS within the limits of the EC-TEC (2.1%; Type 3).

Examination of all the 52 hemispheres with interrupted EC-TECs revealed that the most common source of interruption was the presence of a side-branch (38.5%). In 17 of the 20 cases, the side-branches were located at the anterior end of the ACS. Other variations that led to interruptions to the EC-TECs included the point of merger between the ACS and PCS (26.9%), and an overlap between the disappearance of the ACS and the appearance of the PCS within the limits of the EC-TEC (26.9%). Another occurrence of a double sulci that led to interruptions of the EC-TECs was the occurrence of a connection between the ACS and the

Table 2.2

Patterns of Interruption to the Entorhinal and Transentorhinal Cortices and Its Relationship to the Collateral Sulcus

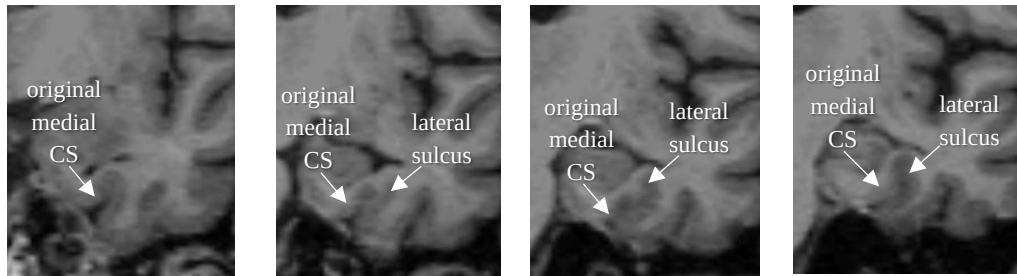
	<u>Hemisphere</u>		
	<u>Left</u> <i>n</i> (%)	<u>Right</u> <i>n</i> (%)	<u>Total</u> <i>n</i> (%)
Patterns of interruptions observed in the EC-TEC			
Uninterrupted Type 1	44 (62.9)	44 (62.9)	88 (62.9)
Interrupted Type 2a: Lateral merger of sulci within the depths of the CS to form a single, deep medial sulcus	12 (17.1)	8 (11.4)	20 (14.3)
Interrupted Type 2b: Lateral skip to a new medial sulcus after former medial CS flattens out	13 (18.6)	16 (22.9)	29 (20.7)
Interrupted Type 3: Medial development of a new more medial sulcus	1 (1.4)	2 (2.9)	3 (2.1)
Source of interruptions along the EC-TEC			
Presence of side-branches	10 (38.5)	10 (38.5)	20 (38.5)
Sulcal merger of ACS and PCS	6 (23.1)	8 (30.8)	14 (26.9)
Unconnected ACS and PCS	7 (26.9)	7 (26.9)	14 (26.9)
Connection to occipitotemporal sulcus	3 (11.5)	1 (3.8)	4 (7.7)
Most medial sulcus at posterior limit of EC-TEC			
ACS	51 (72.9)	53 (75.7)	104 (74.3)
PCS	14 (20.0)	12 (17.1)	26 (18.6)
Both ACS and PCS absent	5 (7.1)	5 (7.1)	10 (7.1)

Note. A total of 140 hemispheres were examined, left = 70, right = 70. EC-TEC: entorhinal cortex and transentorhinal cortex, ACS: anterior collateral sulcus, PCS: posterior collateral sulcus

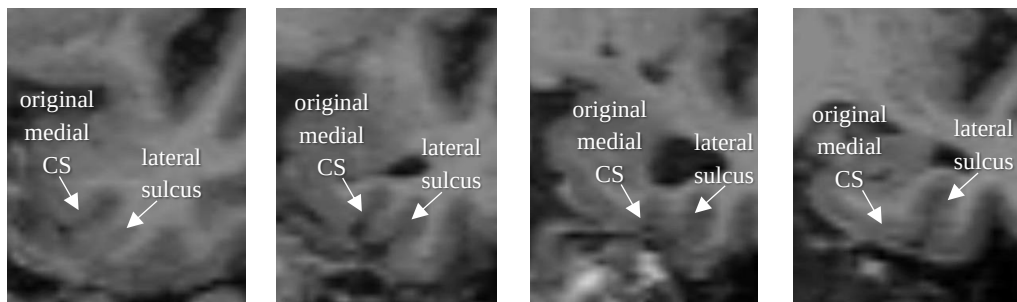
Figure 2.4

Types of Interruptions to the Entorhinal and Transentorhinal Cortices

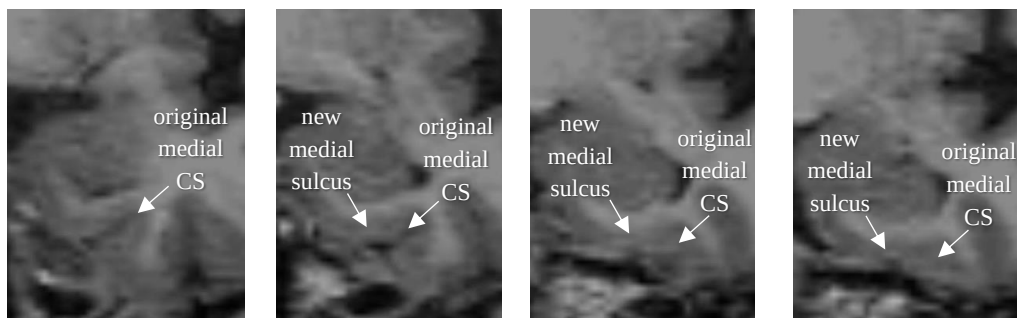
Type 2a



Type 2b



Type 3



Note. These figures illustrate the different types of interruptions to the entorhinal cortex and transentorhinal cortex (EC-TEC) resulting from different folding in the collateral sulcus (CS). Coronal slices for each type are shown from the anterior to posterior (left to right), within the limits of the EC-TEC. Besides Type 1, which is a single, deep CS from the anterior to posterior limits of the EC-TEC (not shown), there are three different types of interruptions. Type 2a: medial CS merges with a lateral sulcus to form one single deep CS, Type 2b: medial CS gradually flattens out and disappears without merging into the lateral sulcus, in which case the more lateral sulcus now becomes the new medial CS, Type 3: an additional sulcus develops out of the medial bank of the original medial CS to form a new medial CS.

OTS, but this was the least common observation (7.7%). Breakdown of the frequency of occurrence for the types of interruptions and its variations can be found in Table 2.2.

In most of the cases, the ACS was the most medial sulcus seen at the posterior limit of the EC-TEC (74.3%). In 18.6% of the cases, the PCS was the most medial sulcus observed at the posterior limit of the EC-TEC. However, when the CS was discontinuous, and the anterior and posterior segments had no overlap (Pattern 1b), occasionally the ACS terminated before the posterior limit of the EC-TEC and no CS was visible on the coronal slices (7.1%). There were no instances in which the ACS was absent in the anterior limit of the EC-TEC. Breakdown of the frequency of the most medial sulcus at the posterior limit of the EC-TEC can be found in Table 2.2.

Chi-square tests revealed neither hemispheric differences in an interrupted or uninterrupted EC-TEC, $\chi^2(1, N = 140) = 0, p = 1.00$, nor hemispheric differences in the types of interruptions (Uninterrupted: Type 1, Interrupted: Type 2a, Type 2b, Type 3), $\chi^2(3, N = 140) = 1.44, p = .70$. There were no hemisphere differences in the sources for the interruptions to an interrupted EC-TEC due to side-branches, sulcal merger, discontinuous CS or connection to OTS, $\chi^2(3, N = 52) = 1.29, p = .73$. There were also no hemispheric differences in which was the most medial sulcus at the posterior limit, be it ACS, PCS, or cases in which neither ACS nor PCS was present, $\chi^2(2, N = 140) = .19, p = .91$.

Discussion

The CS is a major landmark in determining the positions of the entorhinal and transentorhinal cortices. An increasing number of neuroimaging studies have been quantifying the different structures in the MTL, including using volume and cortical thickness measures. Amongst these structures, recent interest has been placed more heavily on the EC-TEC, given their early neuropathological involvement in Alzheimer's disease. Most of the available imaging studies rely on using the CS to determine the borders of the

EC-TEC. However, the CS is known to be a highly variable structure and this variability poses to be one of the main challenges to determining the neuroanatomical location of the EC-TEC.

This study investigated the variability of the CS, detailing its discontinuity between the anterior and posterior segments of the collateral sulcus, as well as the presence of side-branches in the ACS and other connections. The CS plays a crucial role in the localisation of the EC-TEC, and although several studies have addressed the variability of the entire length of CS, none of the patterns of variability described in those studies were specific to the location of the EC-TEC. To address the lack of clarity regarding the variability of the CS in relation to the EC-TEC, this study not only investigated the entire length of the CS, but also focused on the length of CS where the EC-TEC was located and identified the sources of interruptions to the EC-TEC.

Relationship Between Anterior Collateral Sulcus and Posterior Collateral Sulcus

In this study, the examination of the relationship between the ACS and PCS showed three general types of patterns. In order of frequency, Pattern 1 was the variant where there was no connection between ACS and PCS (50.0%), Pattern 2 showed an end-end connection between ACS and PCS to form a continuous CS (32.1%), and Pattern 3 was when ACS and PCS are contiguous but merged together from the sides (17.9%). This classification of CS connections is similar to that used in Huntgeburth and Petrides (2012) who reported frequencies of 63.75%, 2.5% and 33.75% for Patterns 1, 2 and 3 respectively, based on 1.5T MRI brain scans obtained from 40 healthy individuals. Cikla et al. (2016) used a method of categorisation that was similar to that used in Huntgeburth and Petrides (2012). Cikla et al. (2016) observed rates of 57.1%, 25.7% and 17.4% for Patterns 1, 2 and 3 respectively, in a sample of 35 left and 35 right hemispheres from individuals with no known significant medical history and with no significant pathological changes in the brain specimens.

While the estimate for Pattern 1 in this study is lower than that observed by Huntgeburth and Petrides (2012), and to a lesser extent, than that reported by Cikla et al. (2016), it is quite similar to those stated in the studies by Ding and Van Hoesen (2010) and Chau et al. (2014), who found rates of 52% ($N = 130$ hemispheres) and 50% ($N = 30$ hemispheres) respectively. Feczko et al. (2009) had broken down their categorisation of CS more finely and they described that a deep, single-branched CS was the most common, reporting a frequency of 31.9% ($N = 210$ hemispheres), which is similar to the results in this study for Pattern 2. Other studies have categorised the CS simply in terms of separated (Pattern 1) and connected (Pattern 2 and 3 combined). More generally, in terms of a connected CS, the rate of 50.0% reported in this study is slightly higher than the numbers reported in Novak et al. (2002; 33%, $N = 100$ hemispheres) and Ovalioglu et al. (2018; 31.6%, $N = 38$ hemispheres) but is again quite similar to Ding and Van Hoesen (2010) and Chau et al. (2014) estimates of 45% and 50% respectively.

The results also revealed that when ACS and PCS were separate, more often than not, there would be a degree of overlap between the two segments, which concurs with the findings from Ono et al. (1990) and Chau et al. (2014). It was also not uncommon to observe side-branches along the ACS. In this study, side-branches were observed in 32.1% of the hemispheres examined, which falls approximately in the middle of what others have reported (Cikla et al., 2016; Ono et al., 1990). Similar to the observations of Cikla et al. (2016), the most common location of a side branch is at the anterior segment of the ACS. Infrequently, the ACS connected to the OTS (2.9%). This observation was also reported to be an uncommon occurrence in other studies, with reported frequencies ranging from 3% - 8% (Chau et al., 2014; Novak et al., 2002; Ono et al., 1990; Ovalioglu et al., 2018). However, a higher frequency of 18.5% have also been observed (Cikla et al., 2016).

Patterns of Interruption to the Entorhinal and Transentorhinal Cortices

The importance of detailing these patterns of branching and connections of the CS to help understand the impact these variations in morphology may have on the location of the EC-TEC. Studies using cytoarchitectonic analysis or markers of cytoarchitectonic structures generally found that the entorhinal cortex mainly sits in the gyrus ambiens or dorsomedial bank of the parahippocampal gyrus and may extend slightly into the CS, while the transentorhinal cortex occupies the medial branch of the CS and may extend slightly into the lateral bank when the CS is very shallow (Ding & Van Hoesen, 2010; Insausti et al., 1998; Taylor & Probst, 2008; Van Hoesen et al., 2000). Clearly, the CS is an important landmark for determining the location of the EC-TEC *in vivo* and variations in the CS would interrupt the reference points for the borders of the EC-TEC. In this study, an interruption to the reference point was defined as a distinct change in the location of the shoulder or fundus of the medial bank of the CS from one slice to the next contiguous coronal slice.

When examining the length of the CS where EC-TEC is located, in three-quarters of the cases, the EC-TEC is contained entirely within the ACS, regardless of the continuity of the CS. In approximately one-fifth of the hemispheres examined, the posterior section of the EC-TEC would be located within the PCS. Although the examination of the CS showed that in almost 21% of the cases, the ACS is not connected to and does not overlap the PCS (Pattern 1b), in only 10 out of these 29 cases were the CS not present on the coronal slices prior to the posterior limit of the EC-TEC.

In terms of interruptions to the EC-TEC, this study observed that the interruptions often occurred in the presence of a discontinuous or contiguous CS, which accounted for 53.8% of all interruptions. In more than a third of the cases, the EC-TEC would be interrupted by the merger of a lateral sulcus to the medial CS (Type 2). In such cases, most typically the medial CS would flatten out and disappear, and the formerly lateral sulcus

would become the most medial sulcus (Type 2b). However, in other cases, the medial and lateral sulci would conflate within the sulcal depth to form one deep CS (Type 2a). In rare cases, a separate sulcus was observed to develop from the medial bank of the medial CS to become a newer, more medial CS (Type 3).

However, other than a non-continuous CS being a source of EC-TEC interruption, another very common source of interruption to the EC-TEC was the presence of side-branches (38.5%). The side-branches entered or arose from the main medial sulcus, forming a V-shaped convolution and changing the depth of the ACS after merging or bifurcating. Most of the interruptions arising from these side-branches occurred at the anterior segment of the ACS. Infrequently, the EC-TEC might be interrupted by a connection to the OTS, but this type of disruption only occurred 4 out of the 52 cases of interrupted EC-TECs.

Relevance to the Localisation of the Entorhinal and Transentorhinal Cortices

In the development of a segmentation protocol, most studies agree that the boundaries of the EC-TEC, particularly the transentorhinal cortex, should be dependent on the sulcal depth. Furthermore, given the variability of the CS, protocols need to provide a clear decision tree on how to locate the EC-TEC. However, the definitions of both medial and lateral borders of the EC-TEC can be markedly different between protocols.

For example, the lateral boundary of the entorhinal cortex as described by Insausti et al. (1998) varied according to the sulcal depth, whilst most studies have opted for a distinct boundary definition on the medial edge, shoulder or fundus of the collateral sulcus as a guideline (Berron et al., 2017; Goncharova et al., 2001; Kivisaari et al., 2013; Meade et al., 2008; Pruessner et al., 2002). Differences in the methods of defining the boundaries could result in different findings between protocols, as demonstrated in a comparative study by C. C. Price et al. (2010). They compared three published entorhinal protocols and found that besides large differences in entorhinal volumes derived from the three protocols, a lateral

border that varied according to CS depth also resulted in lower inter-rater reliability compared to the protocols with a fixed boundary definition.

The few extant transentorhinal manual segmentation protocols that recognise the need to address the variability of the CS also vary with regard to their medial and lateral borders. Berron et al. (2017) provided detailed instructions for boundary placement that was based on four categories of sulcal depth. On the other hand, Kivisaari et al. (2013) also provided details of boundary placement for the transentorhinal cortex but the rules for border determination were based on only two categories of sulcal depth. While both protocols use distinct boundary definitions, the boundary definitions differ depending on the type of CS encountered. For example, when encountering a Type 2b variation, which is the variation with a shallow medial sulcus that flattens out and disappears posteriorly, the protocol by Berron et al. (2017) would introduce a more gradual but potentially more variable shift in boundary towards the lateral sulcus. Conversely, the protocol by Kivisaari et al. (2013) would require a comparatively more abrupt but distinct change in boundary from the shallower medial sulcus to the deeper lateral sulcus. The comparability of these two protocols is yet to be fully assessed.

Besides *in vivo* manual protocol development, automated segmentation protocols for the MTL can also be vastly discrepant from one to the next. Considering the variability of the CS, it remains unclear whether automated methods using a single atlas approach to the localisation of EC-TEC would suffice (Augustinack et al., 2013; Yushkevich, Pluta, et al., 2015), or it would be more appropriate to develop a multi-atlas approach based on variable CS depths or branching patterns (Xie et al., 2017).

Limitations and Conclusion

A limitation of this study is that it was only based on the 3T MRI scans of older adults with low retention of amyloid beta on neuroimaging, which some may argue is not

generalisable across all populations. For example, the possibility of a single, continuous CS being more prevalent in people with AD and in people with temporal lobe epilepsy was raised (Feczko et al., 2009; Kim et al., 2008), but the relationship between CS variants and pathology remains to be clarified. The fact that the determination of the locations of the EC-TEC was based solely on gross anatomical landmarks and therefore not an accurate representation of its actual location - anything short of a cytoarchitectural investigation would likely be less reliable – may also pose another limitation. However, the selected slices are based on widely recommended anterior and posterior limits in prior histological studies (Ding & Van Hoesen, 2010; Insausti et al., 1998; Insausti et al., 1995), which makes the observations in this study more meaningful in its application to *in vivo* MRI segmentations.

Although several studies have highlighted the variability of the CS in its entirety, there has been no literature found that specifically discussed the rates of CS variability in relation to the location of the EC-TEC. As the findings and many prior researches suggest, the CS is not a static landmark. The shape and depth of the CS changes according to the continuity between the anterior and posterior segments. More importantly, the frequencies of variability can be different between the entire CS and the length of CS where the EC-TEC is located. This study shows that the contiguity of the ACS and PCS is not the only point of consideration when localising the EC-TEC. There are other common sources of interruption, most commonly, the occurrence of side-branches along the anterior length of the CS. The existence of these side-branches emphasises the need for future studies to consider the types and frequencies of variability that might be encountered in the process of defining the location of the EC-TEC on MRI scans, particularly in the development of both manual and automated segmentation protocols.

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 7(3), 270-279.
doi:10.1016/j.jalz.2011.03.008
- Amaral, D. G., Insausti, R., & Cowan, W. M. (1987). The entorhinal cortex of the monkey: I. Cytoarchitectonic organization. *Journal of Comparative Neurology*, 264(3), 326.
doi:10.1002/cne.902640305
- Anstey, K. J., Cherbuin, N., Eramudugolla, R., Sargent-Cox, K., Easteal, S., Kumar, R., & Sachdev, P. (2013). Characterizing mild cognitive disorders in the young-old over 8 years: Prevalence, estimated incidence, stability of diagnosis, and impact on IADLs. *Alzheimers & Dementia*, 9(6), 640-648. doi:10.1016/j.jalz.2012.11.013
- Augustinack, J. C., Huber, K. E., Stevens, A. A., Roy, M., Frosch, M. P., van der Kouwe, A. J., . . . Alzheimer's Disease Neuroimaging, I. (2013). Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. *Neuroimage*, 64, 32-42.
doi:10.1016/j.neuroimage.2012.08.071
- Bayram, E., Caldwell, J. Z. K., & Banks, S. J. (2018). Current understanding of magnetic resonance imaging biomarkers and memory in Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 4, 395-413.
doi:10.1016/j.trci.2018.04.007
- Berron, D., Vieweg, P., Hochkeppeler, A., Pluta, J. B., Ding, S. L., Maass, A., . . . Wisse, L. E. M. (2017). A protocol for manual segmentation of medial temporal lobe subregions in 7 Tesla MRI. *NeuroImage: Clinical*, 15, 466-482. doi:10.1016/j.nicl.2017.05.022

- Bonner-Jackson, A., Mahmoud, S., Miller, J., & Banks, S. J. (2015). Verbal and non-verbal memory and hippocampal volumes in a memory clinic population. *Alzheimer's Research and Therapy*, 7(1), 61. doi:10.1186/s13195-015-0147-9
- Braak, H., & Braak, E. (1985). On areas of transition between entorhinal allocortex and temporal isocortex in the human brain. Normal morphology and lamina-specific pathology in Alzheimer's disease. *Acta Neuropathologica*, 68(4), 325-332. doi:10.1007/bf00690836
- Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4), 239-259. doi:10.1007/bf00308809
- Braak, H., & Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging*, 16(3), 271-278. doi:10.1016/0197-4580(95)00021-6
- Braak, H., Braak, E., & Bohl, J. (1993). Staging of Alzheimer-related cortical destruction. *European Neurology*, 33(6), 403-408. doi:10.1159/000116984
- Brodmann, K., & Garey, L. J. (2006). *Brodmann's localization in the cerebral cortex: The principles of comparative localisation in the cerebral cortex based on cytoarchitectonics*: [New York, NY] : Springer, c2006.
- Chau, A. M., Stewart, F., & Gragnaniello, C. (2014). Sulcal and gyral anatomy of the basal occipital-temporal lobe. *Surgical & Radiologic Anatomy*, 36(10), 959-965. doi:10.1007/s00276-014-1294-6
- Chételat, G., Desgranges, B., de la Sayette, V., Viader, F., Berkouk, K., Landeau, B., . . . Eustache, F. (2003). Dissociating atrophy and hypometabolism impact on episodic memory in mild cognitive impairment. *Brain*, 126(9), 1955-1967. doi:10.1093/brain/awg196

- Cikla, U., Menekse, G., Quraishi, A., Neves, G., Keles, A., Liu, C., . . . Baskaya, M. K. (2016). The sulci of the inferior surface of the temporal lobe: An anatomical study. *Clinical Anatomy*, 29(7), 932-942. doi:10.1002/ca.22767
- Deweert, B., Lehericy, S., Pillon, B., Baulac, M., Chiras, J., Marsault, C., . . . Dubois, B. (1995). Memory disorders in probable Alzheimer's disease: The role of hippocampal atrophy as shown with MRI. *Journal of Neurology, Neurosurgery & Psychiatry*, 58(5), 590. doi:10.1136/jnnp.58.5.590
- Ding, S. L., & Van Hoesen, G. W. (2010). Borders, extent, and topography of human perirhinal cortex as revealed using multiple modern neuroanatomical and pathological markers. *Human Brain Mapping*, 31(9), 1359-1379. doi:10.1002/hbm.20940
- Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., Dekosky, S. T., Barberger-Gateau, P., . . . Scheltens, P. (2010). Revising the definition of Alzheimer's disease: A new lexicon. *The Lancet Neurology*, 9(11), 1118-1127. doi:10.1016/s1474-4422(10)70223-4
- Duvernoy, H. M. (1999). *The human brain: Surface, three-dimensional sectional anatomy with MRI, and blood supply*: Springer.
- Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., . . . Ames, D. (2009). The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International Psychogeriatrics*, 21(4), 672-687. doi:10.1017/S1041610209009405
- Feczko, E., Augustinack, J. C., Fischl, B., & Dickerson, B. C. (2009). An MRI-based method for measuring volume, thickness and surface area of entorhinal, perirhinal, and posterior parahippocampal cortex. *Neurobiology of Aging*, 30(3), 420-431. doi:10.1016/j.neurobiolaging.2007.07.023

- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341-355. doi:10.1016/s0896-6273(02)00569-x
- Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis - II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2), 195-207. doi:10.1006/nimg.1998.0396
- Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4), 272-284. doi:10.1002/(sici)1097-0193(1999)8:4<272::aid-hbm10>3.0.co;2-4
- Franko, E., Insausti, A. M., Artacho-Perula, E., Insausti, R., & Chavoix, C. (2014). Identification of the human medial temporal lobe regions on magnetic resonance images. *Human Brain Mapping*, 35(1), 248-256. doi:10.1002/hbm.22170
- Gloor, P. (1997). *The temporal lobe and limbic system*: New York : Oxford University Press, 1997.
- Goncharova, II, Dickerson, B. C., Stoub, T. R., & deToledo-Morrell, L. (2001). MRI of human entorhinal cortex: a reliable protocol for volumetric measurement. *Neurobiology of Aging*, 22(5), 737-745. doi:10.1016/s0197-4580(01)00270-6
- Honeycutt, N. A., Smith, P. D., Aylward, E., Li, Q., Chan, M., Barta, P. E., & Pearlson, G. D. (1998). Mesial temporal lobe measurements on magnetic resonance imaging scans. *Psychiatry Research*, 83(2), 85-94. doi:10.1016/S0925-4927(98)00035-3
- Huntgeburth, S. C., & Petrides, M. (2012). Morphological patterns of the collateral sulcus in the human brain. *European Journal of Neuroscience*, 35(8), 1295-1311. doi:10.1111/j.1460-9568.2012.08031.x

- Insausti, R., & Amaral, D. G. (2012). Chapter 24: Hippocampal Formation. In *The Human Nervous System* (pp. 896-942).
- Insausti, R., Juottonen, K., Soininen, H., Insausti, A. M., Partanen, K., Vainio, P., . . . Pitkanen, A. (1998). MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *Ajnr: American Journal of Neuroradiology*, 19(4), 659-671.
- Insausti, R., Tunon, T., Sobreviela, T., Insausti, A. M., & Gonzalo, L. M. (1995). The human entorhinal cortex: a cytoarchitectonic analysis. *Journal of Comparative Neurology*, 355(2), 171-198. doi:10.1002/cne.903550203
- Kim, H., Bernasconi, N., Bernhardt, B., Colliot, O., & Bernasconi, A. (2008). Basal temporal sulcal morphology in healthy controls and patients with temporal lobe epilepsy. *Neurology*, 70(22 Pt 2), 2159-2165. doi:10.1212/01.wnl.0000313150.62832.79
- Kivisaari, S. L., Probst, A., & Taylor, K. I. (2013). The perirhinal, entorhinal, and parahippocampal cortices and hippocampus: An overview of functional anatomy and protocol for their segmentation in MR Images. In *fMRI : Basics and Clinical Applications* (pp. 239).
- Kordower, J. H., Chu, Y., Stebbins, G. T., DeKosky, S. T., Cochran, E. J., Bennett, D., & Mufson, E. J. (2001). Loss and atrophy of layer II entorhinal cortex neurons in elderly people with mild cognitive impairment. *Annals of Neurology*, 49(2), 202-213. doi:doi:10.1002/1531-8249(20010201)49:2<202::AID-ANA40>3.0.CO;2-3
- Krimer, L. S., Hyde, T. M., Herman, M. M., & Saunders, R. C. (1997). The entorhinal cortex: an examination of cyto- and myeloarchitectonic organization in humans. *Cerebral Cortex*, 7(8), 722-731. doi:10.1093/cercor/7.8.722
- Kulason, S., Tward, D. J., Brown, T., Sicat, C. S., Liu, C. F., Ratnanather, J. T., . . . Alzheimer's Disease Neuroimaging, I. (2018). Cortical thickness atrophy in the

- transentorhinal cortex in mild cognitive impairment. *NeuroImage: Clinical*, 21, 101617. doi:10.1016/j.nicl.2018.101617
- Li, X. D., Jiao, J. S., Shimizu, S., Jibiki, I., Watanabe, K., & Kubota, T. (2012). Correlations between atrophy of the entorhinal cortex and cognitive function in patients with Alzheimer's disease and mild cognitive impairment. *Psychiatry and Clinical Neurosciences*, 66(7), 587-593. doi:10.1111/pcn.12002
- Maass, A., Lockhart, S. N., Harrison, T. M., Bell, R. K., Mellinger, T., Swinnerton, K., . . . Jagust, W. J. (2018). Entorhinal tau pathology, episodic memory decline, and neurodegeneration in aging. *Journal of Neuroscience*, 38(3), 530-543. doi:10.1523/JNEUROSCI.2028-17.2017
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to 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), 263-269. doi:10.1016/j.jalz.2011.03.005
- Meade, C. E., Bowden, S. C., Whelan, G., & Cook, M. J. (2008). Rhinal cortex asymmetries in patients with mesial temporal sclerosis. *Seizure-European Journal of Epilepsy*, 17(3), 234-246. doi:10.1016/j.seizure.2007.07.010
- Meunier, M., Hadfield, W., Bachevalier, J., & Murray, E. A. (1996). Effects of rhinal cortex lesions combined with hippocampectomy on visual recognition memory in rhesus monkeys. *Journal of Neurophysiology*, 75(3), 1190-1205. doi:10.1152/jn.1996.75.3.1190
- Meyer, P., Feldkamp, H., Hoppstadter, M., King, A. V., Frolich, L., Wessa, M., & Flor, H. (2013). Using voxel-based morphometry to examine the relationship between regional

- brain volumes and memory performance in amnesic mild cognitive impairment.
Frontiers in Behavioral Neuroscience, 7, 89. doi:10.3389/fnbeh.2013.00089
- Mueller, S. G., Schuff, N., Yaffe, K., Madison, C., Miller, B., & Weiner, M. W. (2010). Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease.
Human Brain Mapping, 31(9), 1339-1347. doi:10.1002/hbm.20934
- Novak, K., Czech, T., Prayer, D., Dietrich, W., Serles, W., Lehr, S., & Baumgartner, C. (2002). Individual variations in the sulcal anatomy of the basal temporal lobe and its relevance for epilepsy surgery: An anatomical study performed using magnetic resonance imaging. *Journal of Neurosurgery*(6), 464-473.
doi:10.3171/jns.2002.96.3.0464
- Olsen, R. K., Yeung, L. K., Noly-Gandon, A., D'Angelo, M. C., Kacollja, A., Smith, V. M., . . . Barense, M. D. (2017). Human anterolateral entorhinal cortex volumes are associated with cognitive decline in aging prior to clinical diagnosis. *Neurobiology of Aging*, 57, 195-205. doi:10.1016/j.neurobiolaging.2017.04.025
- Ono, M. M. D., Kubik, S., & Abernathey, C. D. (1990). *Atlas of the cerebral sulci*: Stuttgart : G. Thieme Verlag ; New York : Thieme Medical Publishers, 1990.
- Ovalioglu, A. O., Ovalioglu, T. C., Canaz, G., & Emel, E. (2018). Morphologic variations of the collateral sulcus on the mediobasal region of the temporal lobe: An anatomical study. *World Neurosurgery*, 118, e212-e216. doi:10.1016/j.wneu.2018.06.156
- Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., Hanninen, T., Laakso, M. P., . . . Soininen, H. (2004). Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiology of Aging*, 25(3), 303-310.
doi:10.1016/s0197-4580(03)00084-8
- Peter, J., Lahr, J., Minkova, L., Lauer, E., Grothe, M. J., Teipel, S., . . . Kloeppel, S. (2016). Contribution of the cholinergic system to verbal memory performance in mild

- cognitive impairment. *Journal of Alzheimer's Disease*, 53, 991-1001.
doi:10.3233/JAD-160273
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment - Clinical characterization and outcome. *Archives of Neurology*, 56(3), 303-308. doi:10.1001/archneur.56.3.303
- Price, C. C., Wood, M. F., Leonard, C. M., Towler, S., Ward, J., Montijo, H., . . . Schmalzfuss, I. (2010). Entorhinal cortex volume in older adults: Reliability and validity considerations for three published measurement protocols. *Journal of the International Neuropsychological Society*, 16(5), 846-855.
doi:10.1017/s135561771000072x
- Price, J. L., Ko, A. I., Wade, M. J., Tsou, S. K., McKeel, D. W., & Morris, J. C. (2001). Neuron number in the entorhinal cortex and CA1 in preclinical Alzheimer disease. *Archives of Neurology*, 58(9), 1395-1402. doi:10.1001/archneur.58.9.1395
- Pruessner, J. C., Kohler, S., Crane, J., Pruessner, M., Lord, C., Byrne, A., . . . Evans, A. C. (2002). Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. *Cerebral Cortex*, 12(12), 1342-1353. doi:10.1093/cercor/12.12.1342
- Sperling, R., & Johnson, K. (2013). Biomarkers of Alzheimer disease: Current and future applications to diagnostic criteria. *Continuum*, 19(2 Dementia), 325-338.
doi:10.1212/01.CON.0000429181.60095.99
- Tapiola, T., Pennanen, C., Tapiola, M., Tervo, S., Kivipelto, M., Hänninen, T., . . . Soininen, H. (2008). MRI of hippocampus and entorhinal cortex in mild cognitive impairment: A follow-up study. *Neurobiology of Aging*, 29(1), 31-38.
doi:10.1016/j.neurobiolaging.2006.09.007

- Taylor, K. I., & Probst, A. (2008). Anatomic localization of the transentorhinal region of the perirhinal cortex. *Neurobiology of Aging*, 29(10), 1591-1596.
doi:10.1016/j.neurobiolaging.2007.03.024
- Tward, D. J., Sicat, C. S., Brown, T., Bakker, A., Gallagher, M., Albert, M., & Miller, M. (2017). Entorhinal and transentorhinal atrophy in mild cognitive impairment using longitudinal diffeomorphometry. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 9, 41-50. doi:10.1016/j.dadm.2017.07.005
- Van Hoesen, G. W., Augustinack, J. C., Dierking, J., Redman, S. J., & Thangavel, R. (2000). The parahippocampal gyrus in Alzheimer's disease - Clinical and preclinical neuroanatomical correlates. In H. E. Scharfman, M. P. Witter, & R. Schwarcz (Eds.), *Parahippocampal Region: Implications for Neurological and Psychiatric Diseases* (Vol. 911, pp. 254-274). New York: New York Acad Sciences.
- Velayudhan, L., Proitsi, P., Westman, E., Muehlboeck, J. S., Mecocci, P., Vellas, B., . . . AddNeuroMed, C. (2013). Entorhinal cortex thickness predicts cognitive decline in Alzheimer's disease. *Journal of Alzheimer's Disease*, 33(3), 755-766. doi:10.3233/jad-2012-121408
- Vyhnalek, M., Nikolai, T., Andel, R., Nedelska, Z., Rubinova, E., Markova, H., . . . Hort, J. (2014). Neuropsychological correlates of hippocampal atrophy in memory testing in nondemented older adults. *Journal of Alzheimer's Disease*, 42 Suppl 3, S81-90.
doi:10.3233/jad-132642
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., . . . Alzheimer's Disease Neuroimaging, I. (2015). 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's & Dementia*, 11(6), e1-120. doi:10.1016/j.jalz.2014.11.001

- Wolk, D. A., Das, S. R., Mueller, S. G., Weiner, M. W., Yushkevich, P. A., & Alzheimer's Disease Neuroimaging, I. (2017). Medial temporal lobe subregional morphometry using high resolution MRI in Alzheimer's disease. *Neurobiology of Aging*, 49, 204-213. doi:10.1016/j.neurobiolaging.2016.09.011
- Xie, L., Pluta, J. B., Das, S. R., Wisse, L. E. M., Wang, H., Mancuso, L., . . . Yushkevich, P. A. (2017). Multi-template analysis of human perirhinal cortex in brain MRI: Explicitly accounting for anatomical variability. *Neuroimage*, 144(Pt A), 183-202. doi:10.1016/j.neuroimage.2016.09.070
- Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S. L., Gertje, E. C., . . . Wolk, D. A. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, 36(1), 258-287. doi:10.1002/hbm.22627

Chapter 3

Comparative Utility of Manual versus Automated Segmentation of Hippocampus and Entorhinal Cortex Volumes in a Memory Clinic Sample

This chapter addresses one of the main aims of this study, which is to validate the use of FreeSurfer volume estimation against the gold standard of manual delineation. This chapter consist of the author accepted version of a published pilot validation study that was published in the *Journal of Alzheimer's Disease* in 2019. The validation study evaluated the performance of FreeSurfer in two structures that are associated with Alzheimer's disease, namely the hippocampus and entorhinal cortex. The final publication of this study is available at IOS press through <http://dx.doi.org/10.3233/JAD-181172>.

Found in the appendix of this chapter, is a letter to the editor for the *Journal of Alzheimer's Disease*, published online in March 2019. The letter detailed an issue with the volumetric estimations of the entorhinal and perirhinal cortices (perirhinal cortex being analogous to the transentorhinal cortex as highlighted in Chapter 2) found in the newest version of FreeSurfer (v6.0), which was uncovered during the preparation of the main manuscript. The original online publication of this letter is available on the Journal of Alzheimer's disease website, <https://www.j-alz.com/content/labelling-statistical-output-cytoarchitecture>.

Abstract

Structural neuroimaging is a useful non-invasive biomarker commonly employed to evaluate the integrity of mesial temporal lobe structures that are typically compromised in Alzheimer's disease. Advances in quantitative neuroimaging have permitted the development of automated segmentation protocols (e.g. FreeSurfer) with significantly increased efficiency compared to earlier manual techniques. While these protocols have been found to be suitable for large-scale, multi-site research studies, we were interested in assessing the practical utility and reliability of automated FreeSurfer protocols compared to manual volumetry on routinely acquired clinical scans. Independent validation studies with newer automated segmentation protocols are scarce. Two FreeSurfer protocols for each of two regions of interest – the hippocampus and entorhinal cortex – were compared against manual volumetry. High reliability and agreement were found between FreeSurfer and manual hippocampal protocols; however, there was lower reliability and agreement between FreeSurfer and manual entorhinal protocols. Although based on a the relatively small sample of subjects drawn from a memory clinic ($n=27$), our study findings suggest further refinements to improve measurement error and most accurately depict true regional brain volumes using automated segmentation protocols are required, especially for non-hippocampal mesial temporal structures, to achieve maximal utility for routine clinical evaluations.

Keywords: Alzheimer disease, neuroimaging, hippocampus, entorhinal cortex, biomarker, reproducibility of results, temporal lobe

Comparative Utility of Manual versus Automated Segmentation of Hippocampus and Entorhinal Cortex Volumes in a Memory Clinic Sample

Of the core brain biomarkers of neurodegeneration, structural magnetic resonance imaging (MRI) is arguably the most accessible and cost-effective assessment that can be potentially implemented in clinical practice [1]. Although biomarkers specifying the underlying neuropathology are desirable targets in early detection, there are still challenges in standardization, accessibility and cost to be overcome before they may be recommended for routine use in standard clinical practice [2]. As a means of assessing macroscopic integrity, structural imaging although non-specific is considered to be a biomarker reflective of neuronal loss resulting from pathophysiological processes usually occurring in the clinically overt phases of the disease [3].

Typically, an anatomical brain MRI scan of a patient with symptomatic Alzheimer's disease (AD) reveals atrophy of medial temporal lobe structures. It has been found that the extent and pattern of atrophy of these medial structures have high concordance with neurofibrillary tangle distribution and neuronal loss across AD severity stages [3-5]. A meta-analysis of hippocampal and entorhinal cortex volumes measured in follow-up studies of individuals without dementia has also been shown to be predictive of the risk of incipient AD [6].

Some studies have also suggested that entorhinal cortex volumes may be a better predictor of conversion from mild cognitive impairment (MCI) to AD than hippocampal volume [7]. A study by Pennanen, et al. [8], reported that structural MRI employing manually measured hippocampal volumes was able to correctly identify 90.7% of 48 AD patients from 59 normal control participants, and 82.3% of 65 MCI patients from 48 patients with AD. Despite classification of MCI and controls being comparatively poorer using the entorhinal cortex volume (65.9%) than that of classification of AD from controls using hippocampal

volume, the MCI group had significantly more volume loss in the entorhinal cortex than in the hippocampus (16% and 8% respectively). This finding is therefore in line with the Braak stages that suggests AD pathology involves the entorhinal cortex before progressing to the hippocampus [9].

Manual measurement of structural MRI volumes by an expert operator has generally been considered the “gold standard” of quantitative assessment, which contrasts with routine clinical evaluations for AD that typically employ visual inspection alone to detect regional atrophy. Visual inspection may compromise the reliability of structural neuroimaging as a diagnostic biomarker; however, several practical challenges constrain routine use of manual measurement in clinical practice [10]. The challenges include the time-consuming nature of manual measurements, requirement for specialized expertise in delineation of the region of interest, inter-scan ‘rater drift’, as well as a need for a standardized set of technical procedures and a relevant normative database for the most meaningful interpretation of results in clinical settings [1, 11].

To redress these practical limitations, there has been noteworthy development of automated quantitative assessment in the last decade [12-17]. FreeSurfer, a freely available software package, is the most commonly used automated method in studies performing cortical segmentation, structural measurement and other types of neuroimaging analysis [18]. Earlier studies of the standard segmentation protocol for the hippocampus have concluded sufficiently high reliability and acceptable method agreement with manual measurement for research purposes in large group-level studies, despite a limitation of the protocol tending to over-estimate hippocampal volumes relative to manual volumetry [19-22].

With increased interest in studying the impact of neuropathological processes on smaller but pathophysiologically important structures such as the entorhinal cortex, as well as the subfields of the hippocampal complex, numerous updates and new segmentation

protocols have been made available in the FreeSurfer package. Of primary interest to this study are the newest hippocampal and entorhinal cortex protocols consisting of the ex vivo validated entorhinal protocol [23] and hippocampal subfields protocol [11, 24]. Despite preliminary independent appraisals [25, 26], there is still a lack of studies that include manual measurement as a reference standard. Further, there is no clear indication as to which FreeSurfer protocol is the superior option for each corresponding structure. There are suggestions that the improvements in atlas definitions in the latest available hippocampal subfields package may result in closer agreement with manual measurement as a summated volume than its older, standard FreeSurfer hippocampal protocol. Iglesias, et al. [27] investigated FreeSurfer's (v5.3) ability to discriminate between AD and control groups, and found that the hippocampal subfield protocol provided improved classification accuracy when compared to the standard hippocampal protocol, which suggests that the subfields protocol provided more informative estimates of hippocampal volumes than the standard FreeSurfer automated segmentation.

Compared to the hippocampus, the entorhinal cortex is known to be a more challenging structure to delineate manually due to higher inter-individual anatomical variability and a lack of easily discernible landmarks for the cytoarchitectonic boundaries [28, 29]. Similarly, the more recently developed and histologically validated automated ex vivo entorhinal protocol is believed to show better reliability and closer agreement with manual measurement compared to the original FreeSurfer automated entorhinal protocol.

The primary goal of this study was to compare two versions of automated segmentations of the hippocampus and entorhinal cortex employing FreeSurfer software (v6.0) against manual delineation of the same structures. The degree of agreement and reliability of the automated segmentation techniques was examined simultaneously to identify the most accurate relative to manual measurement. A secondary objective was to carry out a

preliminary evaluation of FreeSurfer's capability to perform segmentation and volume estimation of selected medial temporal structures in routinely acquired clinical diagnostic MRI brain scans.

Materials and Methods

Study Design

This study was a retrospective analysis performed on a sample of consecutive, routinely acquired diagnostic brain MRI scans obtained through the St Vincent's Hospital memory clinic. The study was approved by the St. Vincent's Hospital (Melbourne) Human Research Ethics Committee (HREC-A 037-10).

Sample Criteria

Sixty-five cases were selected from the archival database of patients who had been referred to the memory clinic in St Vincent's Hospital Melbourne between 2004 and 2014. All underwent a 1.5T MRI head scan, full clinical and neuropsychological assessment, which includes the full administration of the complete WMS-III/IV whenever possible, as part of their dementia clinical workup.

Twenty-one of the 65 scans were eliminated through the following exclusionary criteria: absence of comprehensive neuropsychological evaluation, memory difficulties attributable to acquired brain injury, inter-current or longstanding neurological or major psychiatric illness. Patients with a prior or current history of mild-to-moderate anxiety or depression were retained given the naturally increased prevalence of mood-related difficulties in this patient group. Three scans that were obtained on follow-up were also eliminated.

To maximise the opportunity for successful processing by a fully automated FreeSurfer pipeline within a limited sample size, scans were also categorically rated for image quality by visual inspection of the scans, taking into account lesions and artefacts which are known to cause segmentation failure in FreeSurfer [10, 19]. The subjective

categorical rating was conducted a senior neuroimaging scientist, resulting in the exclusion of 14 scans to obtain a final sample size of 27 scans. Ten scans from the final sample were randomly selected for intra-rater and inter-rater reliability analysis. These 10 patients (five females), mean age = 63.7 years ($SD = 10.2$) were made up of five patients with subjective symptoms but neither objective impairment on neuropsychological assessment nor documented functional decline, two patients diagnosed with mild cognitive impairment and three patients with clinically diagnosed dementia.

MRI Acquisition

All T1-weighted scans were acquired coronally along the hippocampal axis with 3D-sequenced MPRAGE on either a 1.5T Siemens Avanto scanner ($n = 14$; 12-channel head coils, $TR = 1.53\text{ms}$, $TE = 3.24\text{ms}$, $FOV = 156\text{mm} \times 250\text{mm}$, flip angle = 8° , voxel size $1 \times 1 \times 1.4\text{mm}^3$, 160 slices, matrix 256×160) or a Siemens Symphony scanner ($n = 16$; CP head array coil, $TR = 1.66\text{ms}$, $TE = 4.13\text{ms}$, $FOV = 156\text{mm} \times 250\text{mm}$, flip angle = 8° , voxel size $1 \times 1 \times 1.4\text{mm}^3$, 160 slices, matrix 256×160).

Volumetric Procedures

Automated measurement. Automated derivation of hippocampal and entorhinal volumes for each case was generated using the publicly accessible FreeSurfer software package (v6.0, Martinos Center for Biomedical Imaging, Harvard-MIT, Boston, MA). Technical documentation on the software can be found on the FreeSurfer website (<http://surfer.nmr.mgh.harvard.edu>).

In FreeSurfer, the “recon-all” command triggers a complete set of built-in commands for the fully automated processing of raw DICOM images. Details of this automated procedure have been described in Fischl, et al. [30]. The standard FreeSurfer automated subcortical segmentation protocol for the hippocampus, which will be referred to as FS_{hv} hereafter, is based on a probabilistic atlas described in Fischl, et al. [30]. The standard

FreeSurfer automated cortical segmentation for entorhinal cortex, which will now be referred to as FS_{erc}, is based on the Desikan-Killiany atlas [31]. Further cortical reconstruction is achieved through registration of each surface vertex in a polygonal tessellated model to a spherical atlas and their respective cortical regions-of-interest following further post-processing and refinement of the 3D image data [32-37].

Automated segmentation of all scans was successfully completed on an Ubuntu Linux workstation over an approximate duration of six to eight hours analysis per scan. Left and right hemisphere standard hippocampal and entorhinal protocol volume statistics were extracted from the corresponding cortical parcellation statistics file [31, 38]. The entorhinal volumes generated using the ex vivo validated protocol (FS_{exvivo}) did not require further processing and the volume statistic could be directly extracted from the corresponding label text file.

To generate hippocampal subfield statistics using the FreeSurfer subfields protocol (FS_{subfield}), an additional 30 to 60 minutes of processing time per scan was required [24]. To match the prescriptions of the manual hippocampal protocol (detailed in the next section) as closely as possible, generated volumes from the hippocampal subfield package labelled as presubiculum and parasubiculum were not included in the summated hippocampal subfield volume for each case.

Manual measurement. Manual delineation of the hippocampal (M_{hv}) and entorhinal cortex (M_{erc}) was performed using the Analyze image analysis software (version 12, Brain Imaging Resource, Mayo Foundation, Rochester, MN). Raw DICOM files were prepared for manual tracing by converting them into Analyze format and transformed by cubic spline into volumes of $0.3 \times 0.3 \times 1.4$ mm voxel size to allow as fine a resolution as possible for manual tracing maps. Slice thickness was preserved; no reformatting was required. The manual protocols used for hippocampus and entorhinal cortex are described in Cook, et al. [39] and

Meade, et al. [40] respectively. An example of manual delineation of these two structures, along with FreeSurfer's automated delineation, is illustrated in Figure 3.1. Manual delineation for hippocampus and entorhinal cortex were performed independently of each other, in which complete hippocampal tracings were not visible during execution of the entorhinal protocol and vice versa.

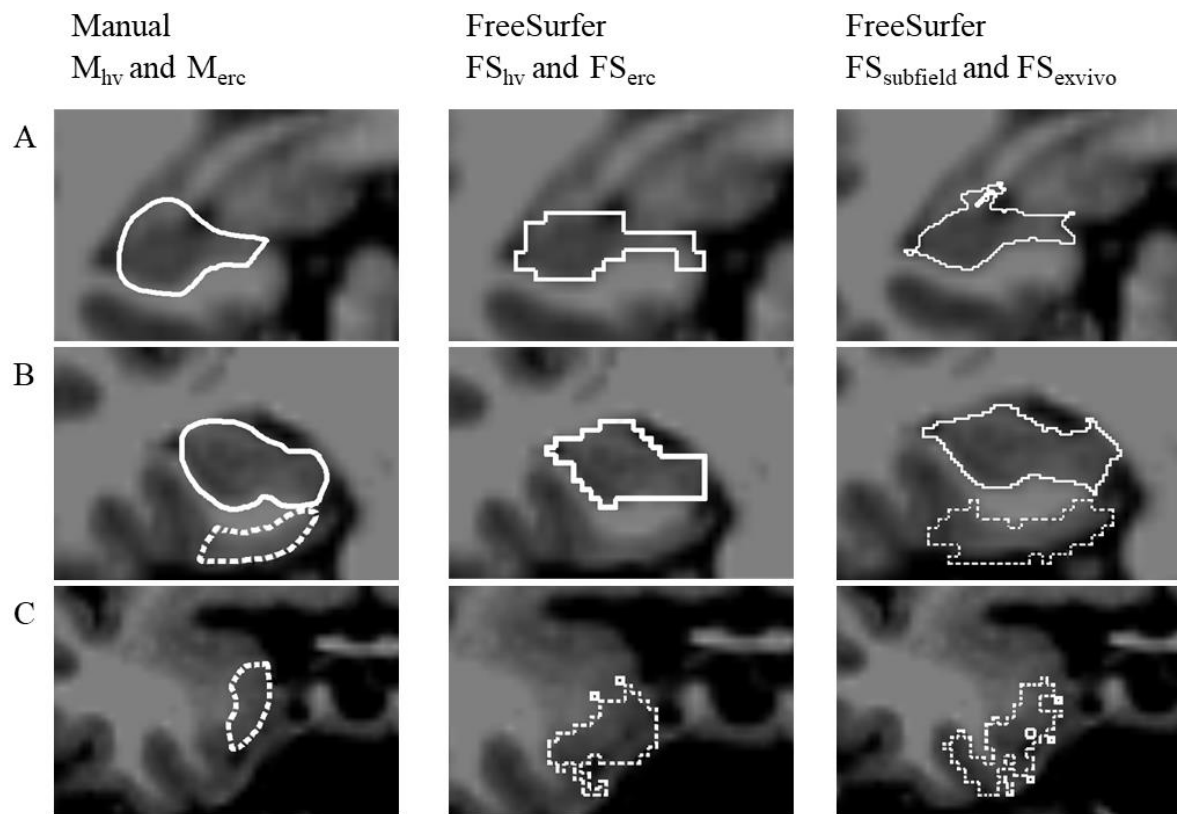
All manual segmentation of hippocampus and entorhinal cortex on the 27 scans was performed by a single trained rater. Two-way, absolute agreement intra-class correlation coefficient (ICC) was used to assess intra- and inter-rater reliability within and between the single trained rater and another rater on 10 randomly selected scans [41-43]. Excellent to good intra-rater reliability was found for both left and right hippocampal volume estimations, single measure ICC = .78 (95% CI: .29, .94) and single measure ICC = .94 (95% CI: .74, .97) respectively; similarly, excellent to good intra-rater reliability was found for both left and right entorhinal volume estimations, single measure ICC = .89 (95% CI: .62, .97) and single measure ICC = .91 (95% CI: .67, .98) respectively. Single measure inter-rater reliability values were found to be excellent between the raters for left and right hippocampal volume estimations, ICC = .92 (95% CI: .73, .98) and ICC = .93 (95% CI: .76, .98) respectively. Single measure inter-rater reliability values were also good for both left and right entorhinal volume estimations, ICC = .78 (95% CI: .36, .94) and ICC = .73 (95% CI = -.07, .94) respectively; refer to Figure 3.2 for Bland-Altman plots illustrating the agreements between the raters.

Statistical Analysis

Statistical analysis was conducted in Stata (version 15). Reliability between manual and automated volume measurements for each region of interest was assessed using two-way absolute agreement intra-class correlation (ICC). The more stringent absolute agreement ICC

Figure 3.1

Key Landmarks of Hippocampal and Entorhinal Manual Protocols



Note. This is a case illustration of key visual landmarks of hippocampal and entorhinal manual protocols from caudal to rostral view on the left most column. The white solid line indicates the borders of the right hippocampus and the white broken line indicates the borders of the right entorhinal cortex.

A: Longest uninterrupted length of fornix (posterior limit of hippocampus).

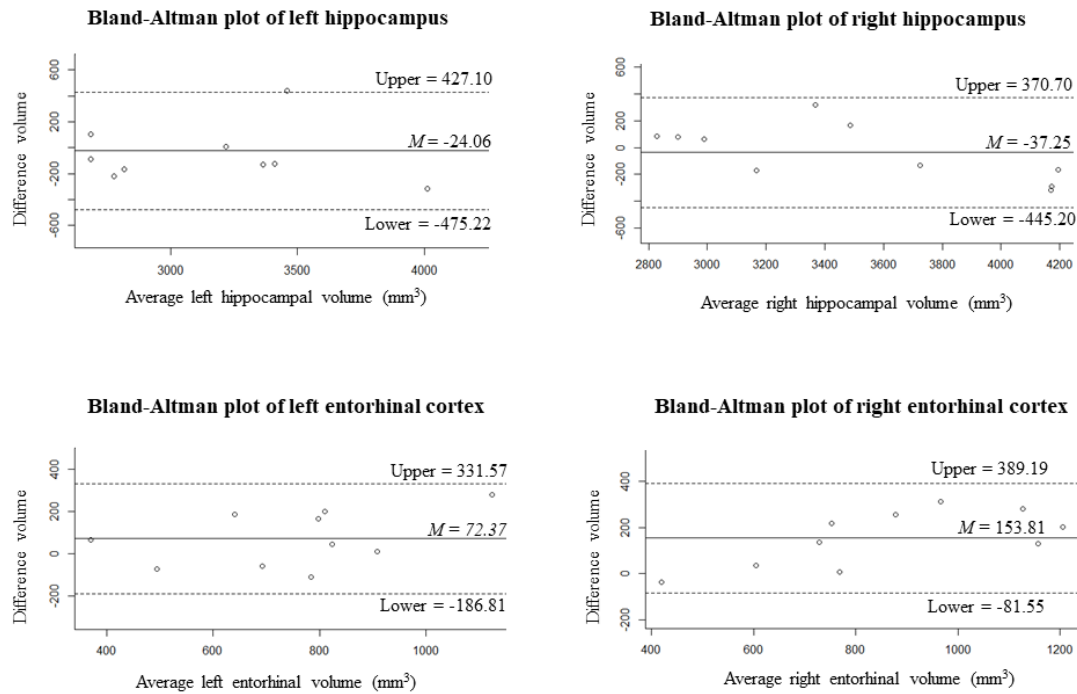
B: Appearance of intralimbic gyrus (posterior limit of entorhinal cortex).

C: Limen insulae continuous with frontal lobe (anterior limit of entorhinal cortex).

Delineations of the right hippocampus and right entorhinal cortex, as determined by FreeSurfer's segmentation protocols, are also illustrated (if present) on the same coronal slices as demonstrated for the manual protocol.

Figure 3.2

Bland-Altman Plots of Hippocampal and Entorhinal Volumes by Two Independent Raters.



Note. This figure shows the Bland-Altman plots of manual volume estimations of left and right hippocampal volume estimations, and left and right entorhinal cortex volume estimations, by two independent raters, on 10 randomly selected scans. Solid lines represent the difference of the mean volumes between the raters. Dashed lines represent the upper and lower 95% limits of agreement.

was used to set a high standard of value agreement between the manual and automated methods.

Pearson correlations were performed to examine the relationship between volumes from each protocol. Correlation is not a measure of agreement [44]; however, calculating a correlation allows a common basis for comparison of the present study's results against previous studies, which vary greatly in methodology and study aims.

The degree of agreement between protocols was assessed with Bland-Altman plots [45]. The difference between volumes derived from each method for all cases are plotted against the mean of both measurements on a graph. The 95% limits of agreement are also plotted, showing the variability of individual differences across the two methods. In view of a small sample size ($n < 60$), the limits of agreement were computed with Lentner's modified formula [46]. In the event that a good fit to the data was not obtained within the limits of agreement, a percentage of the differences was plotted instead of absolute volumes [45-47].

Results

Sample Characteristics

The final analyzed sample consisted of 27 individual memory clinic patients (mean age 64.6 ± 10.4 years). Thirteen were male (mean age 65.3 ± 9.4 years) and 14 were female (mean age 63.9 ± 11.6 years). At time of scan acquisition, 16 of the patients were deemed "normal" with subjective complaints; 14 of these individuals having no abnormal findings on other objective assessments, including detailed neuropsychological testing. Two of those in the normal group scored 1.5 standard deviations below the mean on one of the index scores on the WMS-III/IV; however, both endorsed mild-moderate levels of anxiety and/or depressive symptoms based on the Beck's Depression Inventory-II (BDI-II score > 13) and the State-Trait Anxiety Inventory (STAI scores $> 75\%$). Consequently, their reduced memory performances were conservatively determined by the assessing clinical

neuropsychologist to probably be a result of inter-current anxiety or mood difficulties. In accordance with Petersen, et al. [48] criteria, three patients were classified as having mild cognitive impairment wherein they scored >1.5 standard deviations below the mean on one of the index scores on the WMS-III or IV, with no report of personal functional decline. Eight received a clinical diagnosis of AD, scoring at least 1.5 standard deviations below the mean on at least one of the index scores on the WMS-III or IV, and differentiated from mild cognitive impairment by a corroborated report of significant personal functional decline [49].

Agreement between Protocols

Hippocampal. ICC comparison of left and right hippocampal volume ($n = 27$) as estimated by M_{hv} and FS_{hv} showed a single measure ICC of .64 (95% CI: .21, .84) and .78 (95% CI: .50, .90), respectively. Comparing the left and right hippocampal volumes obtained with M_{hv} and $FS_{subfield}$, the single measure ICC was .52 (95% CI = .11, .77) and .51 (95% CI = -.10, .81), respectively.

Entorhinal. Each individual subject's entorhinal cortex volumes ($n = 27$) as estimated by M_{erc} and FS_{erc} were analyzed by each hemisphere, revealing a low single measure ICC of .03 (95% CI = -.04, .16) for the left and a similarly low single measure ICC of .04 (95% CI = -.04, .18) for the right. The single measure ICC comparing the left and right entorhinal cortex volumes estimated using M_{erc} and FS_{exvivo} was -.09 (95% CI = -.45, .30) and .08 (95% CI = -.10, .33), respectively.

Association between Methods

Hippocampal. Pearson's correlations of left and right hippocampal volumes ($n = 27$) estimated by M_{hv} with that of automated methods were strong for FS_{hv} ($r = .73, p < .001$; $r = .82, p < .001$, respectively), as well as $FS_{subfield}$ ($r = .63, p < .001$; $r = .76, p < .001$, respectively).

Entorhinal. With respect to entorhinal cortex volumes, correlations of left and right volumes between the manual protocol and both automated protocols, FS_{erc} and FS_{exvivo} , did not reach significance.

Agreement between Methods

Hippocampal. The absolute and percentage difference between the mean volumes of each protocol are shown in Table 3.1. The Bland-Altman plots for left and right hippocampal absolute volumes are shown in Figure 3.3 and Bland-Altman plots for left and right entorhinal cortex percentage volumes are shown in Figure 3.4. Assumptions of normality were met for all plotted variables.

Compared to manual hippocampal measurements, FS_{hv} tended to overestimate hippocampal volumes while $FS_{subfield}$ tended to underestimate hippocampal volumes in both left and right hemispheres (Table 3.1). However, the 95% limits of agreement in hippocampal volume estimations for both automated hippocampal protocols were similar (Figure 3.3).

Entorhinal. With respect to entorhinal cortex volume estimations, left and right entorhinal estimations by FS_{erc} were inflated compared to manual measurements, while FS_{exvivo} overestimated the left entorhinal and underestimated the right entorhinal cortex (Table 3.1). Comparing the percentage mean volume difference between hippocampal and entorhinal volumes, the average percentage difference between the hippocampal volume estimations were much smaller than the average percentage difference between entorhinal cortex volumes (Table 3.1). The only notable exception was the average percentage difference between the estimations by left M_{erc} and FS_{erc} , in which the average difference was only -6.2%; however, it also has the widest 95% limits of agreement (Figure 3.4). Both automated hippocampal protocols also showed a smaller coefficient of variation (percentage standard deviation) than the automated entorhinal cortex protocols, suggesting larger measurement error among the entorhinal protocols relative to manual measurement.

Table 3.1*Absolute and Percentage Difference of Mean Volumes Between Protocols*

	M		SD		Maximum		Minimum	
	Volume (mm ³)	Percentage (%)	Volume (mm ³)	Percentage (%)	Volume (mm ³)	Percentage (%)	Volume (mm ³)	Percentage (%)
Hippocampal								
<i>Left hemisphere</i>								
M _{hv} - FS _{hv}	-278.66	-8.26	385.38	11.51	+535.29	+15.68	-995.51	-31.61
M _{hv} - FS _{subfield}	+291.70	+8.99	415.18	12.76	+1179.36	+38.16	-310.34	-11.21
<i>Right hemisphere</i>								
M _{hv} - FS _{hv}	-171.16	-4.73	309.13	8.22	+332.03	+9.67	-936.12	-23.25
M _{hv} - FS _{subfield}	+446.69	+13.6	317.10	9.48	+1138.69	+34.14	-88.27	-2.45
Entorhinal								
<i>Left hemisphere</i>								
M _{erc} - FS _{erc}	-1123.64	-81.17	428.33	25.42	-353.08	-27.91	-2198.60	-134.23
M _{erc} - FS _{exvivo}	-69.34	-6.23	376.53	44.46	+470.03	+85.49	-1027.60	-100.35
<i>Right hemisphere</i>								
M _{erc} - FS _{erc}	-960.13	-70.09	361.32	25.08	-341.00	-23.07	-1968.60	-130.57
M _{erc} - FS _{exvivo}	+303.54	+37.74	263.79	31.64	+817.54	+95.49	-278.60	-42.04

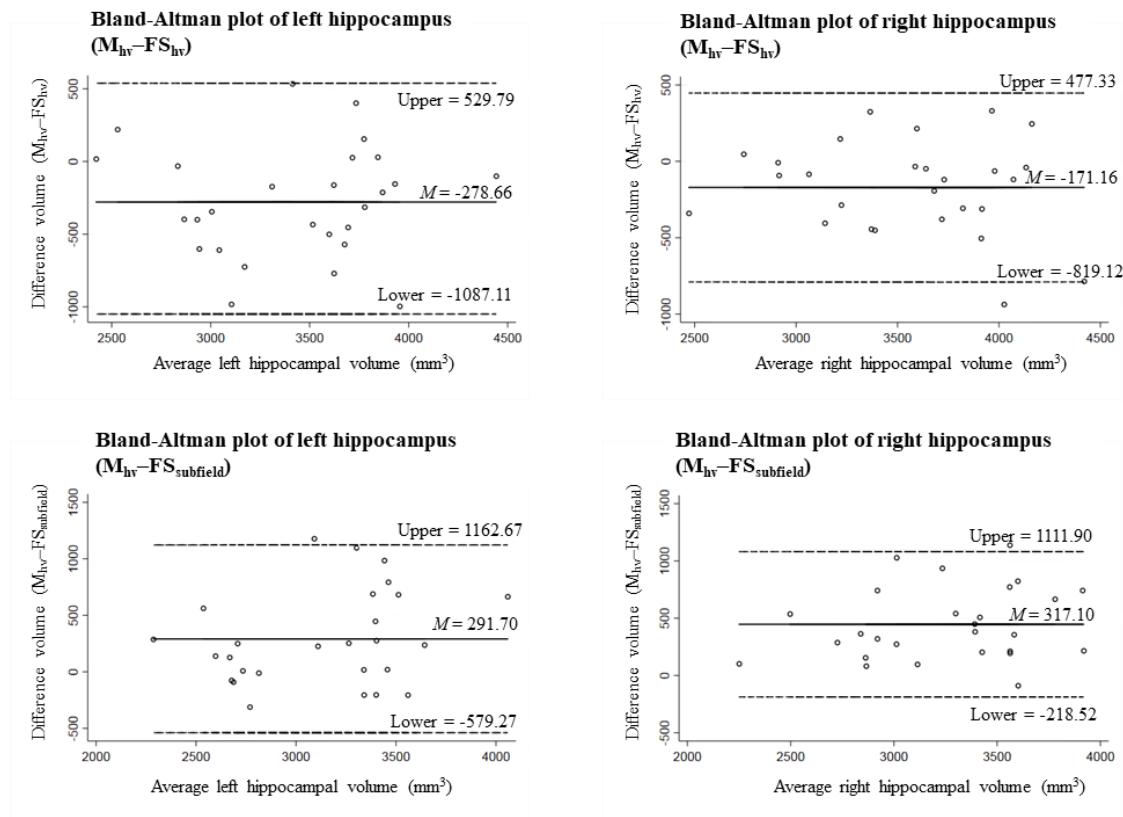
Discussion

The present study compared two FreeSurfer automated segmentation protocols with manual measurement in each of two regions of interest, namely, the hippocampus and entorhinal cortex. The standard hippocampal protocol in FreeSurfer is based on an existing probabilistic atlas described in Fischl, et al. [30], while the standard entorhinal protocol in FreeSurfer is based on the Desikan-Killiany atlas [31]. New automated segmentation protocols for these regions of interest include the hippocampal subfields package [24] and the ex vivo validated entorhinal protocol [23]. To assess the potential clinical translational utility of automated volumetry in a “real world” setting, we undertook a preliminary survey of the potential pitfalls and strengths of an automated approach based on analysis of routinely acquired diagnostic MRI brain scans from a small sample of memory clinic patients. While our relatively limited sample size precludes definitive conclusions, our study findings suggest further refinements to improve measurement error and thereby more accurately depict true regional brain volumes using automated segmentation protocols are required, especially for non-hippocampal mesial temporal structures.

The results of this study indicated that the standard FreeSurfer hippocampal protocol tended to produce over-estimations of volume and the hippocampal subfields protocol tended to under-estimate volumes relative to manual measurement. One study reported inflated standard FreeSurfer (v4.0.5) hippocampal protocol measurements of up to 26% larger in a sample of 403 community dwelling adults and 36% larger in a combined sample of younger bipolar disorder and healthy control patients relative to manual measurement [20]. In contrast and contrary to our findings, one study found that hippocampal subfields protocol overestimated whole hippocampal volumes when comparing against manual measurements using FreeSurfer version 5.1.0 [50]. The differences found in the present study for the hippocampal standard and subfields protocols, particularly the smaller mean differences

Figure 3.3

Bland-Altman plot of Hippocampal Volumetric Difference Between Manual and FreeSurfer



Note. Bland-Altman plot of left and right hippocampus difference volumes against corresponding average manual (M_{hv}) and FreeSurfer volumes as estimated using FS_{hv} and $FS_{subfield}$. Solid lines represent the difference of the mean volumes between methods (see Table 3.1). Dashed lines represent the upper and lower 95% limits of agreement.

found between manual and automated measures, might be indicative of refinements of the automated hippocampal protocols in the newer version of FreeSurfer (v6.0) used in this study suggesting perhaps a more representative delineation of the hippocampal volumes.

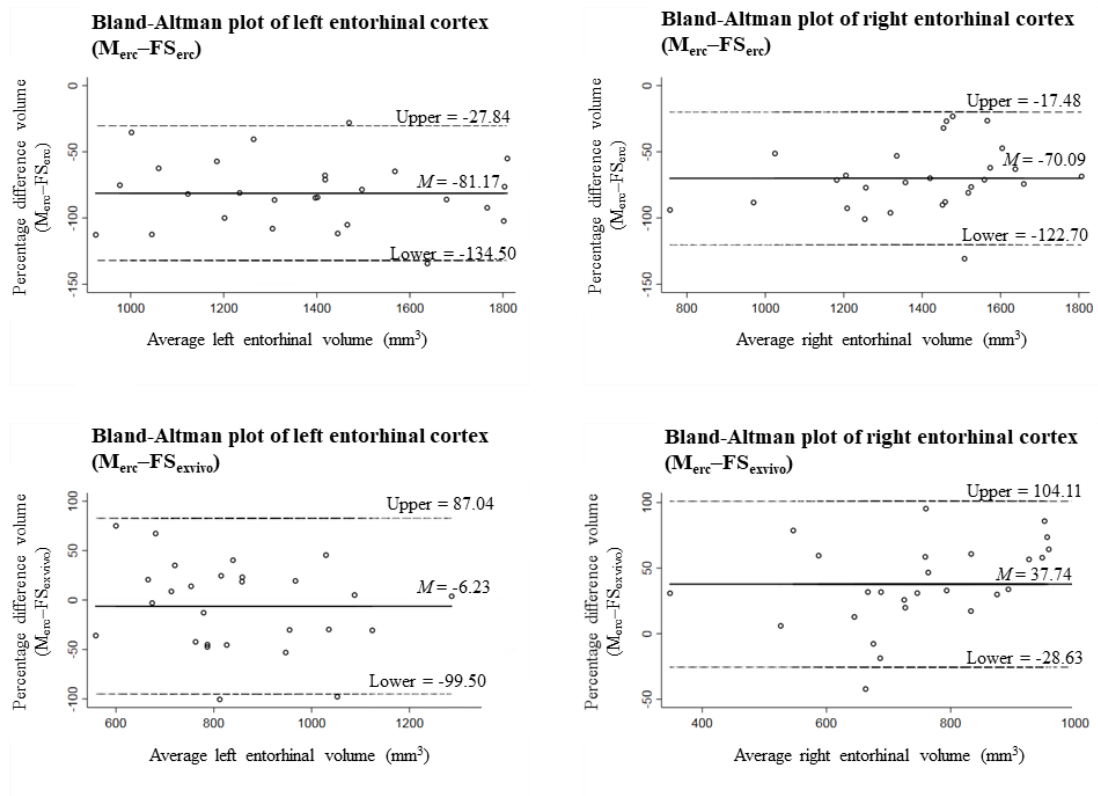
Both FS_{hv} and $FS_{subfield}$ showed similar reliability and agreement to manual hippocampal measurement, with ICC values that suggested moderate to good reliability and limits of agreement that were similar and acceptably narrow. Although the confidence intervals for intra-class correlations were quite wide, the point estimates for hippocampal volumes were

consistent across both hemispheres in both methods, which reassuringly suggests that the wide confidence intervals were likely due to small sample size rather than unreliability in estimates of agreement. In general, correlations between FreeSurfer and manual hippocampal measurement were consistent with those reported in previous studies [19-22, 51]. Fewer studies have evaluated ICC reliabilities. High ICC values of .84 and above have been reported in studies examining the standard FreeSurfer hippocampal protocol. Those studies recruited comparatively younger adults, ranging from 24 to 58 years old and consisted of either healthy controls or patients with major depressive disorder ($n = 41$) such as in Tae, et al. [52] or patients with bipolar disorder ($n = 93$) such as in Doring, et al. [20]. By contrast, we found lower ICCs in our memory clinic cohort. One possibility for this difference may be due to anatomical variability that is more likely to be present in an older clinical population. There have been studies that reported reduced reliability of FreeSurfer's hippocampal estimations in brains with more abnormalities and atrophy [22, 53].

Few studies have directly compared FreeSurfer and manual entorhinal cortex measurement. In a sample of 20 young adults with temporal lobe epilepsy and 20 healthy controls, Scanlon [54] reported higher ICC values of .8 with the standard entorhinal protocol. Weak correlations of between $-.27$ to $.37$ between FreeSurfer and manual entorhinal cortex volume measures have been reported by Lehmann, et al. [51] in a mixed sample of 20 AD and control patients. This study showed poor ICC coefficients for both FreeSurfer standard and ex vivo entorhinal protocols and correlations were all very poor and not significant. In the present study, the standard and newer ex vivo entorhinal protocols produced very poor reliability and agreement given the low ICC values and very wide limits of agreement. Perhaps even more surprising was that the newer ex vivo entorhinal protocol had a poorer reliability and agreement than the standard entorhinal protocol. Given the relatively small size of the present study, further validation studies of the newer FreeSurfer ex vivo entorhinal

Figure 3.4

Bland-Altman Plot of Entorhinal Percentage Difference Between Manual and FreeSurfer



Note. Bland-Altman plot of left and right entorhinal percentage difference volumes against corresponding average manual (M_{hv}) and FreeSurfer volumes as estimated using FS_{erc} and FS_{exvivo}. Solid lines represent the difference of the mean difference volumes between methods (see Table 3.1). Dashed lines represent the upper and lower 95% limits of agreement.

protocol that include comparison with manual measurement are necessary to clarify these discrepancies before any conclusions can be generalised.

It should be noted that ideally, the overall mean difference between volume estimates should be as close to zero as possible if the bias between the two methods is minimal; however, there are no statistical criteria for determining if a novel method can adequately substitute an existing method. According to Bland and Altman [45, 47], acceptability criteria for method substitution will vary according to the context and the clinical significance of the

method to be used in decision-making. In the clinical setting, methodological differences are likely to be acceptable if diagnostic accuracy is not affected overall by variability in measurement between and within individual patients. The current study was not sufficiently powered to allow us to attempt meaningful insights into the potential impact of automated volumetric assessments on the final clinical diagnosis but suggests that any bias in automated hippocampal volumetry may be acceptable while the results for non-manual analysis of entorhinal cortex appear less reassuring.

Method-specific differences between FreeSurfer and manual protocols may encompass several factors. Differences in atlas and landmark definitions across protocols constitute a source of measurement error [10, 29, 55, 56]. One of the largest sources of variance in the standard entorhinal protocol and the ex vivo validated protocol may be attributable to the difficulty in atlas definitions, given that the entorhinal cortex is a smaller structure with less distinct cytoarchitectonic boundaries on MRI and greater inter-individual variability in morphology [23, 57, 58].

In addition to the relatively small sample size of the present study, other possible limitations of this archival study are that the scans were obtained over a span of ten years on two different scanners and all the scans were obtained in the coronal plane non-isotropically, which is not the standard sagittally aligned isotropic acquisition as suggested by developers of FreeSurfer. Another related point would also be that some of the scans may have been acquired with different scanner parameters and there would also have been several scanner upgrades throughout the years, which would result in different scan qualities, which could possibly impact the segmentation results. Nevertheless, we visually inspected all the scans that were included to ensure acceptable scan quality for successful segmentation by FreeSurfer. More importantly, despite the limitations of the scanning parameters, the positive findings of this study, demonstrated that FreeSurfer's automated hippocampal protocol

remains robust with different scanners. This highlights the strong potential of FreeSurfer's automated hippocampal protocols to be applied into clinical practice in which routine MRI scans are rarely interrogated to research standards.

Another potential limitation to the study was that total intracranial volume (i.e. head size) was not utilised. While it has been found that total intracranial volume is correlated with hippocampal size and linear correction for head size have been commonly applied in the past, further investigations have found the relationship between head size and hippocampal size to be non-proportionate [59]. Although methods for generating total intracranial volume along with non-linear corrections are currently available, a lack of consensus regarding their application has generated comparability problems among measurement studies with different correction methods [60, 61].

From a clinical point of view, this pilot study has underscored the potential translational diagnostic utility of FreeSurfer automated segmentation analysis for hippocampal measurement based on routinely acquired diagnostic brain MRI scans, although more assessment and development appears necessary to advance the cautious optimism in relation to entorhinal cortex volumes. If streamlined into a routine pipeline, automated segmentation certainly offers higher efficiency in generating a large quantity of objective data compared to manual measurement from a given scan. Furthermore, processing time with successive iterations of FreeSurfer has improved, with the standard "recon-all" processing time for a single scan coming down from approximately 12-16 hours in the earlier 5.3 version to only 6-8 hours in the current 6.0 version. Although this processing is performed automatically it does not entirely remove the need for an expert operator who is not only familiar with running FreeSurfer but also has the necessary anatomical knowledge to ensure that scans are correctly processed. Nevertheless, such improvements in processing time once incorporated into a routine diagnostic pipeline make automated volumetry practicable. A

useful extension of the present study would be to include quantitative data and qualitative observations on scans with common artifacts such as movement, vascular and other lesions, to better understand segmentation failure rates and how tolerant FreeSurfer will be under “real world” or more realistic clinical conditions. A prospective study using mixed model analysis and a longitudinal framework would also be ideal in assessing FreeSurfer’s ability in detecting volumetric changes over a clinically relevant follow-up period. Understanding demographic and clinical characteristics in the sample is also likely to be an important factor underlying the varying rates of hippocampal volume under- and over-estimation by FreeSurfer.

In summary, good reliability and reasonable agreement of the FreeSurfer hippocampal protocol and the hippocampal subfields package with manual measurement was demonstrated in an independent sample of memory clinic patients. This demonstrates the robustness of FreeSurfer’s hippocampal segmentation algorithms in dealing with variable scanning parameters, scan quality and hippocampal structures, which suggest strong translational diagnostic potential of its use in a clinical setting. In contrast, both entorhinal protocols, standard and ex vivo, performed poorly in all reliability and agreement metrics and suggest that further refinements of these protocols are required before they could reliably transition to the analysis of routine diagnostic scans in the clinical setting.

Acknowledgements: Collins S.J. is supported in part by an NHMRC Practitioner Fellowship (#APP1105784). Fung Y.L. is supported by the Melbourne International Research Scholarship, a graduate research scholarship established by the University of Melbourne.

Conflicts of Interest: The authors have no conflict of interest to report.

References

- [1] McEvoy LK, Brewer JB (2010) Quantitative structural MRI for early detection of Alzheimer's disease. *Expert Review of Neurotherapeutics* **10**, 1675-1688.
- [2] Sutphen CL, Fagan AM, Holtzman DM (2014) Progress update: Fluid and imaging biomarkers in Alzheimer's disease. *Biological Psychiatry* **75**, 520-526.
- [3] Jack CR, Jr. (2011) Alliance for aging research AD biomarkers work group: structural MRI. *Neurobiology of Aging* **32**, S48-S57.
- [4] Whitwell JL, Josephs KA, Murray ME, Kantarci K, Przybelski SA, Weigand SD, Vemuri P, Senjem ML, Parisi JE, Knopman DS, Boeve BF, Petersen RC, Dickson DW, Jack CR, Jr. (2008) MRI correlates of neurofibrillary tangle pathology at autopsy - A voxel-based morphometry study. *Neurology* **71**, 743-749.
- [5] Vemuri P, Whitwell JL, Kantarci K, Josephs KA, Parisi JE, Shiung MS, Knopman DS, Boeve BF, Petersen RC, Dickson DW, Jack CR, Jr. (2008) Antemortem MRI based STructural Abnormality iNDex (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage. *Neuroimage* **42**, 559-567.
- [6] Anstey KJ, Maller JJ (2003) The role of volumetric MRI in understanding mild cognitive impairment and similar classifications. *Aging & Mental Health* **7**, 238-250.
- [7] deTedo-Morrell L, Stoub TR, Bulgakova M, Wilson RS, Bennett DA, Leurgans S, Wu J, Turner DA (2004) MRI-derived entorhinal volume is a good predictor of conversion from MCI to AD. *Neurobiology of Aging* **25**, 1197-1203.
- [8] Pennanen C, Kivipelto M, Tuomainen S, Hartikainen P, Hanninen T, Laakso MP, Hallikainen M, Vanhanen M, Nissinen A, Helkala EL, Vainio P, Vanninen R, Partanen K, Soininen H (2004) Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiology of Aging* **25**, 303-310.

- [9] Braak H, Braak E, Bohl J (1993) Staging of Alzheimer-related cortical destruction. *European Neurology* **33**, 403-408.
- [10] Mulder ER, de Jong RA, Knol DL, van Schijndel RA, Cover KS, Visser PJ, Barkhof F, Vrenken H, Alzheimers Dis Neuroimaging I (2014) Hippocampal volume change measurement: Quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST. *Neuroimage* **92**, 169-181.
- [11] Jack CR (2012) Alzheimer disease: New concepts on its neurobiology and the clinical role imaging will play. *Radiology* **263**, 343-360.
- [12] Gronenschild EHBM, Habets P, Jacobs HIL, 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**.
- [13] Gunter JL, Bernstein MA, Borowski BJ, Ward CP, Britson PJ, Felmlee JP, Schuff N, Weiner M, Jack CR (2009) Measurement of MRI scanner performance with the ADNI phantom. *Medical Physics* **36**, 2193-2205.
- [14] Simmons A, Westman E, Muehlboeck S, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Wahlund L-O, Soininen H, Lovestone S, Evans A, Spenger C, AddNeuroMed C (2009) MRI measures of Alzheimer's disease and the AddNeuroMed study In *Biomarkers in Brain Disease*, Lovestone S, ed., pp. 47-55.
- [15] Simmons A, Westman E, Muehlboeck S, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Wahlund L-O, Soininen H, Lovestone S, Evans A, Spenger C, AddNeuroMed C (2011) The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: Experience from the first 24 months. *International Journal of Geriatric Psychiatry* **26**, 75-82.

- [16] Westman E, Simmons A, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Weiner MW, Lovestone S, Spenger C, Wahlund L-O, AddNeuroMed C, Alzheimer's Dis Neuroimaging I (2011) AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. *Neuroimage* **58**, 818-828.
- [17] Wyman BT, Harvey DJ, Crawford K, Bemstein MA, Carmichael O, Cole PE, Crane PK, DeCarli C, Fox NC, Gunter JL, Hill D, Killiany RJ, Pachai C, Schwarz AJ, Schuff N, Senjem ML, Suhy J, Thompson PM, Weiner M, Jack CR, Alzheimer's Dis Neuroimaging I (2013) Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimers & Dementia* **9**, 332-337.
- [18] Fischl B (2012) FreeSurfer. *Neuroimage* **62**, 774-781.
- [19] Cherbuin N, Anstey KJ, Reglade-Meslin C, Sachdev PS (2009) In vivo hippocampal measurement and memory: A comparison of manual tracing and automated segmentation in a large community-based sample. *Plos One* **4**.
- [20] Doring TM, Kubo TTA, Cruz LCH, Jr., Juruena MF, Fainberg J, Domingues RC, Gasparetto EL (2011) Evaluation of hippocampal volume based on MR imaging in patients with bipolar affective disorder applying manual and automatic segmentation techniques. *Journal of Magnetic Resonance Imaging* **33**, 565-572.
- [21] Morey RA, Petty CM, Xu Y, Hayes JP, Wagner HR, II, Lewis DV, Labar KS, Styner M, McCarthy G (2009) A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage* **45**, 855-866.
- [22] Sanchez-Benavides G, Gomez-Anson B, Sainz A, Vives Y, Delfino M, Pena-Casanova J (2010) Manual validation of FreeSurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer Disease subjects. *Psychiatry Research-Neuroimaging* **181**, 219-225.

- [23] Fischl B, Stevens AA, Rajendran N, Yeo BTT, Greve DN, Van Leemput K, Polimeni JR, Kakunoori S, Buckner RL, Pacheco J, Salat DH, Melcher J, Frosch MP, Hyman BT, Grant PE, Rosen BR, van der Kouwe AJW, Wiggins GC, Wald LL, Augustinack JC (2009) Predicting the location of entorhinal cortex from MRI. *Neuroimage* **47**, 8-17.
- [24] Van Leemput K, Bakkour A, Benner T, Wiggins G, Wald LL, Augustinack J, Dickerson BC, Golland P, Fischl B (2009) Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus* **19**, 549-557.
- [25] Insausti R, Rincon M, Diaz-Lopez E, Artacho-Perula E, Mansilla F, Florensa J, Gonzalez-Moreno C, Alvarez-Linera J, Garcia S, Peraita H, Pais E, Insausti AM (2011) FreeSurfer automatic brain segmentation adaptation to medial temporal lobe structures: Volumetric assessment and diagnosis of mild cognitive impairment In *New Challenges on Bioinspired Applications: 4th International Work-Conference on the Interplay between Natural and Artificial Computation, Iwinac 2011, Part II*, Ferrandez JM, Sanchez JRA, DeLaPaz F, Toledo FJ, eds., pp. 112-119.
- [26] Wisse LEM, Biessels GJ, Geerlings MI (2014) A critical appraisal of the hippocampal subfield segmentation package in FreeSurfer. *Frontiers in Aging Neuroscience* **6**.
- [27] Iglesias JE, Augustinack JC, Nguyen K, Player CM, Player A, Wright M, Roy N, Frosch MP, McKee AC, Wald LL, Fischl B, Van Leemput K, Alzheimers Dis N (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.
- [28] Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkanen A (1998) MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *AJNR Am J Neuroradiol* **19**, 659-671.
- [29] Price CC, Wood MF, Leonard CM, Towler S, Ward J, Montijo H, Kellison I, Bowers D, Monk T, Newcomer W, Schmalfluss I (2010) Entorhinal cortex volume in older

- adults: Reliability and validity considerations for three published measurement protocols. *Journal of the International Neuropsychological Society* **16**, 846-855.
- [30] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341-355.
 - [31] Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968-980.
 - [32] Dale AM, Sereno MI (1993) Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction - a linear approach. *Journal of Cognitive Neuroscience* **5**, 162-176.
 - [33] Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis - I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179-194.
 - [34] Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis - II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195-207.
 - [35] Fischl B, Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11050-11055.
 - [36] Fischl B, Salat DH, van der Kouwe AJW, Makris N, Segonne F, Quinn BT, Dale AM (2004) Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23**, S69-S84.

- [37] Fischl B, Sereno MI, Tootell RBH, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping* **8**, 272-284.
- [38] Bonham LW, Geier EG, Fan CC, Leong JK, Besser L, Kukull WA, Kornak J, Andreassen OA, Schellenberg GD, Rosen HJ, Dillon WP, Hess CP, Miller BL, Dale AM, Desikan RS, Yokoyama JS (2016) Age-dependent effects of APOE epsilon 4 in preclinical Alzheimer's disease. *Annals of Clinical and Translational Neurology* **3**, 668-677.
- [39] Cook MJ, Fish DR, Shorvon SD, Straughan K, Stevens JM (1992) Hippocampal volumetric and morphometric studies in frontal and temporal and temporal-lobe epilepsy. *Brain* **115**, 1001-1015.
- [40] Meade CE, Bowden SC, Whelan G, Cook MJ (2008) Rhinal cortex asymmetries in patients with mesial temporal sclerosis. *Seizure-European Journal of Epilepsy* **17**, 234-246.
- [41] Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* **6**, 7.
- [42] McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychological Methods* **1**, 30-46.
- [43] Shrout PE, Fleiss JL (1979) Intraclass correlations - uses in assessing rater reliability. *Psychological Bulletin* **86**, 420-428.
- [44] Bland JM, Altman DG (1995) Comparing methods of measurement - why plotting difference against standard method is misleading. *Lancet* **346**, 1085-1087.
- [45] Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* **8**, 135-160.

- [46] Ludbrook J (2010) Confidence in Altman-Bland plots: A critical review of the method of differences. *Clinical and Experimental Pharmacology and Physiology* **37**, 143-149.
- [47] Bland JM, Altman DG (2003) Applying the right statistics: analyses of measurement studies. *Ultrasound in Obstetrics & Gynecology* **22**, 85-93.
- [48] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment - Clinical characterization and outcome. *Archives of Neurology* **56**, 303-308.
- [49] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Jr., Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **7**, 263-269.
- [50] de Flores R, La Joie R, Landeau B, Perrotin A, Mezenge F, de La Sayette V, Eustache F, Desgranges B, Chetelat G (2015) Effects of age and Alzheimer's disease on hippocampal subfields: Comparison between manual and FreeSurfer volumetry. *Human Brain Mapping* **36**, 463-474.
- [51] Lehmann M, Douiri A, Kim LG, Modat M, Chan D, Ourselin S, Barnes J, Fox NC (2010) Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. *Neuroimage* **49**, 2264-2274.
- [52] Tae WS, Kim SS, Lee KU, Nam E-C, Kim KW (2008) Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology* **50**, 569-581.

- [53] Oscar-Berman M, Song J (2011) Brain volumetric measures in alcoholics: a comparison of two segmentation methods. *Neuropsychiatric Disease and Treatment* **7**, 65-75.
- [54] Scanlon C (2010) in *Biomedical Science* National University of Ireland, Dublin, Ireland, p. 187.
- [55] Geuze E, Vermetten E, Bremner JD (2005) MR-based in vivo hippocampal volumetrics: 1. Review of methodologies currently employed. *Molecular Psychiatry* **10**, 147-159.
- [56] Konrad C, Ukas T, Nebel C, Arolt V, Toga AW, Narr KL (2009) Defining the human hippocampus in cerebral magnetic resonance images-An overview of current segmentation protocols. *Neuroimage* **47**, 1185-1195.
- [57] Kikinis R, Shenton ME, Gerig G, Martin J, Anderson M, Metcalf D, Guttmann CRG, McCarley RW, Lorensen W, Cline H, Jolesz FA (1992) Routine quantitative - analysis of brain and cerebrospinal-fluid spaces with MR imaging. *Jmri-Journal of Magnetic Resonance Imaging* **2**, 619-629.
- [58] Franko E, Maria Insausti A, Artacho-Perula E, Insausti R, Chavoix C (2014) Identification of the human medial temporal lobe regions on magnetic resonance images. *Human Brain Mapping* **35**, 248-256.
- [59] Barnes J, Ridgway GR, Bartlett J, Henley SMD, Lehmann M, Hobbs N, Clarkson MJ, MacManus DG, Ourselin S, Fox NC (2010) Head size, age and gender adjustment in MRI studies: a necessary nuisance? *Neuroimage* **53**, 1244-1255.
- [60] Nordenskjold R, Malmberg F, Larsson E-M, Simmons A, Brooks SJ, Lind L, Ahlstrom H, Johansson L, Kullberg J (2013) Intracranial volume estimated with commonly used methods could introduce bias in studies including brain volume measurements. *Neuroimage* **83**, 355-360.

- [61] Voevodskaya O, Simmons A, Nordenskjold R, Kullberg J, Ahlstrom H, Lind L, Wahlund L-O, Larsson E-M, Westman E, Alzheimer's Dis N (2014) The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Frontiers in Aging Neuroscience* **6**.

APPENDIX A

Letter to the Editor

Labelling of Statistical Output of Cytoarchitecture-Defined Variables in FreeSurfer Version 6

14 March 2019

In our recent paper, Fung et al. [1], we investigated the validity of automated FreeSurfer protocols in a group of routine clinical brain MRI scans. One structure of interest was the entorhinal cortex, in which we compared manual segmentation against automated segmentation generated by the ex vivo protocols of Fischl et al. [2] and Augustinack et al. [3] found in FreeSurfer version 6.

The standard volume statistics generated from this cytoarchitecturally defined protocol are found in the automatically generated stats files (BA_exvivo.stats and BA_exvivo.thresh.stats); however, during the auditing of our study data, we noted that there was a mismatch in the “annot labels” of the volume statistics with the “probabilistic labels”. Specifically, the structure labelled as “entorhinal cortex” in the annot stats file corresponded more closely to the perirhinal cortex probabilistic label, while the structure labelled as “perirhinal label” in the annot stats file corresponded more closely to the entorhinal cortex probabilistic label. Given our observation that the probabilistic labels were consistent with the expected anatomical locations of these two structures, as a workaround to the mislabelling in the stats files, the annot statistical output that was labelled as “perirhinal cortex” was used in our analysis in place of the FreeSurfer automated ex vivo “entorhinal” segmentation.

We contacted the developers of FreeSurfer and they have acknowledged that this systematic mislabelling was an issue previously unknown to them and agreed that the structures are mislabelled within the annot stats file. The developers have advised to use the work-around of the probabilistic labels for analysis while they work on an update to resolve the error with the annot labelling. At the time of writing we are unaware of available software updates to rectify this problem.

In sum, the aim of this letter is twofold: (1) to bring broad collegiate attention to the currently unaddressed anomalies in the automated stats output found in FreeSurfer v.6, and (2) to highlight the vital importance of researcher vigilance when evaluating the assumptions surrounding the outputs generated by automated algorithms, which are possibly too often taken at face value.

Yi Leng Fung¹, Kelly E.T. Ng¹, Simon J. Vogrin², Catherine Meade², Michael Ngo², Steven J. Collins^{2,3}, Stephen C. Bowden^{1,2}

¹ School of Psychological Sciences, University of Melbourne, Victoria, Australia

² Centre for Clinical Neuroscience and Neurological Research, St Vincent's Hospital, Victoria, Australia

³ Department of Medicine, University of Melbourne, Victoria, Australia

Correspondence to:

Yi Leng Fung, MPsych (Clinical Neuropsychology)/PhD Candidate, School of Psychological Sciences, University of Melbourne, Victoria, Australia. E-mail: yfung@student.unimelb.edu.au

References

[1] [Fung YL, Ng KET, Vogrin SJ, Meade C, Ngo M, Collins SJ, Bowden SC \(2019\) Comparative Utility of Manual versus Automated Segmentation of Hippocampus and Entorhinal Cortex Volumes in a Memory Clinic Sample. *J Alzheimers Dis*, doi: 10.3233/JAD-181172.](#)

[2] Fischl B, Stevens AA, Rajendran N, Yeo BT, Greve DN, Van Leemput K, Polimeni JR, Kakunoori S, Buckner RL, Pacheco J, Salat DH, Melcher J, Frosch MP, Hyman BT, Grant PE, Rosen BR, van der Kouwe AJ, Wiggins GC, Wald LL, Augustinack JC (2009) Predicting the location of entorhinal cortex from MRI. *Neuroimage* **47**, 8-17.

[3] Augustinack JC, Huber KE, Stevens AA, Roy M, Frosch MP, van der Kouwe AJ, Wald LL, Van Leemput K, McKee AC, Fischl B, Alzheimer's Disease Neuroimaging Initiative (2013) Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. *Neuroimage* **64**, 32-42.

Chapter 4

Manual Validation of Entorhinal and Transentorhinal *Ex Vivo* Protocol in FreeSurfer

This chapter is an extension of the published pilot study in Chapter 3. This chapter was specifically designed to investigate the validity of FreeSurfer segmentation of not only the entorhinal cortex, but also the transentorhinal cortex. This extension study was also designed to overcome the previous limitations by expanding the sample size, including control participants and investigating the performance of automated segmentation on both 1.5T and 3T scanners. The entire sample pool available for Chapters 4 – 6 consist of 40 healthy controls and 90 clinical patients. The data obtained from this large sample pool is new and not the same data used in Chapter 3. A subset of the entire sample pool, consisting of 40 healthy controls and 60 clinical patients, was used in the investigation found in this chapter.

Abstract

The entorhinal and transentorhinal cortices are among the earliest regions affected by Alzheimer's disease and they have been identified as potential structural biomarkers to facilitate early detection of the disease. While manual segmentation remains the method of choice in quantifying volumes, advances in automated segmentation provides an appealing alternative due to its efficiency and relative ease of use. There have been recent developments in cytoarchitecturally-validated *ex vivo* protocols for localisation of the entorhinal and transentorhinal cortices in FreeSurfer, but few studies have assessed the validity and practical utility of FreeSurfer in clinical settings. This study compared the entorhinal and transentorhinal automated estimates based on the *ex vivo* protocols against manual estimations generated from 1.5T and 3T MRI scanners. In both scan conditions, there was a control sample ($n = 40$) and a clinical sample ($n = 30$) that consisted of individuals above 50 years of age. Consistency correlation measures and levels of agreement were suboptimal for the entorhinal volumes and poor for the transentorhinal volumes, across all samples and magnetic field strengths. These findings suggest that further refinements are required to improve segmentation accuracy for both these structures before they can be used for diagnostic purposes in clinical evaluations.

Keywords: entorhinal, transentorhinal, FreeSurfer, automated, manual, volume, segmentation, validation, reliability

Manual Validation of Entorhinal and Transentorhinal

Ex Vivo Protocol in FreeSurfer

Alzheimer's disease (AD) is the most common form of progressive neurodegenerative disorder. The prevalence of AD increases with advancing age and much of society is subject to the growing burden of this disease as life expectancy rises (Alzheimer's Association, 2018). The development of AD is highly protracted, including an extended pre-clinical phase (Sperling et al., 2011). Although the clinical picture of AD can appear to emerge aggressively, mounting evidence suggests that cognitive symptoms gradually evolve over several years before dementia is diagnosed and AD-related pathological process may remain dormant decades before crossing the threshold into the symptomatic stages (Braak & Del Tredici, 2012, 2015b; Sperling et al., 2011; Wilson et al., 2012).

Medial Temporal Structures and Alzheimer's Disease Pathology

The human cerebral cortex can be divided into two major portions, the isocortex and the allocortex. The isocortex, which is also called the neocortex and makes up about 96% of the cerebral cortex, is predominantly made up of the frontal and temporal lobes, followed by the parietal and occipital lobes (Mai & Paxinos, 2011). One important division of the allocortex is the archicortex. Although the archicortex makes up only about 3.5% of the cortical surface, it includes important structures in the limbic circuit that functions as the centre for memory and learning (Braak & Del Tredici, 2015b; Mai & Paxinos, 2011). The limbic regions receive unfiltered input from the neocortex, and in order to categorise and retain the data in memory, the incoming data is reorganised within the limbic circuit before output is generated and relayed back to the neocortex (Braak & Del Tredici, 2015b).

The components of the limbic circuit include the hippocampus, the entorhinal cortex and the transentorhinal cortex (Braak & Del Tredici, 2015b). The transentorhinal and entorhinal regions exert major influences on the afferent stimulation that the hippocampus

receives from the neocortex. The transentorhinal region relays the data received from the neocortex into the entorhinal region, which in turn provides the majority of the convergent neocortical input to the hippocampal formation (Braak & Del Tredici, 2015a; Ranganath & Ritchey, 2012; Squire & Zola-Morgan, 1991; van Strien, Cappaert, & Witter, 2009; Young, Otto, Fox, & Eichenbaum, 1997). The transentorhinal and entorhinal cortices thereby form the interface between the neocortex and the hippocampal formation (Braak & Del Tredici, 2015a).

In AD, these limbic regions are one of the earliest regions affected by AD-related pathological lesions. One of the hallmark neuropathological markers of AD is the aggregation of hyperphosphorylated tau protein within the cell soma, known as neurofibrillary tangles (NFTs). Progression of the NFT distribution in AD is remarkably consistent in its temporal sequence across individuals, hence allowing for neuropathological staging of the AD process. In the seminal paper by Braak and Braak (1991), they observed that NFTs develop initially in the transentorhinal cortex (NFT stage I) before progressing into the superficial layers of the entorhinal cortex during NFT stage II. In stages I and II, there is no overt impact on cognition, but the early clinical picture of AD typically emerges in stages III and IV (Braak & Del Tredici, 2015a).

During stage III, the pathology is still mainly restricted within the transentorhinal cortex, entorhinal cortex and hippocampus, with minor involvement of the neighbouring temporal neocortex (Braak, Alafuzoff, Arzberger, Kretschmar, & Del Tredici, 2006; Braak & Braak, 1995). In stage IV, the density of NFTs within those regions become more prominent and extensive, and the frontal lobes begin to be implicated in the NFT propagation process (Braak et al., 2006; Braak & Braak, 1995). People with mild cognitive impairment (MCI) due to AD are commonly identified around stages III and IV, with one autopsy study reporting a mean Braak NFT stage of 3.3 in the MCI patients (Markesbery et al., 2006). As

the density of abnormal tau protein increases in the transentorhinal and entorhinal regions during NFT stages III and IV, the connectivity with the hippocampus becomes disrupted. This eventually leads to the isolation of the hippocampus from the neocortex and the ensuing curtailment of memory and cognitive functions that emerges in and characterises AD (Braak & Del Tredici, 2015a).

Medial Temporal Atrophy as a Biomarker for Alzheimer's Disease

Although molecular evidence of AD-related neuropathology serves as the ideal biomarker for detecting AD, structural surrogate biomarkers that are topographically consistent with AD-related neurodegeneration also provide good supportive evidence for the disease (Dubois, Hampel, et al., 2016; Jack et al., 2015; McKhann et al., 2011; Tan, Yu, & Tan, 2014). The characteristic deposition of NFT has been found to be closely correlate with eventual neuronal loss and numerous studies have observed that appropriate atrophy patterns in the medial temporal structures corresponded well to these NFT-related neuronal changes (Arendt, Bruckner, Morawski, Jager, & Gertz, 2015; Bejanin et al., 2017; Giannakopoulos et al., 2003; Maass et al., 2018; G. Novak & Einstein, 2013).

There is a plethora of neuroimaging studies that have reported significant reductions in entorhinal volumes in people with AD relative to healthy participants as assessed through automated (Ramos Bernardes da Silva Filho et al., 2017; Sone et al., 2017), semi-automated (Dickerson et al., 2009), as well as manual volume estimation (Mueller et al., 2010; Xu et al., 2000). Studies that used manual methodology have also found that the entorhinal volumes were similarly reduced in people with MCI when compared against controls, although the degree of volumetric atrophy was similar between people with early AD and MCI (Stoub, Rogalski, Leurgans, Bennett, & deToledo-Morrell, 2010; Xu et al., 2000).

Even though the above findings suggest that a cross-sectional measure of entorhinal atrophy provides limited utility in differentiating MCI from AD, the longitudinal rates of

atrophy can provide additional information on the likelihood of conversion from MCI to AD. A five-year longitudinal study by Stoub et al. (2010) demonstrated via manual volumetry that the annual hippocampal atrophy rates between MCI participants who experienced cognitive decline on follow-up and MCI participants who remained cognitively stable on follow-up were similar (9.5% and 8.6% respectively). On the other hand, although the rates of entorhinal atrophy between the cognitively stable MCI participants and cognitively stable normal controls were similar (3.3% and 2.7% respectively), the MCI participants who later converted to AD during the five-year follow-up period showed a significantly steeper rate of entorhinal atrophy at 4.7% (Stoub et al., 2010).

Examination of the relationship between transentorhinal cortical volume to cognition and development of AD is a more recent development. A recent three-year longitudinal study by Xie et al. (2018) found significant positive correlations between tau burden and rates of atrophy in the medial temporal regions. In particular, the most prominent atrophy, assessed through cortical thickness that was determined through a multi-atlas segmentation algorithm, was seen in the transentorhinal cortex, which is the earliest cortical site of NFT formation as identified by Braak and Braak (1991). A longitudinal study by Kulason et al. (2018) found that compared to controls, people with MCI had significantly smaller entorhinal and transentorhinal manually-determined volumes at baseline and demonstrated greater rates of atrophy in those areas. Kulason and colleagues (2018) performed a discriminant function analysis that showed that although a combination of the entorhinal and transentorhinal volumes did not add additional discriminant power, transentorhinal thickness by itself had the largest sensitivity and specificity for differentiating MCI from controls.

As the above evidence suggests, there are clear benefits to including the entorhinal and transentorhinal cortical measures as supportive structural biomarkers in the detection of early AD. One of the advantages of using structural biomarkers is that cortical measures such

as thickness and volumes can be determined through brain scans obtained using magnetic resonance imaging (MRI). The use of MRI is non-invasive, widely available and frequently used in routine clinical practice to rule out other causes of clinical impairment. The availability of this neuroimaging capability provides an existing platform for potentially translating research findings into clinical practice.

Volumetric Quantification of the Medial Temporal Structures

The ardent pursuit of *in vivo* volumetric quantification of the medial temporal structures using MRIs can be appreciated through the numerous manual segmentation protocols that have been developed in recent years. Earlier manual protocols have only included the entorhinal cortex (Goncharova et al., 2001; Honeycutt et al., 1998; Insausti et al., 1998; Meade et al., 2008; Pruessner et al., 2002). Development of transentorhinal manual protocols have been hindered by inadequate knowledge about its anatomical boundaries and neuroanatomical variability. However, this knowledge gap has been elucidated by recent neuroanatomical research (Ding & Van Hoesen, 2010; Taylor & Probst, 2008) and development of transentorhinal cortex manual segmentation protocols have subsequently followed (Berron et al., 2017; Kivisaari et al., 2013).

Manual delineation is typically regarded as the gold standard in quantitative volumetric studies (Lehmann et al., 2010; Sanchez-Benavides et al., 2010; Teipel et al., 2010; Wilke, de Haan, Juenger, & Karnath, 2011). Manual segmentation allows for flexible adaptation to variations in the structures or scan quality differences (Jack, 2012). Nonetheless, the labour-intensive nature of manual segmentation limits its practical utility in large research and clinical situations (Jack, Barkhof, et al., 2011; McEvoy & Brewer, 2010). There is also a need for trained expertise in neuroanatomy and manual segmentation in the region of interest, an expertise that is not readily available (Jack, 2012). Furthermore, there may be some concerns with inter- and intra-rater variability, which limits the ability to

generate standardised normative data and derive meaningful interpretation of the values (Despotovic, Goossens, & Philips, 2015; Jack, 2012).

Advancements in automated segmentation capabilities provide an opportunity to overcome the limitations of manual segmentation (Bigler, 2015; Jack, 2012). Automated segmentation allows reduction in time and labour investment through the total, or partial removal of the need for manual intervention (Bigler, 2015; Jack, 2012). Automated quantitative techniques also provide an opportunity for data standardisation through reproducible findings across different settings (Bigler, 2015; Jack, 2012).

FreeSurfer (v6.0) is one of the most commonly used automated image analysis packages that is publicly accessible from the FreeSurfer website (<http://surfer.nmr.mgh.harvard.edu>). To address the crucial need for reliable location and quantification of the entorhinal and transentorhinal cortex, particularly due to the potential roles of these structures in facilitating early detection of AD, additional quantification protocols were introduced into FreeSurfer. Both the new entorhinal and transentorhinal probabilistic mapping pipelines were developed and validated based on cytoarchitectural features, using ultra-high resolution 7T *ex vivo* MRI scans (Augustinack et al., 2013; Fischl et al., 2009). To develop the entorhinal and transentorhinal *ex vivo* protocols, high resolution imaging was completed on autopsied brain samples on a 7T Siemens scanner and subsequently the architectonic features of these structures were labelled on these *ex vivo* MRI scans and mapped on to surface models (Augustinack et al., 2013; Fischl et al., 2009).

Entorhinal and transentorhinal cortex are known to be challenging structures to delineate due to their small size, anatomical variability and proximity to the tentorium cerebelli and oculomotor nerve (Insausti et al., 1998; Wolk et al., 2017; Yushkevich, Pluta, et al., 2015). This complexity therefore makes it critical to evaluate the usefulness of automated algorithms to successfully localise and delineate these challenging structures without

significant manual intervention. Assessment of automated FreeSurfer measures of the transentorhinal cortex appear to have been largely overlooked, with an absence of validation studies on this structure. There is also a paucity of studies assessing the ability of FreeSurfer to reliably segment the entorhinal cortex. In one of the few studies available, McCarthy et al. (2015) compared three edited and unedited entorhinal measures, namely surface area, white matter volume and cortical thickness, estimated using Desikan-Killiany gyral-based atlas in FreeSurfer (v.5.3). Overall, the authors observed less than ideal intra-class correlation coefficients (ICC) on the 1.5T scans, reporting an ICC of .52 for surface area estimation and an ICC of .60 on white matter volume estimation. Compared to surface area and white matter volume ICCs, the ICC for cortical thickness on the 1.5T scans were better at .81; however, the authors noted that even though the ICC for cortical thickness on the 3T scans (ICC = .92) were higher than that found in the 1.5T scans, there were significant differences between the edited and unedited measures of entorhinal cortical thickness on those 3T scans ($p < .001$) that were not present in the 1.5T scans. While disappointing, their findings may not be entirely surprising. McCarthy and colleagues (2015) suggested that their findings highlighted that the entorhinal cortex is a common region for imaging artefacts, which is consistent with other studies that have suggested that the temporal regions are particularly troublesome areas for FreeSurfer to measure (Desikan et al., 2010; Desikan et al., 2006). Moreover, the variability of the sulcal patterns commonly found in the medial temporal lobe are not represented in atlases registered on a single template volume, such as the Desikan-Killiany atlas.

The new entorhinal and transentorhinal *ex vivo* probabilistic mapping approaches were developed to overcome the limitations of single template registration atlases through statistical averaging of multiple sulcal patterns (Augustinack et al., 2013; Fischl et al., 2009). As detailed in the published study in Chapter 3, however a validation study that examined the

entorhinal volumetric outputs using FreeSurfer (v.6.0) *ex vivo* protocol found that there was poor agreement between manual and automated measures (Fung et al., 2019a). The reported ICCs were .37 and .11 on the left and right hemispheres respectively, and very wide limits of agreement on the Bland-Altman plots. It is noteworthy that a control group was not included in the study and only 1.5T scans were used. Therefore, it was possible that the results were impacted by scan resolution, as well as morphological variabilities that were likely more frequently present in clinical samples.

Aims of Current Study

Therefore, the goal of this study was to evaluate the utility of FreeSurfer to reliably segment the entorhinal and transentorhinal cortex using the new *ex vivo* pipelines compared to manual estimations as a reference standard. The performance was assessed on different scanner strengths, 1.5T and 3T scans, in order to understand the impact of different scan resolutions on segmentation accuracy. Besides using scans obtained from healthy older adults, the capabilities of the new segmentation modules were also tested under clinical conditions. Scans that were obtained as part of a routine dementia workup were used, as they represent “real life” conditions in which more variability in scan quality and structural morphology would be expected.

Method

Study Population

Two samples of clinical scans ($n = 60$) were drawn retrospectively from an archival database of patients who were referred to the Cognitive Clinic at St Vincent’s Hospital Melbourne between 2006 – 2018. Patients were consecutive samples for whom all data were available. As part of their diagnostic dementia clinical workup, they undertook a comprehensive neuropsychological assessment that included the full administration of the

complete WMS-III/IV whenever possible. All patients were diagnosed as having mild cognitive impairment or Alzheimer's disease dementia according to recommended diagnostic criteria (Albert et al., 2011; McKhann et al., 2011). Criteria for mild cognitive impairment included evidence of a change in cognition that was not attributable to the medical conditions, coupled with objective evidence of disproportionate cognitive impairment in one or more domains when weighted against the person's age and educational background (Albert et al., 2011). The differentiation of mild cognitive impairment from dementia rested on a clinical judgement of the presence of significant interference to daily function as a result of the cognitive impairment for a clinical diagnosis of dementia (McKhann et al., 2011).

Although the criteria for MCI and Alzheimer's disease dementia have been updated between the years of retrospective data collection, the clinical diagnostic criteria have remained similar. The only differences in terms of the diagnosis of mild cognitive impairment was that this study did not differentiate between single- and multi-domain MCI, which was a distinction previously laid out in the earlier diagnostic criteria (Petersen et al., 1999). Patients with Alzheimer's disease dementia in this study was classified based on the clinical diagnosis guidelines that were basically the same between the original 1984 and revised 2011 guidelines (McKhann et al., 1984; McKhann et al., 2011). The clinical diagnostic criteria for mild cognitive impairment and AD are also in line respectively with the current diagnostic guidelines for mild neurocognitive disorder and major neurocognitive disorder that are laid out in the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013).

All participants who did not meet the diagnostic criteria, were considered as having subjective memory concerns. Patients with a prior or current history of mild to moderate anxiety or depression were retained given the naturally increased prevalence of mood-related difficulties in this patient population.

Clinical 1.5T sample: Patients from this sample underwent a routine MRI head scan on a 1.5T scanner between the years 2006 – 2014 ($n = 30$). Sixteen patients in this sample had subjective memory concerns, five were diagnosed with mild cognitive impairment and nine were diagnosed with AD.

Clinical 3T sample: Patients in this sample underwent a routine MRI head scan on a 3T scanner between the years 2013 – 2018 ($n = 30$). Seventeen patients in this sample were diagnosed to have subjective memory concerns, six were diagnosed with mild cognitive impairment and seven were diagnosed with AD.

General exclusionary criteria for both clinical samples were being below 50 years of age and having a significant history of an acquired head injury with persisting cognitive deficits.

Control sample: A group of healthy control participants was recruited for this study from the years 2017 – 2018 ($n = 40$). All control participants underwent MRI head scans on both a 1.5T and 3T scanner on the same day, and undertook a comprehensive neuropsychological assessment, which included a complete administration of the WMS-IV.

General exclusionary criteria for the control sample were being below 50 years of age, having contraindications to receiving an MRI, having any current medical or neurological condition that would have impact on cognition and any history of an acquired head injury with persisting cognitive deficits. All control participants included in this study had no unusual findings on their cognitive assessments and MRIs. Control participants with a prior or current history of mild to moderate anxiety or depression were retained.

The basic demographic details of the age and gender distribution of the clinical and control samples can be found in Table 4.1. This study was approved by the St. Vincent's Hospital (Melbourne) Human Research Ethics Committee (HREC-A 057-15).

Table 4.1*Summary of Demographics Within in Each Sample*

	Control (<i>n</i> = 40)	Clinical		Chi-square		ANOVA	
		1.5T (<i>n</i> = 30)	3T (<i>n</i> = 30)	χ^2	<i>p</i> - value	<i>F</i> - ratio	<i>p</i> - value
Age	67.8 (5.4)	66.9 (10.7)	68.9 (7.2)	n/a	n/a	0.49	.61
Gender (female)	30	11	17	10.37	.006	n/a	n/a

Note. Standard deviations of mean age are provided in the parenthesis.

MRI Acquisition

All T1-weighted scans were acquired coronally along the hippocampal axis with 3D-sequenced MPRAGE. The 1.5T scans were acquired on a Siemens Symphony scanner and scan parameters are as follows: CP head array coil, TR = 1530, TE = 3.41, FOV = 156mm * 250mm, flip angle = 8°, voxel size 1 * 1 * 1.4mm³, matrix 256 * 160. The 3T scans were acquired on a Siemens Skyra scanner and scan parameters are as follows: 64-channel head coil, TR = 2200, TE = 2.03, FOV = 230 * 230, flip angle = 8°, voxel size = 0.9 * 0.9 * 1mm³, matrix 256 * 256.

There were occasional deviations from these exact parameters that resulted in some differences in total slice numbers, but in-plane resolution did not change. Both clinical and control scans were acquired using the standard clinical protocol routinely used at St Vincent's Hospital Cognitive Clinic. An additional, consecutive T1-weighted MPRAGE was also acquired on the 3T scanner for the control sample with the same scan parameters.

Volumetric Procedures

Automated segmentation. Automated segmentation of scans was completed on a Linux workstation over approximately six to eight hours of analysis per scan on one core. Automated labelling for the entorhinal and transentorhinal volumes was generated using FreeSurfer (v.6.0).

The FreeSurfer cortical reconstruction process was performed on raw DICOM images using the “recon-all” command. Details of this automated process are described in Fischl et al. (2002). Recon-all was executed using a single T1 scan, except the Control Skyra sample, in which the reconstruction process was performed on two 3T T1 scans.

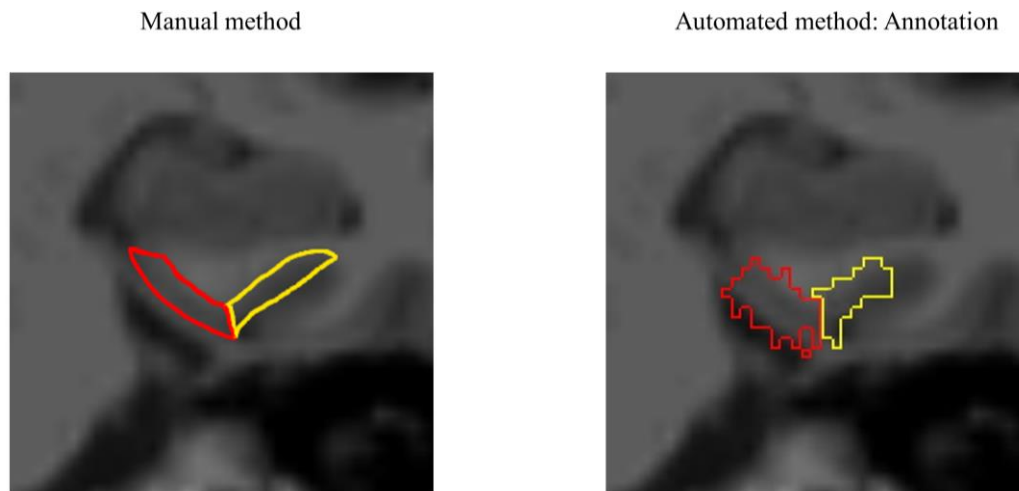
Volume statistics for the entorhinal cortex and transentorhinal cortex were based on the outputs generated by cytoarchitecturally-defined probabilistic labels, a process that is included as part of the standard pipeline (Augustinack et al., 2013; Fischl et al., 2009). Values of the grey volumes were extracted using three different methods detailed below.

Volumes extracted from the annotation files. As part of the standard pipeline, anatomical statistics for the thresholded cytoarchitecturally-defined structures were generated by default from the annotation files and the unique values could be found in the automatically-generated stats file under the corresponding names, “entorhinal cortex” and “perirhinal cortex”, the “perirhinal cortex” being analogous to the transentorhinal cortex. The volumes generated from the annotation files are referred to as *annotation* in this study.

Worth noting was the information that, at the time of this writing, there was mislabelling of the structures in the Freesurfer stats files, in which the details of this issues are described in the appendix at the end of Chapter 3 (Fung et al., 2019b). The “entorhinal cortex” was labelled as the more lateral structure and the “perirhinal cortex” was labelled as the more medial structure, although the reverse would be a more accurate representation of the typical anatomical location. Therefore, to work around the labelling error, the values labelled under “perirhinal cortex” was used in the analysis of the entorhinal segmentation while the values labelled under “entorhinal cortex” was used in the analysis of the transentorhinal segmentation. An example of the boundaries of the annotation labels is shown in Figure 4.1.

Figure 4.1

Example of the Boundaries of Entorhinal and Transentorhinal Cortices Based on Manual and Automated Methods



Note. The red outline represents the entorhinal cortex and the yellow outline represents the transentorhinal cortex. Automated segmentation was completed using FreeSurfer (v. 6.0).

Volumes extracted from the probabilistic label files. Total grey matter volumes of the entorhinal and transentorhinal cortices were also calculated from the respective left and right thresholded probabilistic labels using the command “`mris_anatomical_stats`”. All volumes generated using this method are referred to as *label-plus* in this study.

In contrast to the annotation files that only allocate one vertex to one structure, vertices defined by the probabilistic labels can belong to one or more structures. In other words, the volumes generated from the probabilistic labels using this method may not be unique to the named labels. As there was substantial overlap between the probabilistic labels of the entorhinal and transentorhinal cortex, the volumes generated from the label volumes was effectively a non-unique representation of these two structures. Of note, there appeared to be an over-representation of the entorhinal cortex, in which a portion of the transentorhinal cortex overlaps the entorhinal cortex (Figure 4.2). Therefore, a considerable portion of the

entorhinal cortical volume generated from the probabilistic labels would include the transentorhinal cortical volume. Additional steps were taken to circumvent this non-unique entorhinal volume. The entorhinal and transentorhinal label files were first merged to form a single label (Figure 4.2). The entorhinal volume was then calculated by subtracting the thresholded transentorhinal probabilistic label volume from the newly merged label volume, to give an additional, distinct entorhinal volume. Entorhinal volumes generated using this method are referred to as *label-minus* in this study.

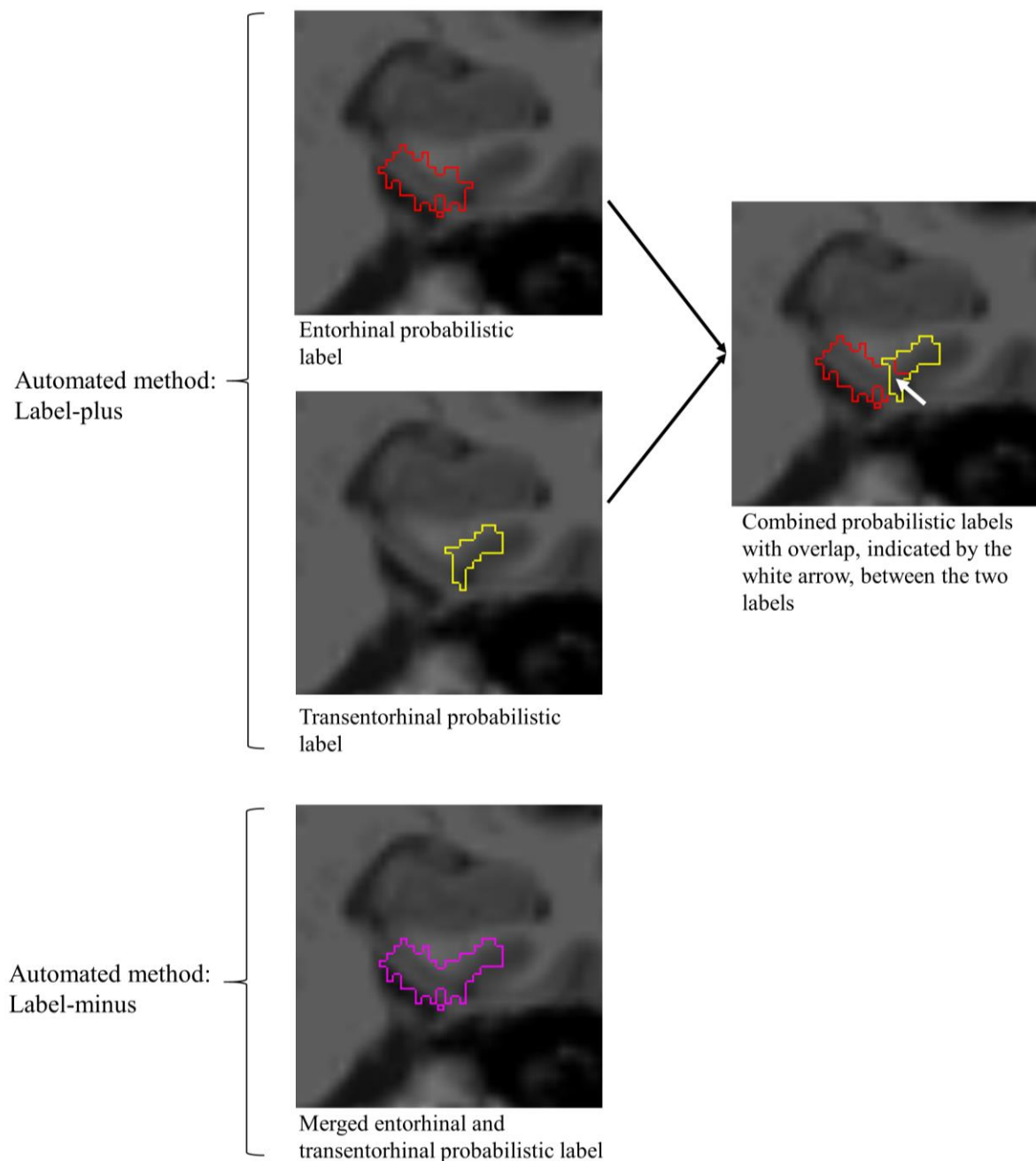
Paired-sample *t*-test found no differences between the transentorhinal volume generated from the annotation files and probabilistic label files, with a standard error of zero in both hemispheres and both scanners. These results indicated that the two automated volumes were identical across all subjects regardless of hemisphere and scanner field strength. Hence, only the *annotation* transentorhinal volumes were used for analysis in the study.

Manual segmentation. Volumes of the entorhinal cortex and transentorhinal cortex were determined by manual delineation on the coronal T1-weighted slices using the ROI measure on Analyze version 12.0, which was released in 2017 (Brain Imaging Resource, Mayo Foundation, Rochester, MN, USA). To prepare for manual tracing, raw DICOM files were converted into volumes of 0.3*0.3*1.4mm voxel size using cubic spline and saved into Analyze format. The x and y dimensions were converted to 0.3 voxels to improve the resolution for manual tracing for both 1.5T and 3T scan, and slice thickness for 3T scans was converted into 1.4mm for comparability with the 1.5T scans. The original slice thickness of 1.4mm was preserved in the 1.5T scans.

The entorhinal and transentorhinal cortices were traced together on the same coronal slice according to the protocol described in Kivisaari et al. (2013). According to this protocol by Kivisaari et al. (2013), the anterior border of the entorhinal and transentorhinal cortices

Figure 4.2

Example of the Automated Segmentation Boundaries Using Probabilistic Labels



Note. The volumes for label-plus entorhinal and transentorhinal volumes were generated based on the respective probabilistic labels. As shown, there was an overlap between the entorhinal (red outline) and transentorhinal (yellow outline) probabilistic labels. An additional label (purple outline) was created by merging the entorhinal and transentorhinal probabilistic labels. The volume generated from the transentorhinal probabilistic label was then subtracted from this merged label to create an additional label-minus entorhinal volume.

were defined as “2 mm posterior to the first anterior slice where the white matter of the limen

insulae is visible” (p. 253) while the posterior border was defined as “1mm posterior to the last slice containing the apex of the intra-limbic gyrus” (p. 254). Due to differences in slice thickness, the anterior limit was chosen to be one slice caudal to the first anterior slice in which the white matter of the limen insulae was visible, and the posterior limit was chosen to be one slice caudal to the last slice where the intra-limbic gyrus was visible. These adjustments to the anterior and posterior limits were done to avoid underestimation of the structures. Figure 4.1 shows an example of the manual delineation of the entorhinal and transentorhinal cortices.

Manual segmentation of the entorhinal and transentorhinal cortices was completed by a single operator who was blinded to their diagnoses. Two-way, absolute agreement intra-class correlation coefficient (ICC) was used to assess intra-rater reliability on 10 randomly selected scans from each clinical sample (Cicchetti, 1994; McGraw & Wong, 1996; Shrout & Fleiss, 1979). In the clinical 1.5T sample ($n = 10$, mean age = 71.5 ± 8.9 years), intra-rater reliability found for manual entorhinal segmentation was moderate to excellent on the left, single-measure ICC = .90 (95% CI: .66, .97), and good to excellent on the right, single-measure ICC = .95 (95% CI: .80, .99). Intra-rater reliability for manual transentorhinal segmentation was good to excellent on the left, single-measure ICC = .96 (95% CI: .84, .99) and excellent on the right, single-measure ICC = .98 (95% CI: .93, .99). Single-measure ICC values found in the clinical 3T sample ($n = 10$, mean age = 69.4 ± 6.9 years) were good to excellent on the left entorhinal manual estimations, ICC = .97 (95% CI: .88, .99), and excellent on the right entorhinal manual estimations, ICC = .98 (95% CI: .90, .99). Intra-rater reliability for manual transentorhinal segmentation was excellent on the left, single-measure ICC = .98 (95% CI: .90, .99), and good to excellent on the right, single-measure ICC = .97 (95% CI: .87, .99).

Statistical Analysis

Statistical analysis was conducted using SPSS (v. 25, released 2017). To test for possible differences in estimated volumes from scans obtained using different field strength, paired-sample *t*-test was used to compare the volume estimates generated from the 1.5T and 3T scans in the control sample, and effect sizes were quantified using Cohen's *d* (Cohen, 1988). Absolute agreement between the volumes obtained from these different scanner strengths were assessed using two-way mixed absolute single-measure ICC.

Reliability between manual and automated volume measurements for each region of interest was assessed using two-way mixed, consistency and absolute agreement ICCs. Both consistency and absolute agreement analysis were performed, as given the paucity of volumetric validation studies on FreeSurfer *ex vivo* entorhinal and transentorhinal segmentation algorithms, the expected variation between the manual and automated methods is unclear. All ICCs reported were based on single-measure ICC values.

A percentage difference Bland-Altman plot, plotted against the mean of both measurements, was used to assess for the level of agreement between each method of volume estimation (Bland & Altman, 1995, 1999). Percentage difference was calculated using the following: $(\text{manual estimate} - \text{automated estimate}) / \text{mean of volume estimates} * 100$. The 95% limits of agreement were plotted as a visual representation of the variability of individual differences across the two methods. Regression lines, with the 95% confidence interval, were included in the Bland-Altman plots to assess for proportional bias.

Results

Descriptive Statistics

There were no significant differences in age between the control sample, clinical 1.5T sample and clinical 3T sample ($p = .61$). Gender distribution between the three samples were significantly different ($p = .006$). Details of the statistical analysis can be found in Table 4.1.

Means and standard deviations of the entorhinal and transentorhinal volume measurements in each sample, namely control 1.5T, control 3T, clinical 1.5T and clinical 3T, can be found in Table 4.2 and Table 4.3. A graphical representation of the values are presented in Figure 4.3 and Figure 4.4. An ANOVA with Tukey's post-hoc testing compared methods of volume estimate, namely manual, *annotation*, *label-plus* and *label-minus*, showed that all automated entorhinal estimates were significantly larger than manual estimates (Table 4.2, Figure 4.3). As gender distribution was significantly different among the different samples, an ANCOVA was conducted as a secondary analysis to covary out the effects of gender differences. In all comparisons, the volume estimates were significantly different among the sample and Bonferroni post hoc analysis showed the same pattern of difference as that seen in the ANOVA analysis, in which automated entorhinal estimates were significantly larger than manual estimates (Table 4.2). Similarly, the comparison between the manual and automated estimations of transentorhinal volume was made using ANOVA and ANCOVA. In both analyses, the automated transentorhinal volumes were significantly larger than manual transentorhinal volumes in all comparisons (Table 4.3, Figure 4.4).

Paired-sample *t*-test was used to test for hemispheric differences within each sample. Results showed that the left entorhinal cortex was smaller than the right in all samples and across all methods of volume estimates, except in clinical 3T manual estimates in which no significant differences were found (Table 4.4). There were no significant differences between the manual estimates of the left and right transentorhinal volumes (Table 4.4). However, in

Table 4.2

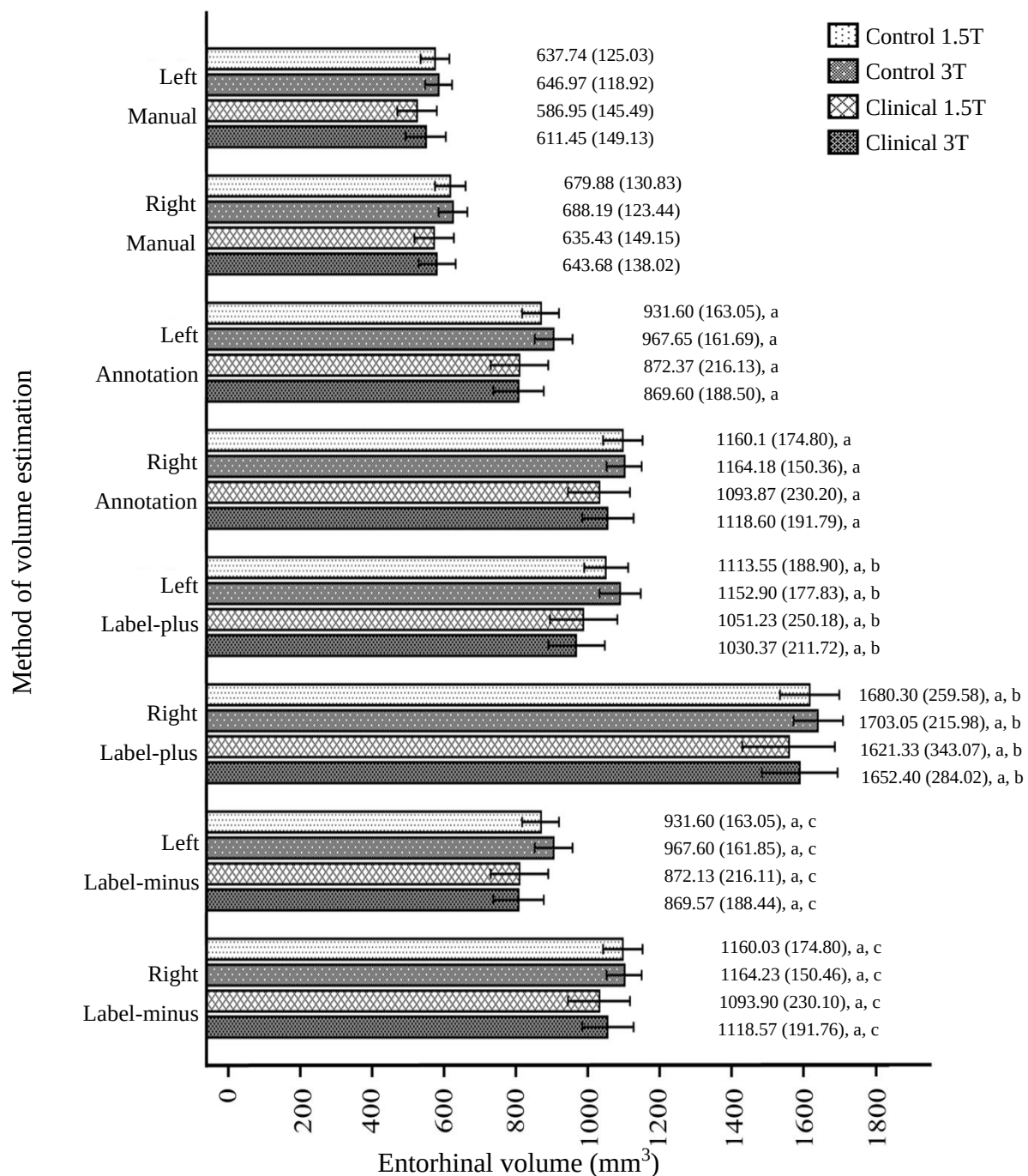
Means and Standard Deviations of the Entorhinal Volumes and Comparisons of these Volumes Across the Different Methods of Volume Estimation

	Method of volume estimation				ANOVA			ANCOVA		
	Manual	Annotation	Label-plus	Label-minus	<i>F</i> -ratio	<i>p</i> -value	η^2	<i>F</i> -ratio	<i>p</i> -value	η^2
Control 1.5T										
Left	637.74 (125.03)	931.60 (163.05)	1113.55 (188.90)	931.60 (163.05)	59.38	< .001	.53	59.33	< .001	.53
Right	679.88 (130.83)	1160.1 (174.80)	1680.30 (259.58)	1160.03 (174.80)	183.44	< .001	.78	183.02	< .001	.78
Control 3T										
Left	646.97 (118.92)	967.65 (161.69)	1152.90 (177.83)	967.60 (161.85)	72.07	< .001	.58	73.01	< .001	.59
Right	688.19 (123.44)	1164.18 (150.36)	1703.05 (215.98)	1164.23 (150.46)	256.86	< .001	.83	260.82	< .001	.83
Clinical 1.5T										
Left	586.95 (145.49)	872.37 (216.13)	1051.23 (250.18)	872.13 (216.11)	24.97	< .001	.39	28.02	< .001	.42
Right	635.43 (149.15)	1093.87 (230.20)	1621.33 (343.07)	1093.90 (230.10)	79.26	< .001	.67	86.20	< .001	.69
Clinical 3T										
Left	611.45 (149.13)	869.60 (188.50)	1030.37 (211.72)	869.57 (188.44)	26.10	< .001	.40	25.87	< .001	.40
Right	643.68 (138.02)	1118.60 (191.79)	1652.40 (284.02)	1118.57 (191.76)	117.65	< .001	.75	122.06	< .001	.76

Note. Each control sample consist of 40 scan and each clinical sample consist of 30 scans. ANOVA analysis shows the differences between the volumes as estimated using different methods, details of post hoc results can be seen in Figure 4.3. In the ANCOVA analysis, the associated results reported are the main effect of method of volume estimation after covarying out the effect of gender. Effects sizes are represented by the partial eta squared, η^2 .

Figure 4.3

Estimated Volumes of the Entorhinal Cortex Based on Different Methods of Segmentations



Note. This figure shows the estimated volumes of the entorhinal cortex with the associated means and standard deviations. The automated FreeSurfer (v6.0) volumes were derived using three different methods: Annotation, Label-plus and Label-minus. The error bars represent the 95% confidence interval around the mean. ANOVA with Tukey post hoc was conducted for each hemisphere within each sample: a. $p < .001$ significantly different from manual estimates; b. $p < .001$ significantly different from Annotation estimates; c. $p < .001$ significantly different from Label-plus estimates.

Table 4.3

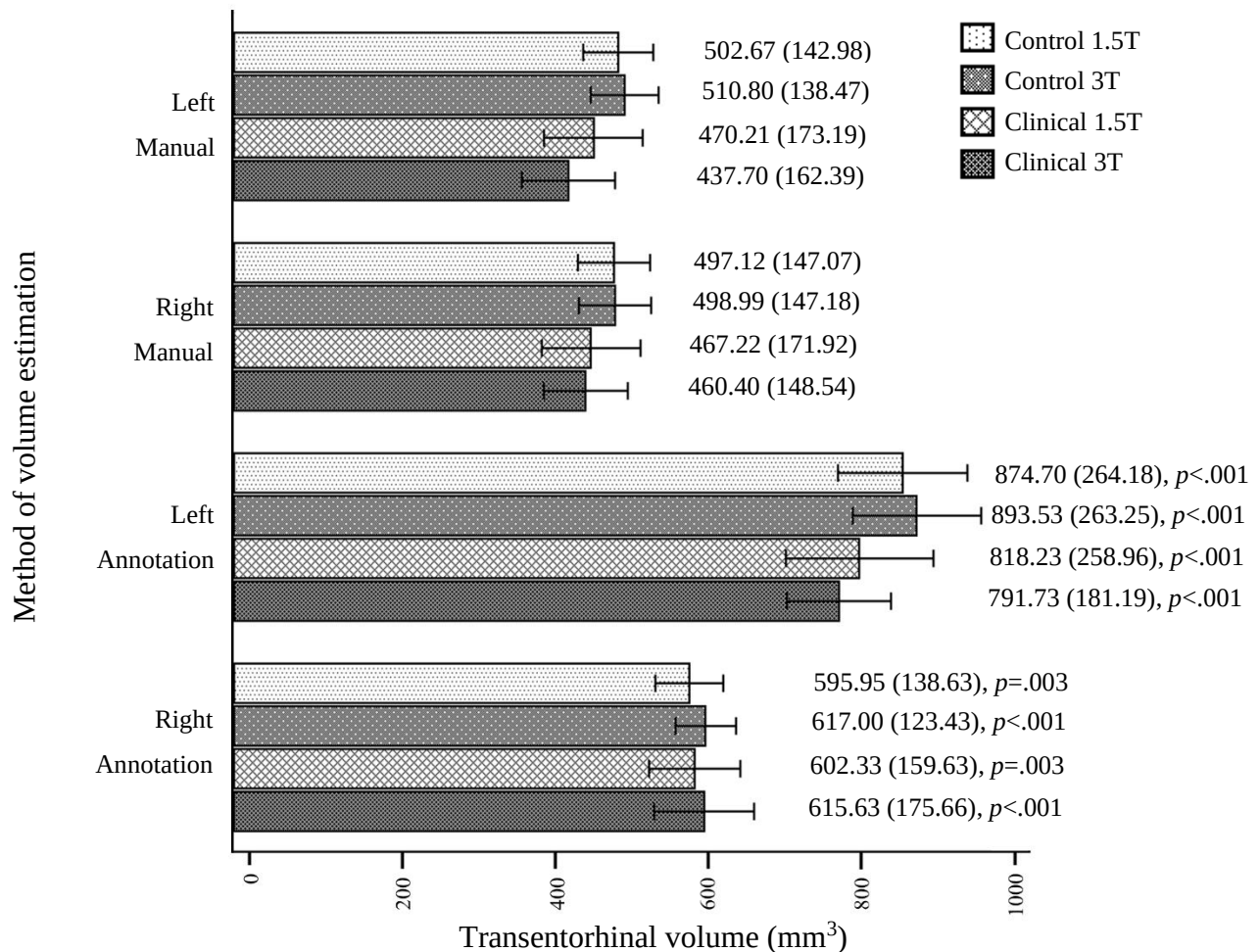
Means and Standard Deviations of the Transentorhinal Volumes and Comparisons of These Volumes Between the Different Methods of Volume Estimation

	Method of volume estimation		ANOVA			ANCOVA		
	Manual	Annotation	F-ratio	p-value	η^2	F-ratio	p-value	η^2
Control 1.5T								
Left	502.67 (142.98)	874.70 (264.18)	61.36	< .001	.44	60.70	< .001	.44
Right	497.12 (147.07)	595.95 (138.63)	9.56	.003	.11	9.68	.003	.11
Control 3T								
Left	510.80 (138.47)	893.53 (263.25)	66.22	< .001	.46	65.65	< .001	.46
Right	498.99 (147.18)	617.00 (123.43)	15.10	< .001	.16	15.18	< .001	.16
Clinical 1.5T								
Left	470.21 (173.19)	818.23 (258.96)	37.44	< .001	.39	37.49	< .001	.40
Right	467.22 (171.92)	602.33 (159.63)	9.95	.003	.15	9.78	.003	.15
Clinical 3T								
Left	437.70 (162.39)	791.73 (181.19)	63.52	< .001	.52	62.54	< .001	.52
Right	460.40 (148.54)	615.63 (175.66)	13.66	< .001	.19	13.46	.001	.19

Note. Each control sample consisted of 40 scan and each clinical sample consisted of 30 scans. ANOVA analysis shows the differences between the volumes as estimated using manual and automated methods. A graphical representation of these results can be seen in Figure 4.4. In the ANCOVA analysis, the associated results reported are the main effect of method of volume estimation after covarying out the effect of gender. Effects sizes are represented by the partial eta squared, η^2 .

Figure 4.4

The Estimated Volumes of the Transentorhinal Cortex Based on Manual and Automated Segmentation



Note. This figure shows the estimated volumes of the transentorhinal cortex with the associated means and standard deviations. The volumes generated using FreeSurfer (v6.0) were obtained from Annotation values. The error bars represent the 95% confidence interval around the mean. The comparisons between manual and Annotation volumes were conducted using ANOVA for each hemisphere within each group. The p -values for each pair of comparison are indicated on the graph.

the automated transentorhinal volume estimates, the right transentorhinal cortex was found to be significantly smaller than the left transentorhinal cortex in all four sample groups (Table 4.4).

Comparability Between Scanners (Control Sample)

Manual measurements. Paired-sample *t*-test showed no significant differences between the right entorhinal volumes obtained from different scanner field strengths. The left entorhinal volume on the 1.5T was 1.4% smaller than that found on 3T, but this difference did not reach significance ($p = .09$) and effect size was small to moderate, $d = 0.27$. There was no significant volume difference found between the left and right transentorhinal estimates on scans obtained from different field strengths and no effect sizes were found to be more than 0.20. Findings from the *t*-tests and percentage difference between the left and right volumes are summarised in Table 4.5. Absolute single-measure ICC values revealed good to excellent levels of agreement between the two scanners for both the entorhinal and transentorhinal cortices (Table 4.5).

Automated measurements. As shown in Table 4.5, the 1.5T left entorhinal volume estimates were significantly smaller than the 3T left entorhinal volume estimates by 3.5% - 3.9%, demonstrating a small to moderate effect size, on all three methods of automated volume calculations. There were no significant differences between the 1.5T and 3T volumes for the right entorhinal cortex (Table 4.5). The estimated volumes of the transentorhinal cortex was smaller by 2.2% on the left and 3.5% on the right on the 1.5T than the 3T scans. However, these differences did not reach significance for either the left ($p = .10$) or right hemisphere ($p = .07$), with small to moderate effect sizes, $d = 0.26$ and $d = 0.29$ respectively (Table 4.5).

Absolute single-measure ICC values are found in Table 4.5. Absolute single-measure ICC found in the *label-plus* right entorhinal volumes was moderate to good, while the ICC

Table 4.4*Comparisons of Left and Right Volume Estimates Using Paired-Sample t-test*

	1.5T Scanner					3T Scanner				
	Mean difference	SD	t-value	p-value	Cohen's <i>d</i>	Mean difference	SD	t-value	p-value	Cohen's <i>d</i>
Entorhinal										
Control										
Manual	-42.14	112.06	-2.38	.022*	0.38	-41.22	113.76	-2.29	.027*	0.36
Annotation	-228.50	123.21	-11.73	<.001*	1.85	-196.53	157.17	-7.91	<.001*	1.25
Label-plus	-566.75	185.46	-19.33	<.001*	3.06	-550.15	212.93	-16.34	<.001*	2.58
Label-minus	-228.43	123.37	-11.71	<.001*	1.85	-196.63	157.46	-7.90	<.001*	1.25
Clinical										
Manual	-48.48	108.20	-2.45	.020*	0.45	-32.23	125.51	-1.41	.170	0.26
Annotation	-221.50	213.80	-5.67	<.001*	1.04	-249.00	146.52	-9.31	<.001*	1.70
Label-plus	-570.10	318.57	-9.80	<.001*	1.79	-622.03	226.24	-15.06	<.001*	2.75
Label-minus	-221.77	213.70	-5.68	<.001*	1.04	-249.00	146.50	-9.31	<.001*	1.70
Transentorhinal										
Control										
Manual	5.55	91.24	0.39	.703	0.06	11.82	88.75	0.84	.405	0.13
Annotation	278.75	213.26	8.27	<.001*	1.31	276.53	227.14	7.70	<.001*	1.22
Clinical										
Manual	2.99	86.11	0.19	.851	0.03	-22.70	93.71	-1.33	.195	0.24
Annotation	215.90	235.59	5.02	<.001*	0.92	176.10	191.40	5.04	<.001*	0.92

Note. Significant differences between the estimates (left – right volumes) are marked with * and effect size was calculated using Cohen's *d*. Control samples were made up of 40 of the same participants in both the 1.5T and 3T scans, while clinical samples were made up of 30 different patients in each scanner group. Corresponding raw volume estimates can be found in Table 4.2 and Table 4.3.

Table 4.5

Comparisons of 1.5T and 3T Manual and Automated Volume Estimates in the Control Sample Using Paired-Sample t-test and Absolute Intraclass Correlation Coefficient (ICC)

	<i>t</i> -test					Single-measure absolute ICC	
	Mean difference (<i>SD</i>)	<i>t</i> -value	<i>p</i> -value	Cohen's <i>d</i>	Percentage difference	ICC value	95% CI
Manual							
Left entorhinal	-9.22 (34.01)	-1.72	.09	0.27	-1.4%	.96	.92, .98
Right entorhinal	-8.31 (47.63)	-1.10	.28	0.17	-1.2%	.93	.87, .96
Left transentorhinal	-8.13 (41.06)	-1.25	.22	0.20	-1.6%	.96	.92, .98
Right transentorhinal	-1.86 (36.74)	-0.32	.75	0.05	-0.4%	.97	.94, .98
Annotation							
Left entorhinal	-36.05 (89.65)	-2.54	.02*	0.40	-3.9%	.83	.68, .91
Right entorhinal	-4.08 (130.32)	-0.20	.84	0.03	-0.4%	.69	.48, .82
Left transentorhinal	-18.82 (71.59)	-1.66	.10	0.26	-2.2%	.96	.92, .98
Right transentorhinal	-21.05 (71.85)	-1.85	.07	0.29	-3.5%	.84	.72, .91
Label-plus							
Left entorhinal	-39.35 (103.15)	-2.41	.02*	0.38	-3.5%	.83	.68, .91
Right entorhinal	-22.75 (168.51)	-0.85	.40	0.14	-1.4%	.75	.60, .86
Label-minus							
Left entorhinal	-36.00 (89.76)	-2.54	.02*	0.40	-3.9%	.83	.68, .91
Right entorhinal	-4.20 (130.56)	-0.20	.84	0.03	-0.4%	.69	.48, .82

Note. Control Sample consist of 40 participants. Volumes reported are in mm³. Automated volume estimates were performed using FreeSrufer (v.6.0).

Significant differences between the volume estimates (1.5T - 3T) are marked with a * and effect size was calculated using Cohen's *d*. ICC values reported are based on two-way mixed, absolute agreement single measure. Percentage difference was calculated using (1.5T – 3T estimates) / 1.5T estimates * 100.

Corresponding raw volume estimates can be found in Table 4.2 and Table 4.3.

values for the *annotation* and *label-minus* right entorhinal volumes were poor to moderate. Across all methods of automated volumetric estimation, ICC values were found to be moderate to excellent for the left entorhinal cortex. The ICC values were excellent on the left transentorhinal cortex, and moderate to excellent on the right transentorhinal cortex, as well as the left entorhinal cortex.

Agreement Between Manual and Automated Methods

Control sample. In both 1.5T and 3T scans, both absolute and consistency single-measure ICC values showed a lack of agreement between the manual and automated measures on both the entorhinal and transentorhinal cortical volumes (Table 4.6). The highest single-measure ICC achieved was single-measure consistency ICC = .30 (95% CI: -.01, .56) in the left *label-plus* entorhinal estimates obtained from 1.5T scans (Table 4.6).

Bland-Altman plots show that the automated measures in both 1.5T and 3T scans tended to overestimate the entorhinal cortical volume, with larger overestimation on the right than the left entorhinal cortex (Figure 4.5 and Figure 4.6). The 95% limits of agreement were very wide regardless of scanner field strength, hemisphere or method of automated estimation, with the narrowest 95% limit of agreement being -113.8% to -55.90% in the 3T right *label-plus* entorhinal cortex to the widest 95% limits of agreement being -79.65% to 5.21% in the 1.5T *annotation* left entorhinal cortex (Figure 4.5 and Figure 4.6).

Similarly, the Bland-Altman plots show that automated measures tended to overestimate the transentorhinal cortex in both 1.5T and 3T scans, but with larger overestimation on the left than the right transentorhinal cortex (Figure 4.7). The 95% limits of agreement were very wide, and wider than that found in the entorhinal cortex, regardless of scanner field strength or hemisphere. The narrowest 95% limits of agreement were -98.0% to 2.9% in the 3T right transentorhinal cortex and the widest -137.9% to 36.7% in the 1.5T left transentorhinal cortex (Figure 4.7).

Table 4.6

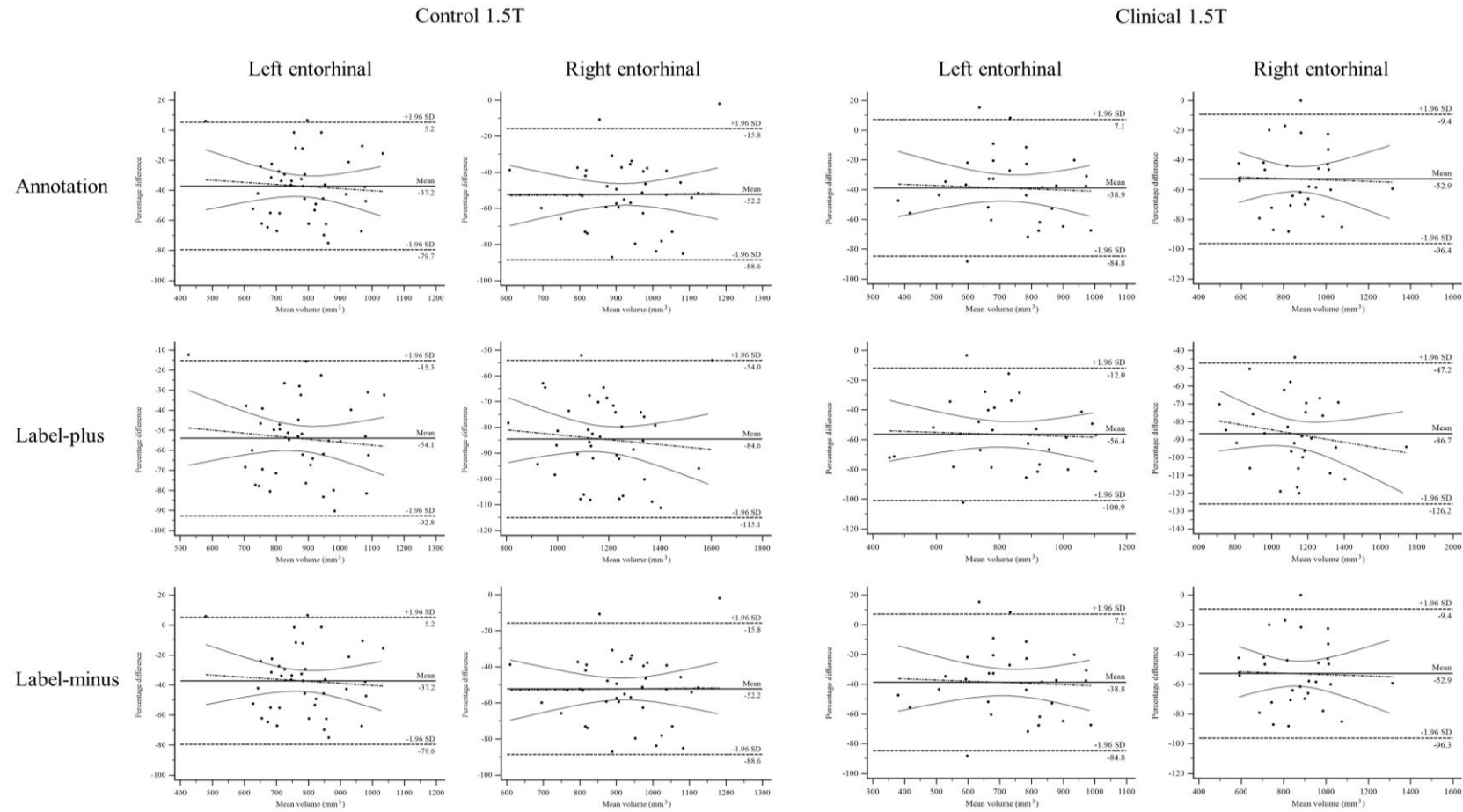
Intraclass Correlation Coefficient (ICC) Between Manual and Automated Estimates of the Entorhinal and Transentorhinal Cortices

	Single-measure ICC values							
	Absolute				Consistency			
	L	95% CI	R	95% CI	L	95% CI	R	95% CI
Entorhinal								
Control 1.5T								
Annotation	.09	-.07, .30	.04	-.04, .17	.26	-.05, .53	.24	-.08, .51
Label-plus	.06	-.05, .22	.02	-.02, .11	.30	-.01, .56	.29	-.02, .55
Label-minus	.09	-.07, .30	.04	-.04, .17	.26	-.05, .53	.24	-.08, .51
Clinical 1.5T								
Annotation	.22	-.10, .55	.10	-.07, .36	.48	.15, .71	.39	.04, .65
Label-plus	.13	-.07, .41	.04	-.03, .17	.45	.11, .69	.28	-.08, .58
Label-minus	.22	-.10, .55	.10	-.07, .36	.48	.15, .71	.39	.04, .65
Control 3T								
Annotation	.05	-.06, .21	.02	-.03, .11	.19	-.13, .47	.16	-.16, .44
Label-plus	.04	-.04, .17	.01	-.01, .07	.26	-.06, .52	.25	-.06, .52
Label-minus	.05	-.06, .21	.02	-.03, .11	.18	-.13, .47	.16	-.16, .44
Clinical 3T								
Annotation	.25	-.10, .59	.13	-.04, .43	.53	.21, .74	.63	.36, .81
Label-plus	.14	-.07, .45	.04	-.02, .17	.51	.19, .73	.42	.08, .68
Label-minus	.25	-.10, .59	.13	-.04, .43	.53	.21, .74	.63	.35, .81
Transentorhinal								
Control 1.5T								
Annotation	-.11	-.27, .14	-.14	-.39, .14	-.28	-.54, .03	-.18	-.46, .14
Clinical 1.5T								
Annotation	-.13	-.32, .17	-.14	-.40, .18	-.29	-.59, .07	-.18	-.50, .19
Control 3T								
Annotation	-.13	-.28, .16	-.20	-.43, .09	-.34	-.59, -.04	-.26	-.53, .05
Clinical 3T								
Annotation	-.15	-.26, .20	-.23	-.49, .11	-.45	-.69, -.11	-.33	-.61, .03

Note. Control sample consist of $n = 40$, while each clinical sample consist of $n = 30$. Automated volume estimates were performed using FreeSurfer (v.6.0). L = left hemisphere, R = right hemisphere

Figure 4.5

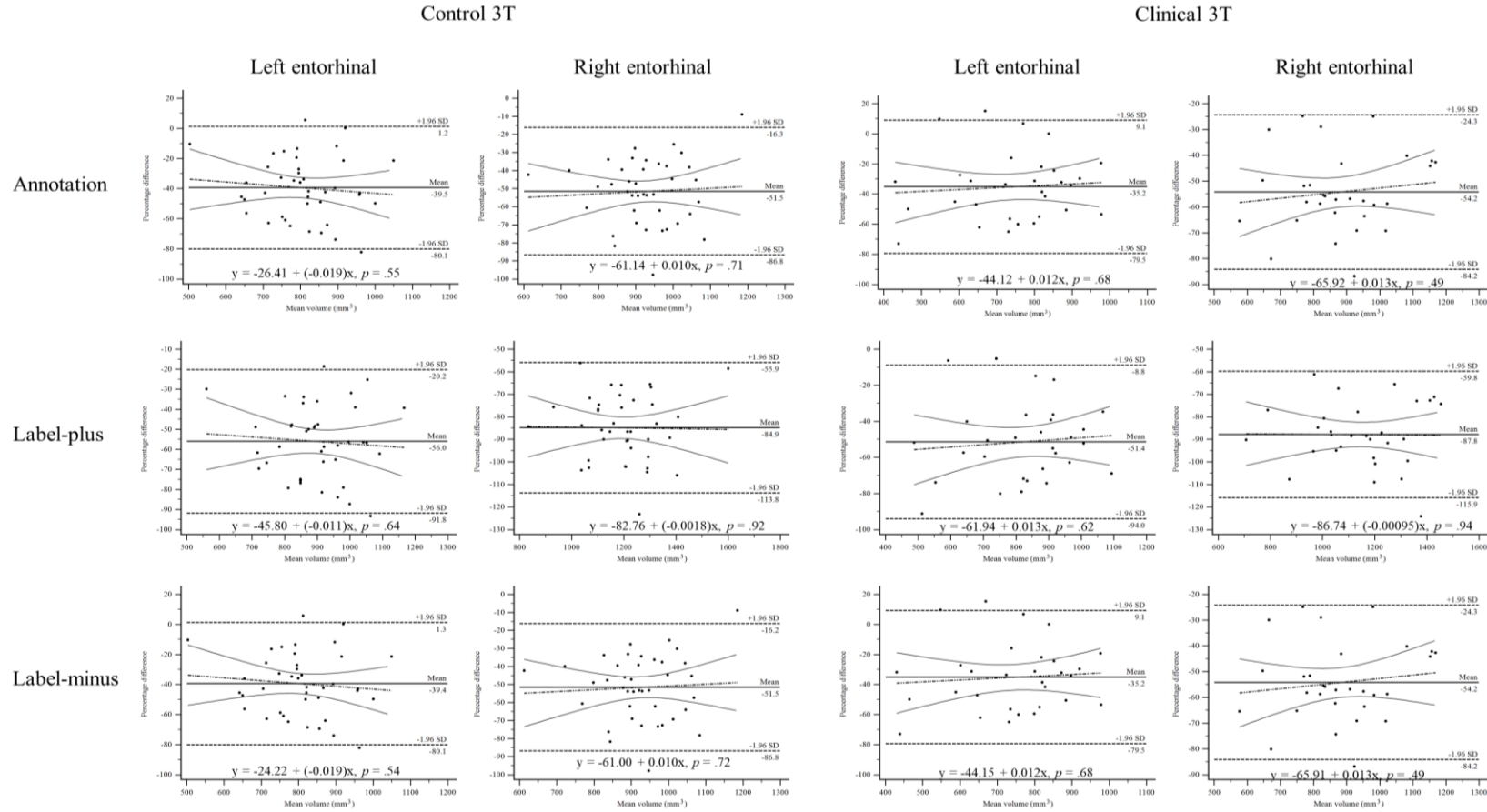
Bland-Altman Plots for Bias Estimation Between Manual and Automated Methods of Entorhinal Segmentation in the 1.5T Control and Clinical Samples



Note. This figure shows the Bland-Altman plots for bias in entorhinal volume estimations in the control 1.5T (n = 40) and clinical 1.5T (n = 30) samples. The percentage difference scores were calculated using (manual-automated)/mean of volumes * 100. The solid line represents the mean, the horizontal broken lines represent the 95% limits of agreement (± 1.96 * SD). Regression lines with the 95% confidence interval, represented by the broken line flanked with solid curved lines, were drawn to show potential proportional bias in estimation.

Figure 4.6

Bland-Altman Plots for Bias Estimation Between Manual and Automated Methods of Entorhinal Segmentation in the 3T Control and Clinical Samples



Note. This figure shows the Bland-Altman plots for bias between the entorhinal volume estimations between manual and automated methods of entorhinal segmentation in the control 3T (n = 40) and clinical 3T (n = 30) samples. The percentage difference scores were calculated using (manual-automated)/mean of volumes * 100. The solid line represents the mean, the horizontal broken lines represent the 95% limits of agreement ($\pm 1.96 * SD$). Regression lines with the 95% confidence interval, represented by the broken line flanked with solid curved lines, were drawn to show potential proportional bias in estimation.

Regression analysis of the relationship between the percentage difference score of the automated and manual entorhinal volume estimates, and the mean volume of the entorhinal cortex was not suggestive of a proportional bias. However, as seen in Figure 4.7, in the left transentorhinal cortex, regression line suggests a proportionate bias, such that the larger the mean of the transentorhinal cortex, the larger the difference between the two manual and automated measurements. The bias was observed on both 1.5T scans, $\beta = -0.15$, $p = .006$, and 3T scans, $\beta = -0.18$, $p = .001$.

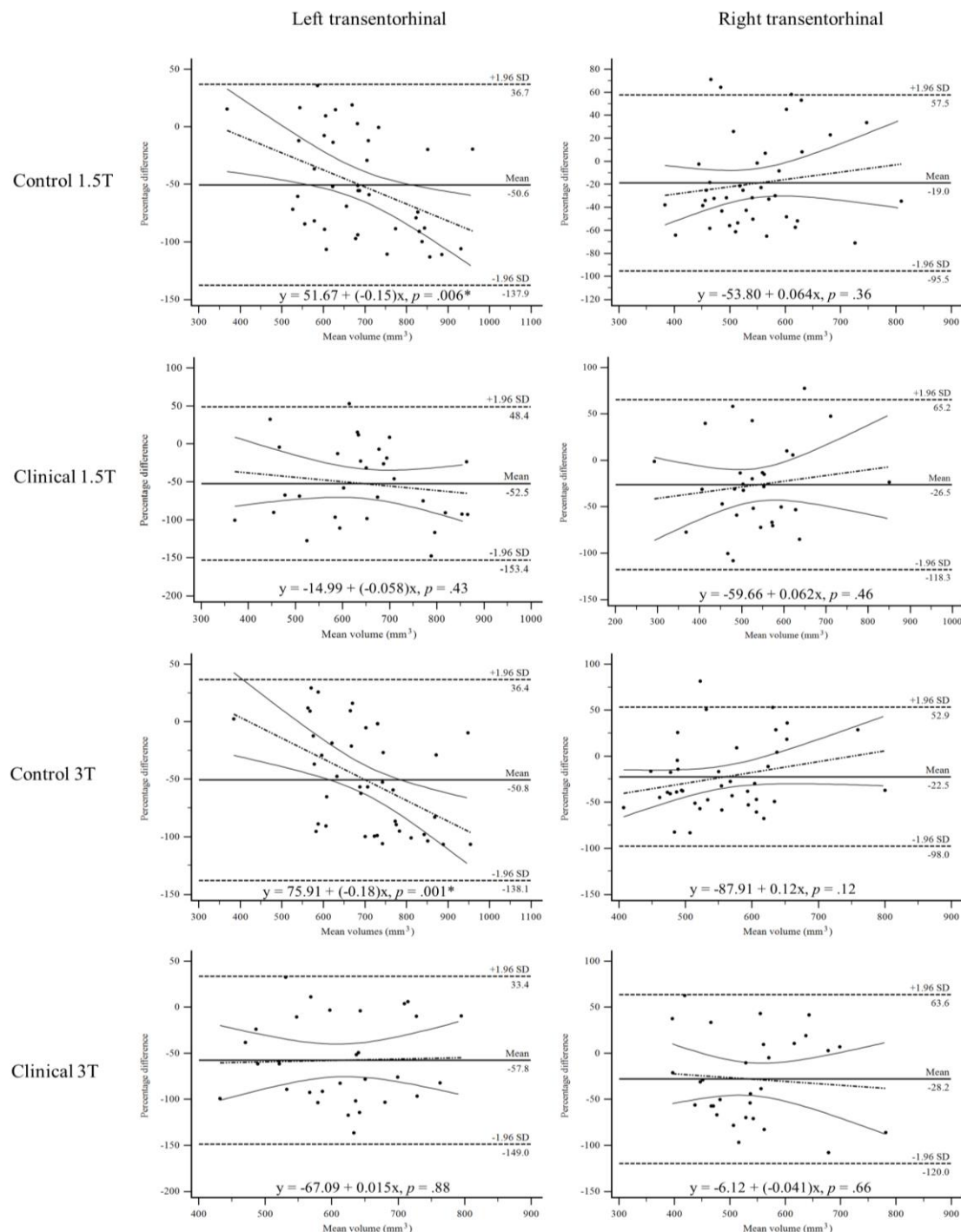
Clinical 1.5T. Single-measure ICC values showed a lack of agreement between all automated estimation of the entorhinal and transentorhinal cortical volumes with manual measurements. As detailed in Table 4.6, the highest single-measure ICC produced was found in the left entorhinal cortical volume estimated by *annotation* and *label-minus*, with identical consistency ICCs of .48 (95% CI: .15, .71).

Bland-Altman plots in Figure 4.5 show that automated measures tended to overestimate the entorhinal cortex, more so on the right than the left entorhinal cortex. Similar to the observation in the 1.5T control sample, the 95% limits of agreement were very wide regardless of hemisphere or method of automated estimation, with the narrowest being -126.2% to -47.2% in the right *label-plus* entorhinal cortex, to the widest being -84.8% to 7.2% in the *label-minus* left entorhinal cortex (Figure 4.5). The automated estimations of the transentorhinal cortex showed the same pattern of overestimation and wide 95% limits of agreement (Figure 4.7). Simple regression analysis was not suggestive of a proportional bias in either structure.

Clinical 3T. The ICCs were slightly higher than that found in the control sample and clinical 1.5T sample, with the highest single-measure consistency ICC = .63 found in both the *annotation* and *label-minus* estimates (Table 4.6). However, the 95% confidence intervals were very wide, ranging from very poor (ICC = .36) to good (ICC = .81).

Figure 4.7

Bland-Altman Plots for Bias Estimation Between Manual and Automated Methods of Transientorhinal Segmentation



Note. Bland-Altman plots for bias estimation between manual and automated methods of transientorhinal segmentation in the control ($n = 40$) and clinical samples ($n = 30$). All percentage difference scores were calculated using $(\text{manual} - \text{automated}) / \text{mean of volumes} \times 100$. The solid line represents the mean, the horizontal broken lines represent the 95% limits of agreement ($\pm 1.96 \times \text{SD}$). Regression line with its 95% confidence interval, represented by the broken line flanked with solid curved lines, were drawn to show potential proportional bias in estimation. Regression analysis that showed significant proportional biases are marked with an asterisk.

Similar to the findings in the control and clinical 1.5T samples, Bland-Altman plots show that automated measures tended to overestimate the entorhinal cortex, more so on the right than the left entorhinal cortex (Figure 4.6). Similar to the observation in the control 3T sample, the 95% limits of agreement were very wide regardless of hemisphere or method of automated estimation, with the narrowest being -115.9% to -59.8% in the right *label-plus* entorhinal cortex to the widest being -79.5% to 9.1% in the *label-minus* left entorhinal cortex (Figure 4.6). Automated measures in the transentorhinal cortex showed the same pattern of overestimation and wide 95% limits of agreement (Figure 4.7). Regression analysis was not suggestive of a proportional bias in either the entorhinal or transentorhinal cortex.

Discussion

Automation of brain volumetric measurements is an appealing potential addition to the clinical diagnostic toolbox. Compared to manual techniques, automated measurement is time-efficient and can be more easily adapted to different settings. With neuropathological findings supporting the selective involvement of the entorhinal and transentorhinal cortices early in the AD process, the ability to incorporate entorhinal and transentorhinal measures as supportive structural biomarkers could prove to play an important role for early diagnosis. To address this need for automated entorhinal and transentorhinal measurements, new *ex vivo* modules for both the entorhinal and transentorhinal cortex have been included in FreeSurfer. Unfortunately, very few studies have assessed the reliability and validity of these new pipelines. Hence, my study aimed to address this gap in the understanding of the performance of these new segmentation modules found in FreeSurfer (v.6.0) by pitting them against the “gold standard” of manual segmentation.

This study assessed the impact of scanner field strength on the manual and FreeSurfer segmentation results in a group of healthy older adults. Overall, all measurements on the 3T scanners tended to be larger than that found on the 1.5T scanner. Although the manual

tracings the entorhinal and transentorhinal volumes were similar on both the 1.5T and 3T scans, there was a small to moderate effect size noted in the left entorhinal cortex, in which the 1.5T scans were 1.4% smaller than that obtained on the 3T scans. The impact of scanner strength on volume estimation was more noticeable in automated segmentations, in which the left entorhinal cortex was significantly smaller by 3.5% to 3.9% on the 1.5T compared to the 3T scans across all automated measurements. Although not significantly different in volumes, effect sizes were small to moderate on the automated transentorhinal estimations, which translated to volumes obtained on the 1.5T scans being smaller by 2.2% on the left transentorhinal cortex and 3.5% on the right transentorhinal cortex when compared to the volumes obtained on the 3T scans.

Similar effects of scanner strength were not observed in the right automated volumes of the entorhinal cortex, but the effect was likely attenuated by reduced reliability, as evident in the rather discrepant measurements as assessed by ICC (Cohen, 1988; Portney & Watkins, 2013). Furthermore, with the exception of the left transentorhinal cortex, all absolute single-measure ICCs between 1.5T and 3T automated measures had wide confidence intervals and did not reach the recommended clinically reliable level of .90 (Portney & Watkins, 2013). These findings suggest a lack of interchangeability between the measures obtained from these two scanners. A note of interest is that although automated entorhinal volumes generated from *label-plus* was significantly different from that generated from *annotation* and *label-minus*, the pattern of differences in volumes between the 1.5T and 3T were similar across methods and hemispheres.

These differences in volumes based on magnetic field strength suggests that the ability to detect entorhinal and transentorhinal atrophy would easily be undermined by changes in scan quality. As mentioned earlier, the five-year longitudinal study by Stoub et al. (2010), showed different rates of entorhinal atrophy in participants who did and did not

experience further cognitive decline. The authors reported similar annual atrophy rates of $2.7 \pm 0.4\%$ and $3.3 \pm 0.9\%$ in stable controls ($n = 26$) and stable MCI ($n = 18$) respectively, but a significantly higher annual rate of entorhinal atrophy of $4.7 \pm 1.1\%$ in people with MCI who converted to AD ($n = 11$). A more recent three-year longitudinal study by Kulason et al. (2018) found that the average rate of reduction in entorhinal and transentorhinal volume per year, measured by manual segmentation, was 1.3% in healthy controls ($n = 21$) and 5.9% in people with MCI ($n = 35$). They also found that compared to controls, people with MCI had entorhinal and transentorhinal cortical volumes that were 22.7% smaller than controls at baseline.

As the findings from this study suggest, due to the sensitivity of volumes to scan quality, using rates of atrophy as an indicator of disease might not be a reliable measure over time given its relatively small range in percentage difference. This limitation is particularly true in clinical settings where scan quality would change over time due to scanner replacements, scanner software upgrades and general lack of control over scan parameters. The larger differences between volumes found in people with and without MCI would suggest that cross-sectional measures may be less sensitive to scanner differences. However, there is no available standardised norms for meaningful comparison at present. In addition, the suboptimal reliability of automated measurements between different scanners makes interpretation of automated volumetric outputs for the entorhinal and transentorhinal cortex less precise.

The main objective of this study was to evaluate the performance of the new entorhinal and transentorhinal *ex vivo* automated segmentation modules. To that end, this study assessed the validity of the automated volume estimations against the “gold standard” of manual estimations. The ICCs achieved for the transentorhinal volumes were all in the negative range, or statistically equivalent to an ICC of zero (James, Demaree, & Wolf, 1984;

LeBreton & Senter, 2008). These ICC values indicates a lack of agreement, as observed in the Bland-Altman plots for the right transentorhinal estimations, which showed a very large range in the limits of agreement between the two measures. An ICC in the negative range could also indicate a systematic disagreement between the two measures (LeBreton & Senter, 2008). Inspection of the Bland-Altman plots between the left manual and automated transentorhinal volumes in the control sample revealed a clear systematic disagreement, in which the larger the mean transentorhinal cortical volume, the more divergent the two measurements were. Although a clear systematic disagreement was not seen in the left transentorhinal Bland-Altman plots in the clinical samples, the 95% confidence intervals of the plots were wider than that seen in the control sample, which would reduce statistical ability to capture systematic differences.

The ICC values of automated entorhinal volume estimates appeared to be slightly more precise than the transentorhinal volumes. However, none of the values was close to the recommended criteria for what might be considered reasonable clinical measurements of .90 or even good reliability of .75 - .90 (Portney & Watkins, 2013). Perhaps a slightly unusual observation was that the reliability appeared slightly better for the clinical than the control sample, but this was likely a function of the larger variance in both manual and automated entorhinal volume estimates in the clinical samples. All the 95% confidence intervals were also very wide, with majority of the point estimates lying well below .75, indicating a low degree of similarity between manual and automated measures. There was also very little agreement between the manual and automated entorhinal volumes obtained, with the automated methods tending to overestimate the volumes. The degree of overestimation was variable between hemispheres and between methods of automated volume estimations. Slightly more encouraging than the findings for the transentorhinal volume estimations, was that the 95% limits of agreement for the entorhinal cortex were approximately two times

narrower; however, this reduction in the limits of agreement still resulted in a standard deviation of 14.3% at best, in the right entorhinal cortex of the 3T clinical *label-plus* sample, or at its worst, a standard deviation of 23.5% in the left entorhinal cortex of the 1.5T clinical *annotation* and *label-minus* samples. This large degree of variability makes it difficult to make precise interpretation of the entorhinal volumes when comparing against a baseline or a control group.

Although the new *ex vivo* protocols in FreeSurfer (v.6.0) appear promising, smaller structures like the entorhinal and transentorhinal cortices are inherently more challenging to measure as a small error in structural localisation would translate to a large error in volume estimation. The published pilot study in Chapter 3 that assessed the entorhinal *ex vivo* segmentation protocol in a mixed sample of 27 1.5T clinical scans reported very poor ICCs of .37 and .11, and very wide limits of agreement between manual and FreeSurfer volumes (Fung et al., 2019a). This current study echoes the unsatisfactory entorhinal findings in both control and clinical samples, and on both 1.5T and 3T scans. In addition to the small size of the entorhinal cortex, entorhinal segmentation is also complicated by its proximity to the tentorium cerebelli, in which cortical tissue and dura is not easily distinguishable in T1-weighted MRI (Augustinack et al., 2013; Fischl et al., 2009; Franko et al., 2014; Yushkevich, Pluta, et al., 2015).

The localisation of the transentorhinal cortex poses a slightly different problem. The transentorhinal cortex lies medial to the collateral sulcus, which is known to be very variable between and within individuals (Augustinack et al., 2013; Wolk et al., 2017; Xie et al., 2017; Yushkevich, Pluta, et al., 2015). The *ex vivo* protocol attempts to overcome the limitations of the anatomical variability through statistical averaging, but it was found that the Hausdorff distance – a mathematical measure of the distance between the automatically mapped structure and the manually defined structure – was much more variable in the transentorhinal

label than the entorhinal label (Augustinack et al., 2013; Augustinack et al., 2014; Fischl et al., 2009). Against manual segmentation as a reference standard, the lack of agreement between the transentorhinal volumes found in this study served to illustrate the comparatively more challenging nature of this structure.

The addition of a T2-weighted scan is recommended to increase contrast and therefore improve the ability of FreeSurfer to discern the entorhinal and the transentorhinal cortices (Augustinack et al., 2013; Fischl et al., 2009). However, T2 scans were not available for this study, which might have inevitably limited the performance of FreeSurfer segmentation. In a clinical setting, the feasibility of acquiring high-quality T2 scans is low compared to T1 scans as T2 scan parameters are more susceptible to movement artefacts due to the longer pulse sequence (Habas, 2018). Re-scans are also usually not practicable due to the significant time constraints on the MRI scanners. Nonetheless, understanding the degree to which an additional scanning would improve FreeSurfer segmentation accuracy would be an interesting line of investigation for future studies.

Other limitations of this study include the non-isotropic acquisition parameters and the re-sampling required to match the dimensions used for manual segmentations and automated segmentation. However, strict adherence to the manual protocol would have limited the potential biases to systematic differences between the manual and automated volumetric estimations. Herein also lies a main strength of this study, which is the comparability of both control and clinical scans that were all acquired using standard scan parameters. This data acquisition strategy mimics realistic clinical conditions and allows evaluation of FreeSurfer to overcome morphological irregularities that are likely to be present in clinical populations, while reducing variability due to differences in scan quality.

The main goal of this study was to assess the validity of the current FreeSurfer *ex vivo* segmentation modules and explore the potential of this FreeSurfer version as a diagnostic

tool. Using manual volumetric estimates as the reference standard on both the 1.5T and 3T scans, performance of the entorhinal *ex vivo* protocol was suboptimal and performance of the transentorhinal *ex vivo* protocol was very poor on the consistency and agreement metrics. These findings highlight the inherently challenging nature of automatically estimating volumes of the entorhinal and transentorhinal cortices even under ideal, clinical-like conditions. There is therefore a need for additional refinement to increase the segmentation accuracy of these two structures before these automated algorithms can be confidently translated into use for routine, diagnostic purposes in a clinical setting.

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to 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), 270-279. doi:10.1016/j.jalz.2011.03.008
- Alzheimer's Association. (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 14(3), 367-425. doi:10.1016/j.jalz.2018.02.001
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Arendt, T., Bruckner, M. K., Morawski, M., Jager, C., & Gertz, H. J. (2015). Early neurone loss in Alzheimer's disease: cortical or subcortical? *Acta Neuropathologica Communications*, 3, 10. doi:10.1186/s40478-015-0187-1
- Augustinack, J. C., Huber, K. E., Stevens, A. A., Roy, M., Frosch, M. P., van der Kouwe, A. J., . . . Alzheimer's Disease Neuroimaging, I. (2013). Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. *Neuroimage*, 64, 32-42. doi:10.1016/j.neuroimage.2012.08.071
- Augustinack, J. C., Magnain, C., Reuter, M., van der Kouwe, A. J., Boas, D., & Fischl, B. (2014). MRI parcellation of ex vivo medial temporal lobe. *Neuroimage*, 93 Pt 2, 252-259. doi:10.1016/j.neuroimage.2013.05.053
- Bejanin, A., Schonhaut, D. R., La Joie, R., Kramer, J. H., Baker, S. L., Sosa, N., . . . Rabinovici, G. D. (2017). Tau pathology and neurodegeneration contribute to cognitive impairment in Alzheimer's disease. *Brain*, 140(12), 3286-3300. doi:10.1093/brain/awx243

- Berron, D., Vieweg, P., Hochkeppler, A., Pluta, J. B., Ding, S. L., Maass, A., . . . Wisse, L. E. M. (2017). A protocol for manual segmentation of medial temporal lobe subregions in 7 Tesla MRI. *NeuroImage: Clinical*, 15, 466-482. doi:10.1016/j.nicl.2017.05.022
- Bigler, E. D. (2015). Structural image analysis of the brain in neuropsychology using magnetic resonance imaging (MRI) techniques. *Neuropsychology Review*, 25(3), 224-249. doi:10.1007/s11065-015-9290-0
- Bland, J. M., & Altman, D. G. (1995). Comparing methods of measurement - why plotting difference against standard method is misleading. *Lancet*, 346(8982), 1085-1087. doi:10.1016/s0140-6736(95)91748-9
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135-160. doi:10.1191/096228099673819272
- Braak, H., Alafuzoff, I., Arzberger, T., Kretschmar, H. A., & Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathologica*, 112(4), 389-404. doi:10.1007/s00401-006-0127-z
- Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4), 239-259. doi:10.1007/bf00308809
- Braak, H., & Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging*, 16(3), 271-278. doi:10.1016/0197-4580(95)00021-6
- Braak, H., & Del Tredici, K. (2012). Alzheimer's disease: pathogenesis and prevention. *Alzheimer's & Dementia*, 8(3), 227-233. doi:10.1016/j.jalz.2012.01.011
- Braak, H., & Del Tredici, K. (2015a). *Neuroanatomy and pathology of sporadic Alzheimer's Disease*: Springer.

- Braak, H., & Del Tredici, K. (2015b). The preclinical phase of the pathological process underlying sporadic Alzheimer's disease. *Brain*, 138(Pt 10), 2814-2833.
doi:10.1093/brain/awv236
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 7. doi:10.1037/1040-3590.6.4.284
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed. ed.): L. Erlbaum Associates.
- Desikan, R. S., Cabral, H. J., Settecase, F., Hess, C. P., Dillon, W. P., Glastonbury, C. M., . . . Fischl, B. (2010). Automated MRI measures predict progression to Alzheimer's disease. *Neurobiology of Aging*, 31(8), 1364-1374.
doi:10.1016/j.neurobiolaging.2010.04.023
- Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Killiany, R. J. (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. doi:10.1016/j.neuroimage.2006.01.021
- Despotovic, I., Goossens, B., & Philips, W. (2015). MRI segmentation of the human brain: challenges, methods, and applications. *Computational and Mathematical Methods in Medicine*, 2015, 450341. doi:10.1155/2015/450341
- Dickerson, B. C., Feczko, E., Augustinack, J. C., Pacheco, J., Morris, J. C., Fischl, B., & Buckner, R. L. (2009). Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of Aging*, 30(3), 432-440. doi:10.1016/j.neurobiolaging.2007.07.022

- Ding, S. L., & Van Hoesen, G. W. (2010). Borders, extent, and topography of human perirhinal cortex as revealed using multiple modern neuroanatomical and pathological markers. *Human Brain Mapping*, 31(9), 1359-1379. doi:10.1002/hbm.20940
- Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., . . . Washington Dc, U. S. A. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 12(3), 292-323. doi:10.1016/j.jalz.2016.02.002
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341-355. doi:10.1016/s0896-6273(02)00569-x
- Fischl, B., Stevens, A. A., Rajendran, N., Yeo, B. T., Greve, D. N., Van Leemput, K., . . . Augustinack, J. C. (2009). Predicting the location of entorhinal cortex from MRI. *Neuroimage*, 47(1), 8-17. doi:10.1016/j.neuroimage.2009.04.033
- Franco, E., Insausti, A. M., Artacho-Perula, E., Insausti, R., & Chavoix, C. (2014). Identification of the human medial temporal lobe regions on magnetic resonance images. *Human Brain Mapping*, 35(1), 248-256. doi:10.1002/hbm.22170
- Fung, Y. L., Ng, K. E. T., Vogrin, S. J., Meade, C., Ngo, M., Collins, S. J., & Bowden, S. C. (2019a). Comparative utility of manual versus automated segmentation of hippocampus and entorhinal cortex volumes in a memory clinic sample. *Journal of Alzheimer's Disease*, 68(1), 159-171. doi:10.3233/JAD-181172
- Fung, Y. L., Ng, K. E. T., Vogrin, S. J., Meade, C., Ngo, M., Collins, S. J., & Bowden, S. C. (2019b, March 14). Labelling of Statistical Output of Cytoarchitecture-Defined Variables in FreeSurfer Version 6. Retrieved from <https://www.j-alz.com/content/labelling-statistical-output-cytoarchitecture>

- Giannakopoulos, P., Herrmann, F., Md, M. P. H., Bussiere, T., Bouras, C., Kovari, E., . . . Hof, P. (2003). Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology*, 60(9), 1495-1500.
doi:10.1212/01.WNL.0000063311.58879.01
- Goncharova, II, Dickerson, B. C., Stoub, T. R., & deToledo-Morrell, L. (2001). MRI of human entorhinal cortex: a reliable protocol for volumetric measurement. *Neurobiology of Aging*, 22(5), 737-745. doi:10.1016/s0197-4580(01)00270-6
- Habas, C. (2018). *The neuroimaging of brain diseases : Structural and functional advances*: Springer.
- Honeycutt, N. A., Smith, P. D., Aylward, E., Li, Q., Chan, M., Barta, P. E., & Pearlson, G. D. (1998). Mesial temporal lobe measurements on magnetic resonance imaging scans. *Psychiatry Research*, 83(2), 85-94. doi:10.1016/S0925-4927(98)00035-3
- Insausti, R., Juottonen, K., Soininen, H., Insausti, A. M., Partanen, K., Vainio, P., . . . Pitkanen, A. (1998). MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *Ajnr: American Journal of Neuroradiology*, 19(4), 659-671.
- Jack, C. R., Jr. (2012). Alzheimer disease: new concepts on its neurobiology and the clinical role imaging will play. *Radiology*, 263(2), 344-361. doi:10.1148/radiol.12110433
- Jack, C. R., Jr., Barkhof, F., Bernstein, M. A., Cantillon, M., Cole, P. E., Decarli, C., . . . Foster, N. L. (2011). Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. *Alzheimer's & Dementia*, 7(4), 474-485 e474. doi:10.1016/j.jalz.2011.04.007
- Jack, C. R., Jr., Barnes, J., Bernstein, M. A., Borowski, B. J., Brewer, J., Clegg, S., . . . Weiner, M. (2015). Magnetic resonance imaging in Alzheimer's Disease Neuroimaging Initiative 2. *Alzheimer's & Dementia*, 11(7), 740-756.
doi:10.1016/j.jalz.2015.05.002

- James, R. L., Demaree, G. R., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69(1), 85-98. doi:10.1037/0021-9010.69.1.85
- Kivisaari, S. L., Probst, A., & Taylor, K. I. (2013). The perirhinal, entorhinal, and parahippocampal cortices and hippocampus: An overview of functional anatomy and protocol for their segmentation in MR Images. In *fMRI : Basics and Clinical Applications* (pp. 239).
- Kulason, S., Tward, D. J., Brown, T., Sicat, C. S., Liu, C. F., Ratnanather, J. T., . . . Alzheimer's Disease Neuroimaging, I. (2018). Cortical thickness atrophy in the transentorhinal cortex in mild cognitive impairment. *NeuroImage: Clinical*, 21, 101617. doi:10.1016/j.nicl.2018.101617
- LeBreton, J., & Senter, L. J. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815-852. doi:10.1177/1094428106296642
- Lehmann, M., Douiri, A., Kim, L. G., Modat, M., Chan, D., Ourselin, S., . . . Fox, N. C. (2010). Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. *Neuroimage*, 49(3), 2264-2274. doi:10.1016/j.neuroimage.2009.10.056
- Maass, A., Lockhart, S. N., Harrison, T. M., Bell, R. K., Mellinger, T., Swinnerton, K., . . . Jagust, W. J. (2018). Entorhinal tau pathology, episodic memory decline, and neurodegeneration in aging. *Journal of Neuroscience*, 38(3), 530-543. doi:10.1523/JNEUROSCI.2028-17.2017
- Mai, J. K., & Paxinos, G. (2011). *The Human Nervous System*. Saint Louis, UNITED STATES: Elsevier Science & Technology.

- Markesbery, W. R., Schmitt, F. A., Kryscio, R. J., Davis, D. G., Smith, C. D., & Wekstein, D. R. (2006). Neuropathologic substrate of mild cognitive impairment. *Archives of Neurology*, 63(1), 38-46. doi:10.1001/archneur.63.1.38
- McCarthy, C., 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. doi:10.3389/fnins.2015.00379/full
- McEvoy, L. K., & Brewer, J. B. (2010). Quantitative structural MRI for early detection of Alzheimer's disease. *Expert Review of Neurotherapeutics*, 10(11), 1675-1688. doi:10.1586/ern.10.162
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46. doi:10.1037/1082-989X.1.1.30
- McKhann, G. M., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7), 939-944. doi:10.1212/wnl.34.7.939
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to 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), 263-269. doi:10.1016/j.jalz.2011.03.005
- Meade, C. E., Bowden, S. C., Whelan, G., & Cook, M. J. (2008). Rhinal cortex asymmetries in patients with mesial temporal sclerosis. *Seizure-European Journal of Epilepsy*, 17(3), 234-246. doi:10.1016/j.seizure.2007.07.010

- Mueller, S. G., Schuff, N., Yaffe, K., Madison, C., Miller, B., & Weiner, M. W. (2010). Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease. *Human Brain Mapping*, 31(9), 1339-1347. doi:10.1002/hbm.20934
- Novak, G., & Einstein, S. G. (2013). Structural magnetic resonance imaging as a biomarker for the diagnosis, progression, and treatment of Alzheimer disease. In R. A. McArthur (Ed.), *Translational Neuroimaging* (pp. 87-129): Academic Press.
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment - Clinical characterization and outcome. *Archives of Neurology*, 56(3), 303-308. doi:10.1001/archneur.56.3.303
- Portney, L. G., & Watkins, M. P. (2013). *Foundations of clinical research: Applications to practice* (Vol. Third edition). Harlow: Pearson.
- Pruessner, J. C., Kohler, S., Crane, J., Pruessner, M., Lord, C., Byrne, A., . . . Evans, A. C. (2002). Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. *Cerebral Cortex*, 12(12), 1342-1353. doi:10.1093/cercor/12.12.1342
- Ramos Bernardes da Silva Filho, S., Oliveira Barbosa, J. H., Rondinoni, C., Dos Santos, A. C., Garrido Salmon, C. E., da Costa Lima, N. K., . . . Moriguti, J. C. (2017). Neuro-degeneration profile of Alzheimer's patients: A brain morphometry study. *NeuroImage: Clinical*, 15, 15-24. doi:10.1016/j.nicl.2017.04.001
- Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, 13(10), 713-726. doi:10.1038/nrn3338
- Sanchez-Benavides, G., Gomez-Anson, B., Sainz, A., Vives, Y., Delfino, M., & Pena-Casanova, J. (2010). Manual validation of FreeSurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer Disease

- subjects. *Psychiatry Research*, 181(3), 219-225.
doi:10.1016/j.psychresns.2009.10.011
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations - uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428. doi:10.1037/0033-2909.86.2.420
- Sone, D., Imabayashi, E., Maikusa, N., Okamura, N., Furumoto, S., Kudo, Y., . . . Matsuda, H. (2017). Regional tau deposition and subregion atrophy of medial temporal structures in early Alzheimer's disease: A combined positron emission tomography/magnetic resonance imaging study. *Alzheimer's & Dementia*, 9, 35-40. doi:10.1016/j.dadm.2017.07.001
- Sperling, R., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., . . . Phelps, C. H. (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. doi:10.1016/j.jalz.2011.03.003
- Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, 253(5026), 1380. doi:10.1126/science.1896849
- Stoub, T. R., Rogalski, E. J., Leurgans, S., Bennett, D. A., & deToledo-Morrell, L. (2010). Rate of entorhinal and hippocampal atrophy in incipient and mild AD: Relation to memory function. *Neurobiology of Aging*, 31(7), 1089-1098. doi:10.1016/j.neurobiolaging.2008.08.003
- Tan, C. C., Yu, J. T., & Tan, L. (2014). Biomarkers for preclinical Alzheimer's disease. *Journal of Alzheimer's Disease*, 42(4), 1051-1069. doi:10.3233/JAD-140843
- Taylor, K. I., & Probst, A. (2008). Anatomic localization of the transentorhinal region of the perirhinal cortex. *Neurobiology of Aging*, 29(10), 1591-1596. doi:10.1016/j.neurobiolaging.2007.03.024

- Teipel, S. J., Ewers, M., Wolf, S., Jessen, F., Kolsch, H., Arlt, S., . . . Hampel, H. (2010). Multicentre variability of MRI-based medial temporal lobe volumetry in Alzheimer's disease. *Psychiatry Research*, 182(3), 244-250.
doi:10.1016/j.psychresns.2010.03.003
- van Strien, N. M., Cappaert, N. L., & Witter, M. P. (2009). The anatomy of memory: An interactive overview of the parahippocampal–hippocampal network. *Nature Reviews Neuroscience*, 10(4), 272. doi:10.1038/nrn2614
- Wilke, M., de Haan, B., Juenger, H., & Karnath, H. O. (2011). Manual, semi-automated, and automated delineation of chronic brain lesions: a comparison of methods. *Neuroimage*, 56(4), 2038-2046. doi:10.1016/j.neuroimage.2011.04.014
- Wilson, R. S., Segawa, E., Boyle, P. A., Anagnos, S. E., Hize, L. P., & Bennett, D. A. (2012). The natural history of cognitive decline in Alzheimer's disease. *Psychology & Aging*, 27(4), 1008-1017. doi:10.1037/a0029857
- Wolk, D. A., Das, S. R., Mueller, S. G., Weiner, M. W., Yushkevich, P. A., & Alzheimer's Disease Neuroimaging, I. (2017). Medial temporal lobe subregional morphometry using high resolution MRI in Alzheimer's disease. *Neurobiology of Aging*, 49, 204-213. doi:10.1016/j.neurobiolaging.2016.09.011
- Xie, L., Das, S. R., Wisse, L. E. M., Ittyerah, R., Yushkevich, P. A., Wolk, D. A., & Alzheimer's Dis Neuroimaging, I. (2018). Early tau burden correlates with higher rate of atrophy in transentorhinal cortex. *Journal of Alzheimer's Disease*, 62(1), 85-92.
doi:10.3233/jad-170945
- Xie, L., Pluta, J. B., Das, S. R., Wisse, L. E. M., Wang, H., Mancuso, L., . . . Yushkevich, P. A. (2017). Multi-template analysis of human perirhinal cortex in brain MRI: Explicitly accounting for anatomical variability. *Neuroimage*, 144(Pt A), 183-202.
doi:10.1016/j.neuroimage.2016.09.070

- Xu, Y., Jack, C. R., Jr., O'Brien, P. C., Kokmen, E., Smith, G. E., Ivnik, R. J., . . . Petersen, R. C. (2000). Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. *Neurology*, 54(9), 1760-1767. doi:10.1212/wnl.54.9.1760
- Young, B. J., Otto, T., Fox, G. D., & Eichenbaum, H. (1997). Memory representation within the parahippocampal region. *The Journal of Neuroscience*, 17(13), 5183-5195. doi:10.1523/jneurosci.17-13-05183.1997
- Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S. L., Gertje, E. C., . . . Wolk, D. A. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, 36(1), 258-287. doi:10.1002/hbm.22627

Chapter 5

Manual Validation of FreeSurfer Hippocampal Subfields Protocol in a Clinical Population

Similar to the study design in Chapter 4, this chapter serves as an extension to the Chapter 3. Again, like Chapter 4, this chapter was designed to overcome the previous sample limitations in the pilot study (Chapter 3). This study specifically validated the FreeSurfer hippocampal volumes against manual segmentation, using data obtained from the entire sample pool that consist of 40 healthy controls and 90 clinical patients.

Abstract

One of the best structural biomarkers for identification of Alzheimer's disease is hippocampal atrophy. Manual delineation remains the gold standard for structural volumetry, but this technique is time consuming and requires extensive training, which limits its use in routine clinical workup. The use of automated neuroimaging software would facilitate the inclusion of hippocampal volumes as supportive evidence in clinical diagnosis, but its feasibility in a clinical population requires clarification and validation. In this study, automated hippocampal segmentation using the FreeSurfer (v.6.0) hippocampal subfield protocol was validated against manual segmentation on 170 scans obtained from 130 older adults. This included 40 control 1.5T and 3T scans, 60 routine clinical 1.5T scans, and 30 routine clinical 3T scans. The results showed that compared to manual measurements, automated measurements tended to systematically underestimate hippocampal volumes to similar degrees across different scans. Consistency single-measure ICCs however, demonstrated good reliability between manual and automated measurements. These findings suggest that the FreeSurfer (v.6.0) hippocampal subfield protocol is robust against varying scan parameters and hippocampal morphologies, hence demonstrating good translational potential for use in a clinical context.

Keywords: hippocampus, FreeSurfer, automated, manual, volume, segmentation, validation, reliability

Manual Validation of FreeSurfer Hippocampal Subfields Protocol in a Clinical Population

The most common form of dementia is Alzheimer's disease (AD), which is an unrelenting neurodegenerative disorder, which in its early symptomatic stages typically begins as insidious memory decline but can involve other cognitive difficulties, and culminates into losses of functional abilities as disease severity increases (Calderon-Garciduenas & Duyckaerts, 2017; Lane, Hardy, & Schott, 2018). As the worldwide population grows progressively older, AD is recognised as a public health priority, especially given the exponential rise in the risk of developing AD with older age (Fiest et al., 2016; Lane et al., 2018).

Alzheimer's Disease Pathology

Although the understanding of the neuropathology and pathogenesis of AD has dramatically increased over recent decades, attempts at the developing an effective disease-modifying treatment has been unsatisfactory thus far (Reynolds, 2019). One of the key neuropathogenic process involved in AD is the intraneuronal misfolding of abnormal tau protein into neurofibrillary tangles (NFTs), and autopsy studies have shown that the majority of cognitively normal adults have traces of the abnormal AD-related tau protein (Braak & Del Tredici, 2011; Braak, Thal, Ghebremedhin, & Del Tredici, 2011).

Even though the process that initiates the propagation of the originally quiescent, abnormal tau protein remains debatable, it is agreed that the downstream effects of the AD-related neuropathological changes become challenging to mitigate once they are well-established in the later stages of AD and when numerous other deleterious pathogenic cascades, such as oxidative stress, neuroinflammation, metal dyshomeostasis, have been triggered (Braak & Del Tredici, 2015b; Serrano-Pozo, Frosch, Masliah, & Hyman, 2011). Therefore, it is imperative to reliably identify people in the earliest stages of AD, even pre-

symptomatically, given the long preclinical evolution of AD neuropathologically. Early identification not only provides the opportunity for patients and their family members to make plans for future care and services, but it also provides the greatest likelihood for patients to derive a meaningful clinical benefit from disease-modifying therapeutic drugs when such interventions do become available (Doody, 2017; Miyagawa & Iwatsubo, 2017).

Hippocampal Atrophy as a Biomarker for Alzheimer's Disease

There is broad consensus that a multi-modal approach to the diagnostic process is the most effective way to bring forward the AD detection window (Alberdi, Aztiria, & Basarab, 2016; Bauer et al., 2018; Belleville, Fouquet, Duchesne, Collins, & Hudon, 2014). Germane to that goal, is the development of effective tools for the detection of relevant pathophysiological biomarkers (Albert et al., 2011; Bayram et al., 2018; Dubois et al., 2010; Frisoni et al., 2017; McKhann et al., 2011; Sperling & Johnson, 2013). Of the available tools, neuroimaging techniques are arguably one of the most readily accessible and effective methods for identifying biomarkers in early stage AD. Several common neuroimaging biomarkers has been identified in AD including amyloid loading through use of positron emission tomography with amyloid ligands such as Pittsburgh compound B, or the regional rates of hypometabolism through fluorodeoxyglucose positron emission tomography, and the most established of those neuroimaging biomarkers is hippocampal volumetry (Tan et al., 2014).

Neurofibrillary tangles (NFTs) are associated with synapse loss and neuronal death, and the increasing loss of functional synapses and neurons will eventually lead to the curtailment of cognitive functions associated with the structure inflicted by these tangles (Braak & Del Tredici, 2015a). The hippocampus, being a relatively early preferential target of these NFTs, is an anatomical structure that is closely related to the pathological expression of the disease and its severity (Braak & Braak, 1995; Braak & Del Tredici, 2015b; Calderon-

Garciduenas & Duyckaerts, 2017). Reductions in hippocampal volumes have been shown to be a good proxy to determining the severity of neuronal and synaptic loss (Atiya, Hyman, Albert, & Killiany, 2003; Kantarci & Jack, 2004; Mueller et al., 2010). The reported age-related hippocampal atrophy rates in healthy adults above the age of 70, ranged from .3% - 1.3% (Cover et al., 2016; Fraser, Shaw, Anstey, & Cherbuin, 2018; Schmidt et al., 2018). On the other hand, the reported hippocampal atrophy rates in people above 70 years old who had been diagnosed with MCI and AD were comparatively larger, with estimates ranging from 2.0% - 3.5% in MCI and 3.6% - 4.2% in AD (Cover et al., 2016; Mulder et al., 2014). Hippocampal atrophy has been demonstrated to be evident even in cognitively asymptomatic individuals carrying the disease and the atrophy continues even into the severe stages of AD (Dubois, Hampel, et al., 2016; Jack, 2011; Mak et al., 2017; Ramos Bernardes da Silva Filho et al., 2017).

From a clinical standpoint, particular attention has been placed on people in the prodromal stage of AD, typically categorised as people with mild cognitive impairment (MCI), a condition in which there is evidence of a noteworthy objective cognitive decline in at least one domain (typically explicit memory) but maintenance of functional independence (Albert et al., 2011; Petersen et al., 1999). From the voluminous research on MCI, it has been repeatedly shown that reductions in hippocampal volumes, coupled with cognitive deficits typical for AD, are one of the best prognostic biomarkers of future conversion from MCI into AD (Clerx et al., 2013; Pini et al., 2016; Schroder & Pantel, 2016; Teipel, Kurth, Krause, & Grothe, 2015; Trzepacz et al., 2014; Weiner et al., 2015). Although it is not advocated for clinicians to use the presence of hippocampal atrophy for routine diagnostic purposes, its use would likely enhance the specificity of a diagnosis of AD-related cognitive decline (Albert et al., 2011; Jack, Albert, et al., 2011; McKhann et al., 2011).

Volumetric Quantification of the Hippocampus

One of the greatest advantages of using hippocampal atrophy as a supportive biomarker is that it could be determined non-invasively through use of structural magnetic resonance imaging (MRI). Typically, hippocampal atrophy is determined qualitatively through routine visual assessment of MRI brain scans. Although visual inspection for regional atrophy increases confidence in the clinical diagnosis of AD, it is not an assessment method that can be reliably reproduced (Jack, 2012).

To provide quantification of hippocampal volumes, manual delineation of structural MRI scans by an expert operator is viewed as the gold standard (Lehmann et al., 2010; Sanchez-Benavides et al., 2010; Teipel et al., 2010). Experienced raters can adapt easily to variations in pathological and physiological shape of the structure, which is likely to occur regularly in a clinical cohort (Jack, 2012). However, this manual method is associated with several constraints. Manual segmentation requires trained expertise in the region of interest, which is a skill not readily available (Jack, 2012). It is also a very laborious task and therefore not economical, time-efficient and is neither realistic for routine assessment in a clinical setting nor feasible for dealing with large datasets in research environments (Jack, Barkhof, et al., 2011; McEvoy & Brewer, 2010). Furthermore, measurements by operators trained on the same set of manual protocols could differ significantly between each other and differ within themselves across time (Despotovic et al., 2015). Standardised normative data, based on manual measurements, are also lacking, limiting meaningful interpretation of such stated hippocampal volumes (Jack, 2012).

With the advent of automated methods developed to parcellate grey matter without manual interventions, it provides an opportunity to improve reproducibility of findings across different settings and standardisation of the extraction of quantitative information while also being much less labour intensive (Jack, 2012). Some have also suggested that compared to

manual segmentations, automated measures are more reproducible and therefore more sensitive to gradual hippocampal atrophy over time (Cover et al., 2018; Cover et al., 2016).

FreeSurfer is one of the most commonly used open-source brain analysis software tools, documented and publicly accessible from the FreeSurfer website (<http://surfer.nmr.mgh.harvard.edu>), which has gained considerable popularity in recent years. Existing validations on the FreeSurfer automated hippocampus segmentation have, despite some limitations particularly around similarity in absolute volumes between manual and automated estimations, concluded with some encouraging results with regards to the relative volumetric agreement. Overall, studies have demonstrated the FreeSurfer systematically inflated hippocampal volumes, although reported overestimation ranged from 4% to 75% (Akudjedu et al., 2018; Cherbuin, Anstey, Reglade-Meslin, & Sachdev, 2009; Morey et al., 2009; Sanchez-Benavides et al., 2010; Schmidt et al., 2018; Schoemaker et al., 2016). Reported intra-class correlation coefficients (ICC) ranged from moderate to good, although occasionally values as poor as .08 have been reported (Akudjedu et al., 2018; Schmidt et al., 2018; Schoemaker et al., 2016; Tae, Kim, Lee, Nam, & Kim, 2008; Wenger et al., 2014). However, it should be noted that most of these validation studies were based on older versions of FreeSurfer and there is evidence that the accuracy of automated segmentation of the medial temporal structures is improving with every software update (Clerx et al., 2015; Schmidt et al., 2018).

Recently, an updated hippocampal subfields segmentation protocol was introduced into FreeSurfer. This segmentation module is a statistical atlas developed from high resolution ex vivo elderly MRI data, designed to be robust to different magnetic field strengths and sequence acquisition, and was reported to produce more accurate representations of hippocampal volumes compared to the standard subcortical segmentation protocol (Iglesias et al., 2015). In a validation study by Schmidt et al. (2018), the new

FreeSurfer hippocampal subfields protocol was evaluated on 664 1.5T scans obtained from healthy adults ranging from ages 38 to 84 years. These authors concluded that the whole hippocampal subfields overestimations were smaller than the overestimations produced by the standard subcortical segmentation. The authors also reported better consistency ICCs of .72 and .70 in the whole hippocampal subfields protocol, compared to lower consistency ICCs of .34 and .32 in the standard subcortical segmentation protocol. However, their manual protocol excluded the hippocampal tail, and when they compared the manual estimations against the subfield estimates that excluded the hippocampal tail values, the automated values underestimated the hippocampal volumes and yielded poorer consistency ICCs of .51 and .48. The discrepancy in consistency ICCs between whole hippocampal volume estimates and hippocampal volume estimates that excluded the tail was not explored in the study. Moreover, their study only used scans of 1.5T field strength and given the wide age range of their sample, it is not implausible that biases could be present particularly in the younger populations (Wenger et al., 2014).

While FreeSurfer was designed to be robust against differing field strengths, the impact of field strength may be minor but still noteworthy. A study that evaluated the scans of 2790 individuals found that age, estimated total intracranial volume, and gender predicted more than 43% of the variance in the hippocampal volumes generated by FreeSurfer (v.5.3), while magnetic field strength accounted for 3.3% and 5.3% of the variance found in the left and right hippocampal volumes (Potvin, Mouiha, Dieumegarde, & Duchesne, 2016; Potvin, Mouiha, Dieumegarde, Duchesne, & Alzheimer's Disease Neuroimaging, 2016). The degree to which the variation for magnetic field strength would impact hippocampal estimations from a validation point of view remains unclear.

Aims of Current Study

This study was designed to investigate the performance of the new FreeSurfer (v.6.0) hippocampal subfield protocol and its ability to deal with hippocampal variability often found in clinical older populations. Firstly, direct validation analyses of both 1.5T and 3T scans would provide valuable additions to the current understanding of the impact of higher resolution scans on the hippocampal segmentation accuracy of FreeSurfer (v.6.0). Hence, this study investigated whether there were differences in the manual and automated hippocampal segmentation volumes of scans acquired on different field strengths. Secondly, the consistency in differences between FreeSurfer and manual estimations on routine clinical scans was evaluated, in which less ideal scan quality and more variability in hippocampal size and morphometry in a clinical cohort was expected.

Method

Study Population

A total of 90 clinical samples were drawn retrospectively from an archival database of memory clinic patients from St Vincent's Hospital Melbourne between the years 2006 – 2018. As part of their dementia clinical workup, they also undertook a comprehensive neuropsychological assessment that included the full administration of the complete WMS-III/IV whenever possible. Patients were diagnosed as having mild cognitive impairment or dementia according to recommended clinical diagnostic criteria (Albert et al., 2011; McKhann et al., 2011). Criteria for mild cognitive impairment would include evidence of insidious objective deterioration in a cognitive domain (typically >1.5 standard deviations), not attributable to other medical conditions. This change in cognition had to be coupled with objective evidence of impairment in one or more cognitive domains that was disproportionate to the person's age and educational background but without interference to daily function (Albert et al., 2011). The differentiation of dementia from mild cognitive impairment rested

on the neurologist's clinical judgement on whether there was significant interference with daily function as a result of the cognitive impairments (McKhann et al., 2011).

Patients who did not meet the diagnostic criteria for mild cognitive impairment or dementia, were considered as to have subjective memory concerns. Patients with a prior or current history of mild to moderate anxiety or depression were retained given the naturally increased prevalence of mood-related difficulties in this patient group. Consequently, these clinical patients were separated into three different groups for analysis according to the scanners used.

Clinical Symphony Sample: Patients from this sample underwent a routine MRI head scan on a 1.5T Symphony scanner between the years 2006 – 2014 ($n = 30$, mean age = 66.9 ± 10.7 years; 11 females). Sixteen patients in this sample were diagnosed as experiencing subjective memory concerns, five were diagnosed with mild cognitive impairment and nine were diagnosed with dementia.

Clinical Avanto Sample: Patients from this sample underwent a routine MRI head scan on a 1.5T Avanto scanner between the years 2007 – 2017 ($n = 30$, mean age = 66.8 ± 9.1 years; 20 females). Thirteen patients in this sample were considered to have subjective memory concerns, nine were diagnosed with mild cognitive impairment and eight were diagnosed with dementia.

Clinical Skyra Sample: Patients in this sample underwent a routine MRI head scan on a 3T scanner between the years 2013 – 2018 ($n = 30$, mean age = 68.9 ± 7.2 years; 17 females). Seventeen patients in this sample were deemed to be experiencing subjective memory concerns, six were diagnosed with mild cognitive impairment and seven were diagnosed with dementia.

General exclusionary criteria for all clinical groups were being below 50 years of age and having a significant history of an acquired head injury with persisting cognitive deficits.

Control Sample: A group of healthy control participants was recruited for this study from 2017 – 2018 ($n = 40$, mean age = 67.8 ± 5.4 years; 30 females). All control participants underwent neuroimaging on both the 1.5T Symphony scanner and 3T Skyra scanner on the same day, as well as a comprehensive neuropsychological assessment, which included a complete administration of the WMS-IV.

General exclusionary criteria for the control group were being below 50 years of age, having contraindications to undertaking an MRI, having significant current medical or neurological condition that would have impact on cognition and significant history of an acquired head injury with persisting cognitive deficits. All control participants included in this study had no unexpected findings on their cognitive assessments and MRIs. Analogous to the clinical group, control participants with a prior history or current symptoms of mild to moderate anxiety or depression were retained. This study was approved by the St. Vincent's Hospital (Melbourne) Human Research Ethics Committee (HREC-A 057-15).

MRI Acquisition

All T1-weighted scans were acquired coronally along the hippocampal axis with 3D-sequenced MPRAGE. The scan parameters for the three scanners used in this study were as the following: (1) 1.5T Siemens Avanto scanner: 12-channel head coils, TR = 1530, TE = 3.24, FOV = 156mm * 250mm, flip angle = 8° , voxel size $1 * 1 * 1.4\text{mm}^3$, 160 slices, matrix 256 * 160; (2) 1.5T Siemens Symphony scanner: CP head array coil, TR = 1530, TE = 3.41, FOV = 156mm * 250mm, flip angle = 8° , voxel size $1 * 1 * 1.4\text{mm}^3$, matrix 256 * 160; (3) 3T Siemens Skyra scanner, 64-channel head coil, TR = 2200, TE = 2.03, FOV = 230 * 230, flip angle = 8° , voxel size = $.9 * .9 * 1\text{mm}^3$, matrix 256 * 256.

There were occasional deviations from these exact parameters that resulted in some differences in total slice numbers, but in-plane resolution did not change. Both clinical and control scans were acquired using the same standard, clinical protocol. An additional,

consecutive T1-weighted MPRAGE was also acquired on the 3T scanner in the control group using the same scan parameters.

Volumetric Procedures

Automated segmentation. Automated segmentation of all scans was performed on a Linux workstation on one core over an approximate duration of seven to nine hours per scan. Automated segmentation of the hippocampus was executed using FreeSurfer (v6.0).

The FreeSurfer cortical reconstruction process was performed on raw DICOM images using the “recon-all” command. Details of this automated process is described in Fischl et al. (2002). An additional flag, “-hippocampal-subfields-T1”, was appended after the recon-all command to generate the values of the hippocampal subfields. The developers of this hippocampal subfield algorithm suggested that this tool provides a better estimate of whole hippocampal volumes compared to the volumes generated by the standard subcortical segmentation protocol and this subfield protocol was used in this study for validation (Iglesias et al., 2015). The volume statistics for the subfield protocol were extracted directly from the automatically generated text file. To more closely approximate the automated volume to the manual protocol used in this study (details in the following section), the hippocampal tail was subtracted from the whole hippocampal volume to generate the final whole hippocampal estimate used in this study.

Recon-all was executed using a single T1 scan, with the exception of the control Skyra sample, which had hippocampal volume estimates generated based on two 3T T1 scans to improve automated segmentation.

Manual segmentation. Volumes of the hippocampus were determined by manual delineation on the coronal T1-weighted slices using the ROI measure on Analyze 12.0 (Brain Imaging Resource, Mayo Foundation, Rochester, MN, USA).

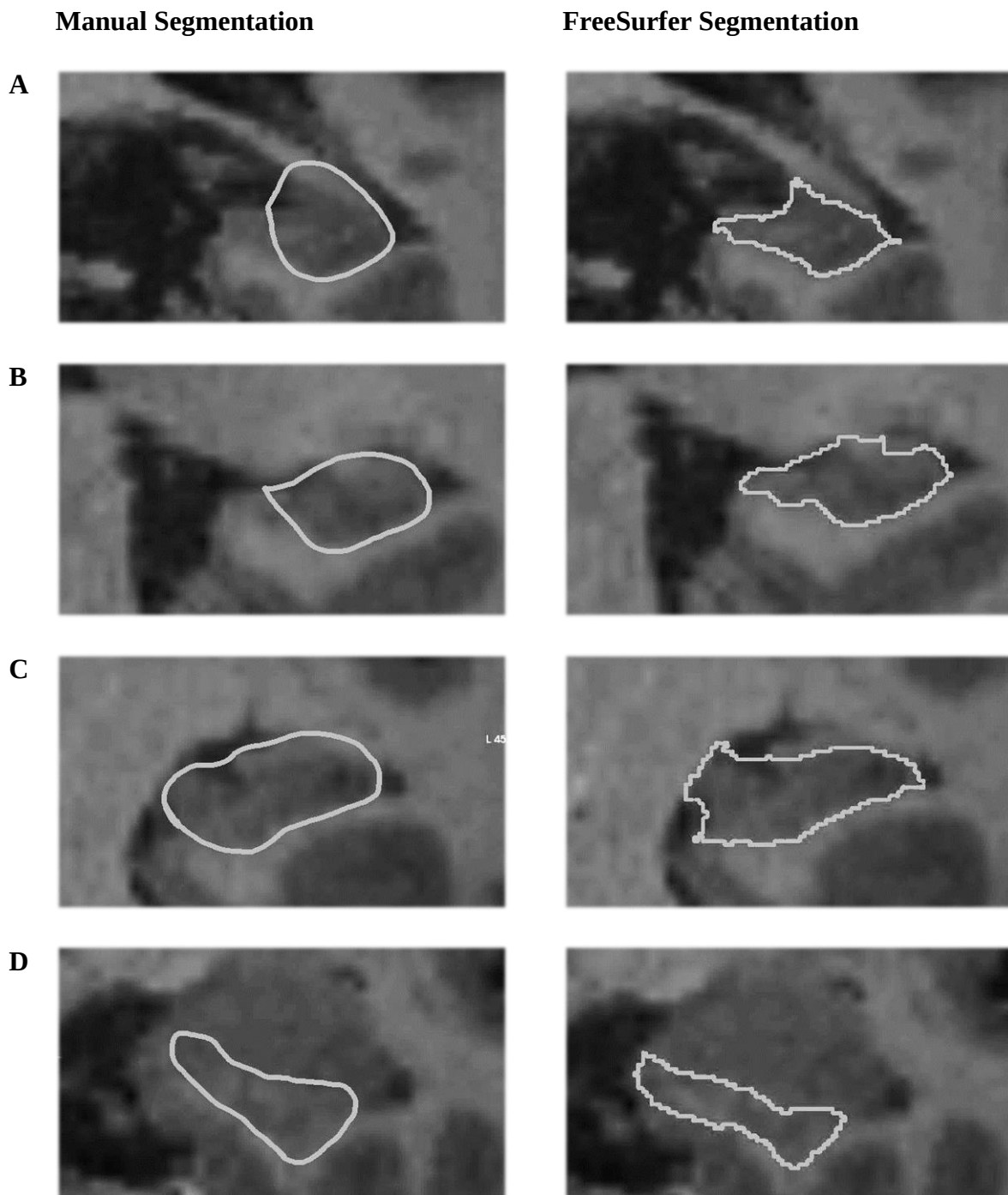
To prepare for manual tracing, raw DICOM files were converted into volumes of 0.3*0.3*1.4mm voxel size using cubic spline and saved into Analyze format. The x and y dimensions were converted to 0.3mm voxel size to improve the resolution for manual tracing on both 1.5T and 3T scans. Slice thickness for 3T scans was converted into 1.4mm for comparability with the 1.5T scans. Original slice thickness of 1.4mm was preserved in 1.5T scans.

The hippocampus was traced according to the protocol described in Cook, Fish, Shorvon, Straughan, and Stevens (1992). In short, the posterior limit of the hippocampus was the longest length of the unbroken fornix and the anterior limit was the last slice where the hippocampus could be discerned. The outer border of the hippocampus was determined by the alveus. The hemisphere in which manual segmentation began was counterbalanced to avoid bias. Figure 5.1 shows an example of the delineation of the hippocampus by both manual method and FreeSurfer.

Manual segmentation of the hippocampus was completed by a single investigator who was blinded to the diagnosis of the participant. Ten scans were randomly selected in each of the Clinical Symphony ($n = 10$, mean age = 71.5 ± 8.9 years), Clinical Avanto ($n = 10$, mean age = 70.8 ± 7.1 years) and Clinical Skyra ($n = 10$, mean age = 69.4 ± 6.9 years) samples, to assess for intra-rater reliability using intra-class correlation coefficient (ICC), using a two-way mixed, absolute agreement for a single measure (Cicchetti, 1994; Portney & Watkins, 2013). Good to excellent single-measure intra-rater ICCs of .97 to .99 was achieved in all three clinical samples on both left and right hemispheres. Inter-rater reliability was also assessed with another trained operator on the same group of scans and inter-rater ICCs in all three clinical samples ranged from moderate to excellent (ICCs = .86 – .97). Values of the single-measure ICCs and the 95% confidence interval can be found in Table 5.1.

Figure 5.1

Illustration of the Manual and FreeSurfer Automated Segmentation of the Hippocampus



Note. This figure illustrates the segmentation of the hippocampus from the caudal to rostral coronal slices. A: Longest uninterrupted length of fornix (posterior limit of hippocampus); B: Body of the hippocampus; C: Appearance of intralimbic gyrus; D: Slice containing the hippocampal head.

Table 5.1

*Intra- and Inter-rater Reliability Assessed by Intra-class Correlation Coefficient (ICC)
Between Manual Hippocampal Volume Estimations*

Group	Hemisphere	Intra-rater ICC	Inter-rater ICC
Clinical Avanto			
	Left	.97 (95% CI: .84, .99)	.88 (95% CI: .60, .97)
	Right	.98 (95% CI: .93, .99)	.89 (95% CI: .54, .97)
Clinical Symphony			
	Left	.98 (95% CI: .93, .99)	.86 (95% CI: .54, .97)
	Right	.99 (95% CI: .96, .99)	.93 (95% CI: .62, .98)
Clinical Skyra			
	Left	.99 (95% CI: .95, .99)	.97 (95% CI: .87, .99)
	Right	.99 (95% CI: .97, .99)	.93 (95% CI: .77, .98)

Note. Each group consist of 10 scans and all ICC values reported are based on two-way mixed, absolute agreement single-measure ICCs

Statistical Analysis

Statistical analysis was conducted using SPSS (v.25). To evaluate the similarity of volumes obtained on 1.5T Symphony scanner and 3T scanner in the control sample, the respective manual and automated hippocampal volumes were compared using paired-sample *t*-test.

Reliabilities between (1) 1.5T and 3T hippocampal volumes, and (2) automated and manual volume estimations, were evaluated using two-way mixed agreement ICC. The use of ICC was regarded as the appropriate estimation of reliability because it captures both the degree of correspondence and agreement among ratings in a single index (Portney & Watkins, 2013; Shrout & Fleiss, 1979). ICC values reported for (1) 1.5T and 3T hippocampal volumes, were based on absolute single measures, as the volume estimations were based on the scans from the same subject and therefore the hippocampal size was expected to be the same. The ICC values reported for (2) automated and manual volume estimations, were based

on both absolute and consistency values. Both values were reported as most studies have found relative differences between the manual and automated estimations. However, manual protocols utilised the various studies varied and the similarities in the absolute values between this current manual protocol and FreeSurfer estimations have not been previously determined.

Bland-Altman plots were used to visualise the level of agreement between (1) 1.5T and 3T hippocampal volumes, and (2) automated and manual volume estimations. For the Bland-Altman plots, the percentage difference between the two volumes for comparison was plotted against the mean of both volume estimations (Bland & Altman, 1999). The 95% limits of agreement were plotted to show the range of variability about the means and a simple regression line was drawn to illustrate whether there is proportional bias between the two methods.

Results

Descriptive Statistics

Control and clinical samples. When broken down into their different scanner groups, there were no significant differences in age between the control and three clinical samples ($F(3,126) = 0.41, p = .75$). However, the gender distribution between the four samples was significantly different $\chi^2(3, N = 130) = 11.25, p = .01$.

As shown in Table 5.2, ANOVA revealed significant differences of the left manual hippocampal estimates between the control and three clinical samples, Clinical Symphony, Clinical Avanto and Clinical Skyra, with a medium effect size, $\eta^2 = .08$ (Cohen, 1988). Although significant differences were not found in the right manual hippocampal estimates and total hippocampal estimates between the groups, small to medium effect sizes were demonstrated. Significant differences reflecting medium to large effects were found between the samples for all automated hippocampal estimates (Table 5.2).

Table 5.2

Comparisons of Manual and Automated Hippocampal Volume Estimations (mm³) in the Control and Clinical Samples

	Control sample (<i>n</i> = 40)		Clinical samples						ANOVA	
	<i>M</i>	<i>SD</i>	Symphony (<i>n</i> = 30)		Avanto (<i>n</i> = 30)		Skyra (<i>n</i> = 30)		<i>p</i> -value	η^2
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Manual										
Left hemisphere	3609.39	411.29	3315.82	601.73	3280.25 ^a	571.76	3275.91 ^a	480.61	.016*	0.08
Right hemisphere	3672.41	434.14	3477.52	560.47	3419.66	591.92	3492.91	475.47	.182	0.03
Total (left + right)	7281.80	828.01	6793.34	1124.32	6699.91	1129.61	6768.83	923.26	.053	0.05
FreeSurfer										
Left hemisphere	2774.43	300.97	2564.06	434.58	2386.17 ^a	458.09	2505.83 ^a	359.42	.001*	0.13
Right hemisphere	2819.40	294.03	2657.23	376.31	2463.87 ^a	446.57	2642.64	329.67	.001*	0.11
Total (left + right)	5593.83	582.31	5221.29	787.66	4850.04 ^a	876.42	5148.46	656.00	.001*	0.13

Note. All automated volume estimations were performed using FreeSurfer (v.6.0). Volumes from the Control sample was calculated from the average of the volumes obtained from the Symphony and Skyra scans. ANOVA with Tukey post hoc was conducted for groups with significantly different means, as indicated by an asterisk. Means that are significantly from the control group at $p < .05$ in the post hoc test is indicated with a superscript “a”. η^2 = effect size.

Comparability Between Scanners

Using the control sample volume estimations, as shown in Table 5.3, paired sample *t*-test showed no significant difference between the hippocampal volumes derived from manual measures between both scanners, Symphony and Skyra, on both the left and right hemispheres and effect sizes were all trivial. Also shown in Table 5.3, the ICC values for left, right and total hippocampal manual estimates were all excellent, achieving single measure absolute ICCs of .96, .98, .98 respectively. Bland-Altman plots of these structures did not indicate any proportional bias (Figure 5.2).

However, FreeSurfer hippocampal estimates were significantly different between volumes obtained from the 1.5T scanner and 3T scanner. As detailed in Table 5.3, paired-sample *t*-test demonstrated that automated volume estimates on Symphony 1.5T were significantly smaller than automated volume estimates on the 3T scans for the left hippocampus by 1.38% ($p = .004$), right hippocampus by 1.89% ($p < .001$) and total hippocampal volumes by 1.63% ($p < .001$), with small to large effects, Cohen's $d = 0.49$, 0.79 and 0.77 , respectively. Absolute single-measure ICC values were excellent on the left hippocampal volumes, but more variable on the right and total hippocampal volumes, ranging from good to excellent (Table 5.3). Bland-Altman plots were not suggestive of a proportional bias between these two scanners (Figure 5.2).

Reliability and Agreement Between Manual and Automated Measures

Control sample. All absolute and consistency single-measure ICC values, with the corresponding 95% confidence intervals, for the control sample can be found in Table 5.4. Absolute single-measure ICCs between manual and automated estimates obtained from the Symphony scans, for left, right and total hippocampal volumes was poor to moderate. However, when consistency rather than absolute single-measure ICC was used, the ICC values improved and ranged from moderate to good on the left hippocampus, and moderate to

Table 5.3

Comparisons of Manual and Automated Hippocampal Volume Estimations Obtained on Different Scanner Magnetic Field Strengths in the Control Sample

	Symphony		Skyra		Paired-sample <i>t</i> -test				Absolute ICC	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Percentage difference ^a	<i>t</i> -value	<i>p</i> -value	Cohen's <i>d</i>	Single-measure	95% confidence interval
Manual										
Left hippocampus	3605.16	416.02	3613.63	414.36	-0.23%	-0.47	.640	-0.07	.96	.93 , .98
Right hippocampus	3675.92	435.22	3668.90	436.59	0.19%	0.57	.574	0.09	.98	.97 , .99
Total hippocampus	7281.07	833.58	7282.54	831.53	-0.02%	-0.05	.958	-0.01	.98	.96 , .99
FreeSurfer										
Left hippocampus	2755.21	303.01	2793.64	304.13	-1.38%	-3.07	.004*	-0.49	.96	.91 , .98
Right hippocampus	2792.56	293.12	2846.25	298.80	-1.89%	-5.01	<.001*	-0.79	.96	.82 , .98
Total hippocampus	5547.77	583.66	5639.90	587.16	-1.63%	-4.84	<.001*	-0.77	.97	.87 , .99

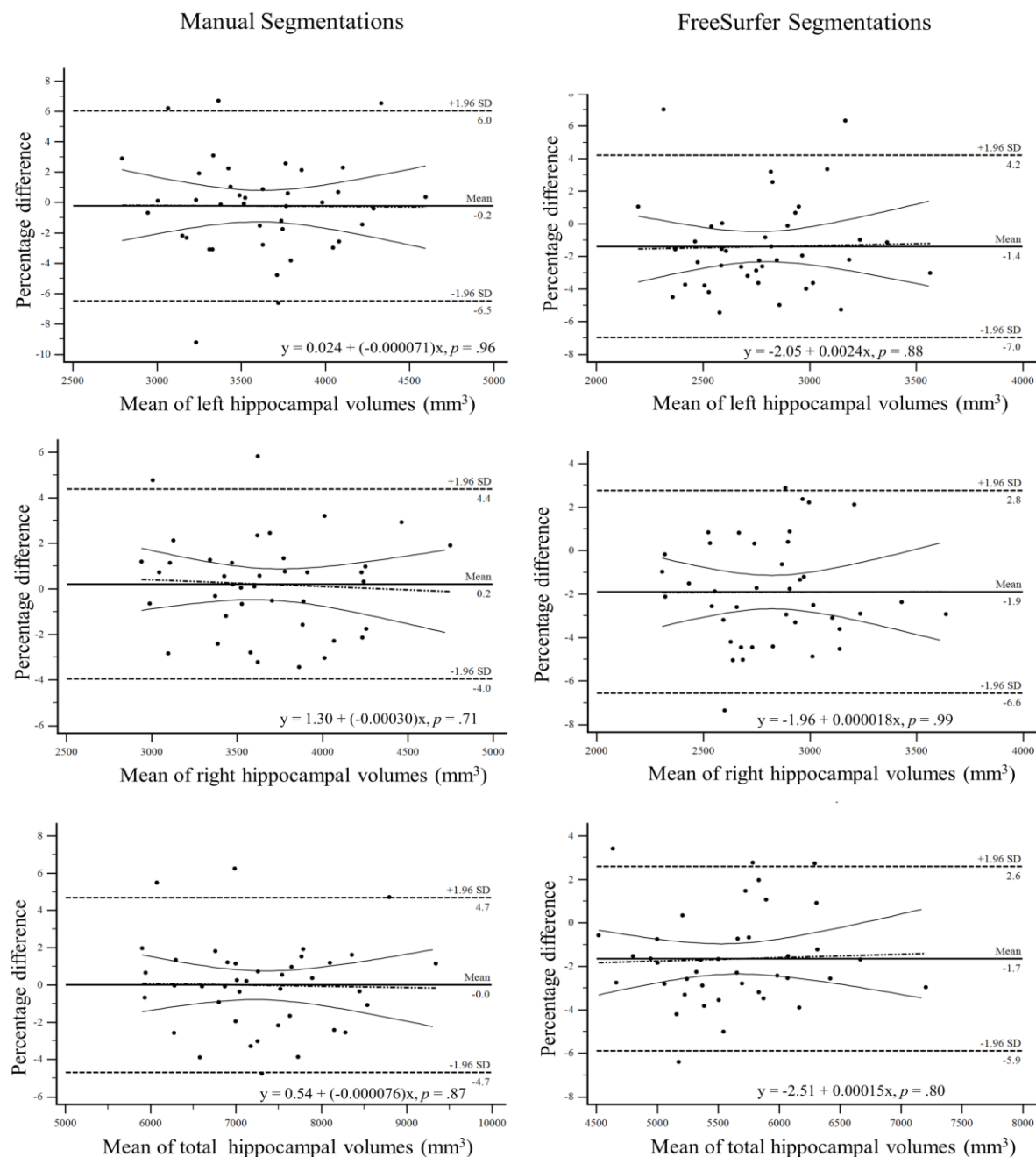
Note. All reported volumes are in mm³. All automated volume estimations were performed using FreeSurfer (v.6.0). Separate manual and automated volumes from the control sample (*n* = 40) was obtained from two scanners, 1.5T scanner Symphony and 3T scanner Skyra.

Comparability of volume estimations was assessed using paired-sample *t*-test and two-way mixed, absolute intraclass correlation coefficient (ICC).

^a Percentage difference was calculated using (Symphony estimates – Skyra estimates) / (Skyra estimates) * 100

Figure 5.2

Bland-Altman Plot of Agreement Between Hippocampal Volumes from Different Scanners in the Control Sample



Note. Bland-Altman plots of agreement between the left, right and total hippocampal volumes obtained on different scanners in the control sample ($n = 40$), based on manual and FreeSurfer estimates. Scans were obtained from the same control subjects on each scanner. The percentage difference (Symphony – Skyra estimates) was plotted against the average Symphony versus Skyra difference estimates. Solid lines represent the mean percentage difference between the scanners, dashed lines represent the upper and lower 95% limits of agreement. Regression line is represented by the broken line flanked by its 95% confidence interval.

excellent on the right and total hippocampal volumes (Table 5.4). Automated volume estimates underestimated the left hippocampal volume by 26.6% ($SD = 6.4$), underestimated the right hippocampal volume by 27.1% ($SD = 5.7$) and overall underestimation of total sum of the left and right hippocampal volume was 26.9% ($SD = 5.4$). There were a few outliers observed on the Bland-Altman plots for left, right and total hippocampal volume estimations, but overall regression analysis did not suggest a proportional bias for any of the volumes (Figure 5.3).

The ICC values from the Skyra scans in the control sample were similar to those found in the Symphony scans. Absolute single-measure ICC values obtained from the control Skyra scans ranged from poor to moderate in left, right and total hippocampal volumes. When consistency single-measure ICCs were used instead, ICC values obtained from the comparison of estimates in the control Skyra scans improved to moderate to good in the left, right and total hippocampal volumes. Again, like the findings in the control sample with the Symphony scans, automated segmentation underestimated left, right and total hippocampal volumes by 25.5% ($SD = 6.7$), 25.1% ($SD = 6.0$) and 25.3% ($SD = 5.9$) respectively. As seen the Bland-Altman plots, there were a few outliers, with the distribution of outliers similar to that seen in the Symphony scans in the control sample (Figure 5.3). Again, regression analysis was not suggestive of any proportional bias (Figure 5.3).

Clinical samples. All absolute and consistency single-measure ICC values, with the corresponding 95% confidence intervals, for the clinical samples can be found in Table 5.4. Absolute single-measure ICC values in the clinical Symphony sample ranged from poor to good in the left and total hippocampal volumes, and poor to moderate in right hippocampus (Table 5.4). Comparatively, consistency single-measure ICCs between the manual and automated measures in this sample was better than absolute single-measure ICC, demonstrating moderate to excellent degrees of consistency on all left, right and total

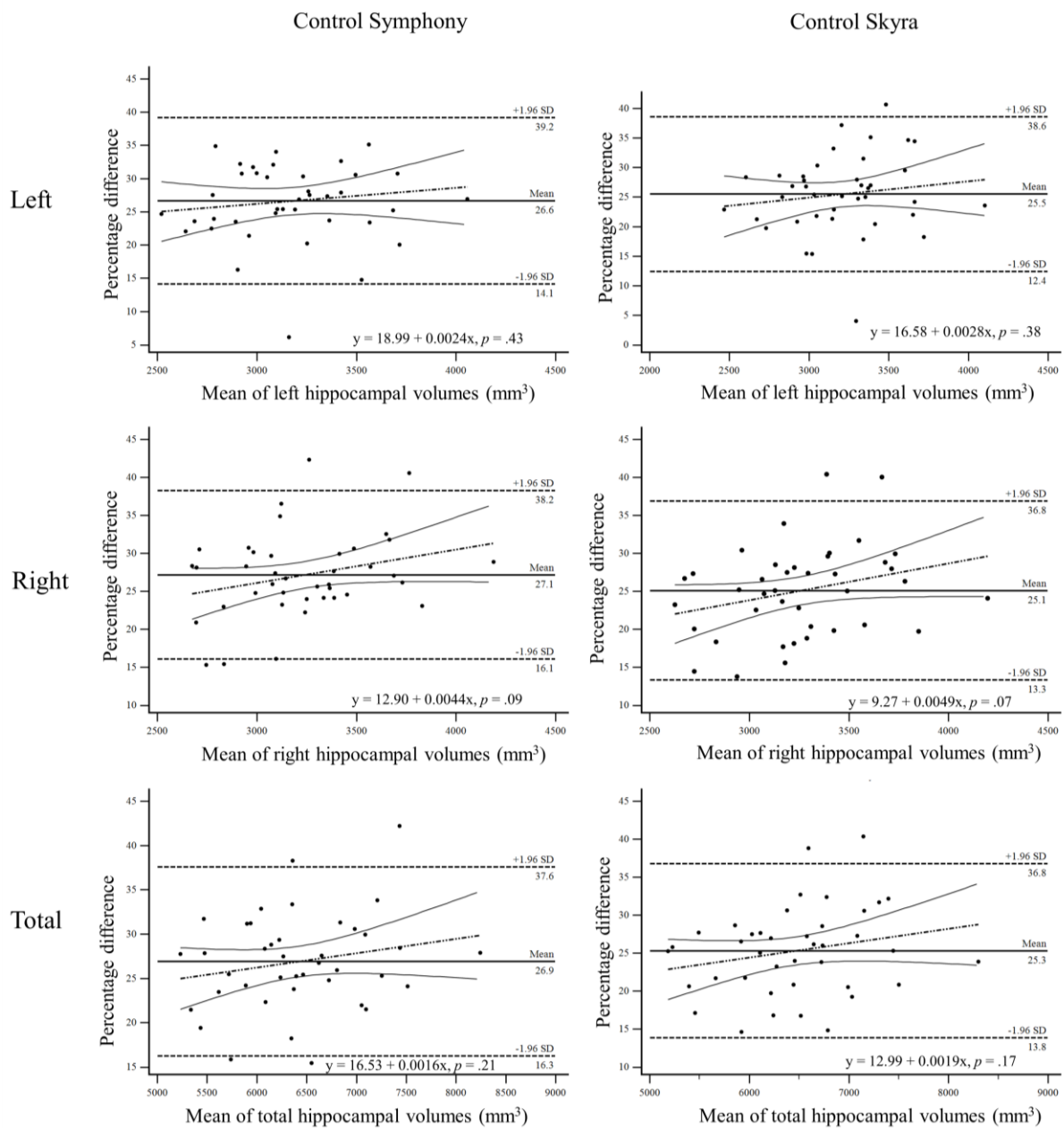
Table 5.4*Intra-class Correlation Coefficient Between Manual and Automated Hippocampal Volume Estimations*

	Single-measure ICC values											
	Absolute						Consistency					
	L	95% CI	R	95% CI	Total	95% CI	L	95% CI	R	95% CI	Total	95% CI
Control												
Symphony (<i>n</i> = 40)	.21	-.04, .57	.21	-.03, .58	.21	-.03, .58	.78	.63, .88	.81	.67, .90	.82	.68, .90
Skyra (<i>n</i> = 40)	.22	-.04, .58	.23	-.04, .61	.22	-.03, .59	.77	.60, .87	.80	.65, .89	.80	.65, .89
Clinical												
Symphony (<i>n</i> = 30)	.43	-.05, .79	.34	-.05, .73	.38	-.04, .76	.87	.74, .94	.84	.70, .92	.87	.75, .94
Avanto (<i>n</i> = 30)	.36	-.03, .75	.34	-.03, .73	.34	-.03, .73	.89	.78, .95	.90	.80, .95	.91	.81, .95
Skyra (<i>n</i> = 30)	.30	-.05, .68	.26	-.04, .65	.26	-.04, .65	.79	.60, .89	.82	.66, .91	.80	.63, .90
1.5T scanner (<i>n</i> = 100)	.35	-.04, .72	.31	-.04, .68	.32	-.04, .70	.86	.80, .90	.85	.78, .89	.87	.81, .91
3T scanner (<i>n</i> = 70)	.29	-.05, .66	.26	-.04, .63	.26	-.04, .63	.81	.71, .88	.82	.72, .88	.82	.73, .89

Note. Automated volume estimations were performed using FreeSurfer (v.6.0). Reliability between manual and automated measures were assessed using mixed model intraclass correlation coefficient (ICC). Both absolute and consistency single measure ICCs were reported, as well as the corresponding 95% confidence interval (95% CI). The 1.5T scans were obtained from Symphony and Avanto scanners, the 3T scans were obtained from Skyra scanner.

Figure 5.3

Bland-Altman Plot of Agreement Between Manual and Automated Hippocampal Volumes in the Control Sample



Note. Bland-Altman plot of agreement between manual and automated estimates of the left, right and total hippocampal volumes of the control sample ($n = 40$), based on different scanners. Scans were obtained from the same subjects on each scanner. The percentage difference plot (manual – automated estimation) was plotted against the averaged manual versus automated volumes. Solid lines represent the mean percentage difference volumes between methods, while dashed lines represent the upper and lower 95% limits of agreement. Regression line is represented by the broken line flanked by its 95% confidence interval.

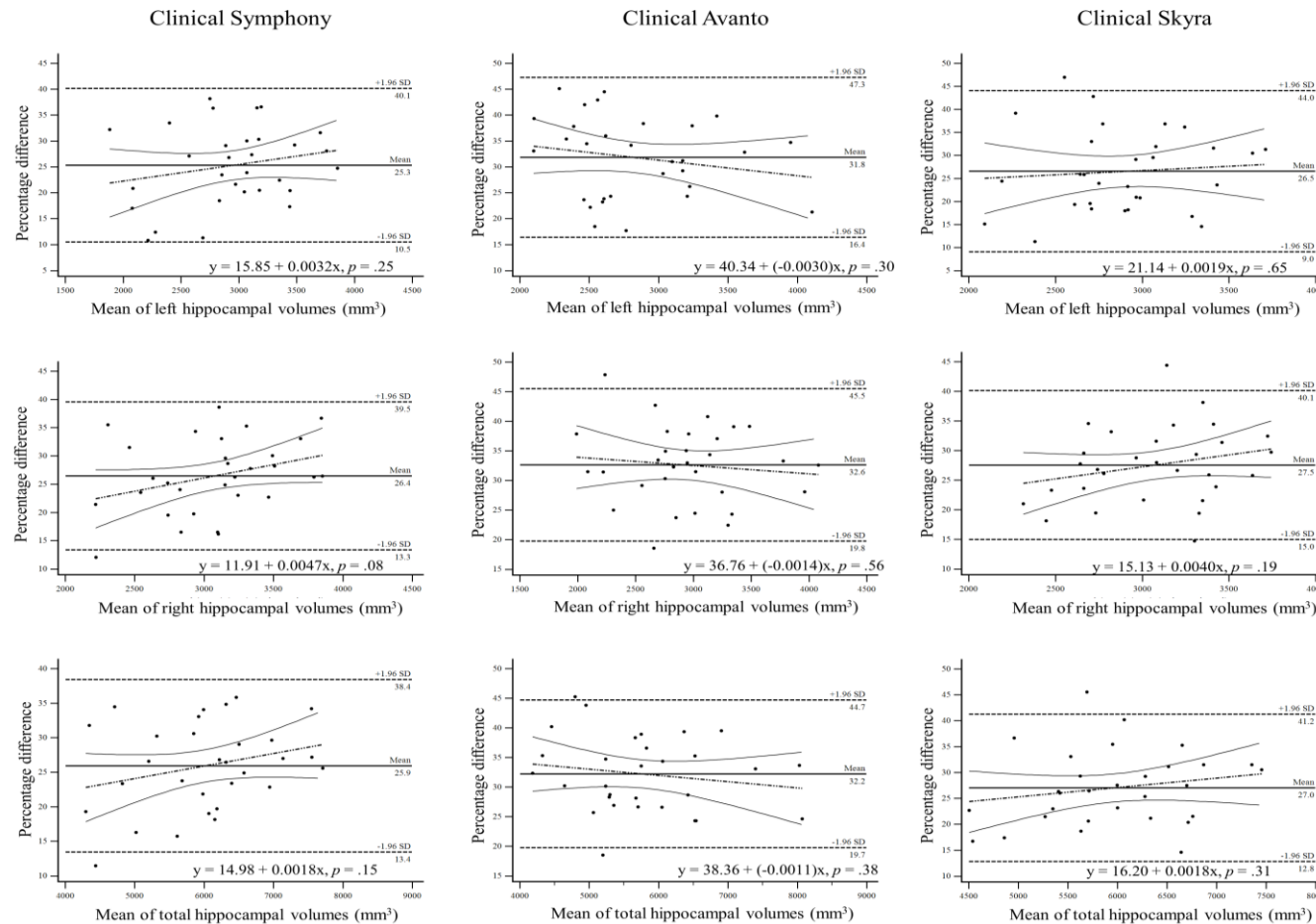
hippocampal volume estimations (Table 5.4). On average, in the clinical Symphony sample, automated volume segmentation underestimated the left hippocampus by 25.3% ($SD = 7.5$), underestimated the right hippocampus by 26.4% ($SD = 6.7$) and underestimated total hippocampal volume by 25.3% ($SD = 6.4$). There were no suggestions of proportional biases and very few outliers were observed in the respective Bland-Altman plots (Figure 5.4).

Estimates of absolute agreement between manual and automated measures in the clinical Avanto sample generally produced poor to moderate absolute single-measure ICCs. However, consistency measures were relatively better, producing good to excellent consistency single-measure ICCs on all comparisons (Table 5.4). Automated volume estimates underestimated the hippocampal volumes on the left by 31.8% ($SD = 7.9$), on the right by 32.6% ($SD = 6.6$), and on overall total volume by 32.2% ($SD = 6.4$). On the Bland-Altman plots, there were a few outliers observed, one in the right hippocampal volumes plot and two in the total hippocampal volume plot, but overall, there were no suggestion of proportional bias in any of the graphs (Figure 5.4).

Lastly, absolute single-measure ICCs in the clinical Skyra sample were poor to moderate. However, consistency single-measure ICCs were relatively better, with values ranging from moderate to good in the left hippocampal estimates, and moderate to excellent in the right and total hippocampal estimates (Table 5.4). In general, hippocampal estimations by FreeSurfer was smaller than that of manual estimations, underestimating hippocampal volumes by 26.5% ($SD = 8.9$) in the left hemisphere, 27.5% ($SD = 6.4$) in the right hemisphere, and 27.0% ($SD = 7.2$) in the total sum of the left and right hippocampal volumes. The Bland-Altman plots of the left, right and total hippocampal volumes showed limited outliers and regression lines were not suggestive of proportional biases (Figure 5.4).

Figure 5.4

Bland-Altman Plots of Agreement Between Manual and Automated Hippocampal Volumes in the Clinical Sample



Note. Bland-Altman plot of left, right and total hippocampal volumes of the different clinical samples ($n = 30$ in each sample). The percentage difference plot (manual – automated estimation) was plotted against the averaged manual versus automated volumes. Solid lines represent the mean percentage difference volumes between methods, while dashed lines represent the upper and lower 95% limits of agreement. Regression line is represented by the broken line flanked by its 95% confidence interval.

Agreement Between Measurements of Different Scanner Strength

1.5T scanner. All absolute and consistency single-measure ICC values, with the corresponding 95% confidence intervals for the estimates obtained from 1.5T scans can be found in Table 5.4. These volume estimates comprised of all the scans ($n = 100$) from the clinical and control samples that were acquired either on the Avanto or Symphony scanners. Absolute single-measure ICCs between manual and automated hippocampal estimates were poor to moderate on the left, right and total hippocampal estimates. However, consistency single-measure ICCs were higher, ranging from good to excellent in the left and total hippocampal estimates, and good in the right hippocampal estimates (Table 5.4).

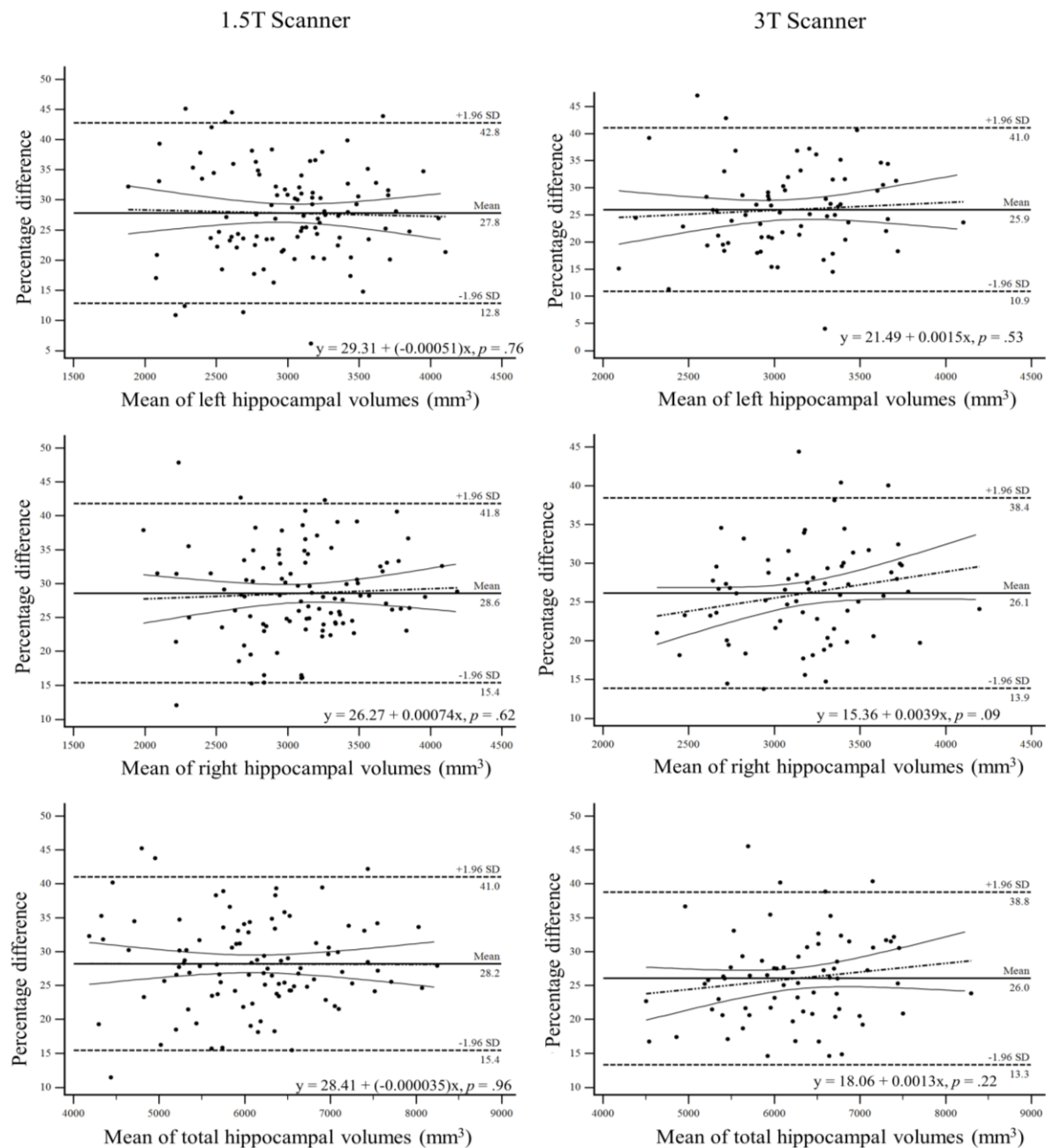
Automated volume measurements underestimated the left hippocampus by 27.8% ($SD = 7.9$), underestimated the right hippocampus by 28.6% ($SD = 7.8$), and underestimated total hippocampus by 28.2% ($SD = 7.1$). The Bland-Altman plots with the respective regression lines did not suggest any proportional biases (Figure 5.5).

3T scanner. All absolute and consistency single-measure ICCs, with the corresponding 95% confidence intervals for the estimates obtained from 3T scans can be found in Table 5.4. These volume estimates consisted of all the scans ($n = 70$) from the Clinical Skyra and Control Skyra samples. Similar to the findings from the 1.5T scanner, absolute single-measure ICCs between manual and automated methods were poor to moderate on the left, right and total hippocampal volume comparisons. On the other hand, consistency single-measure ICCs were better, ranging from moderate to good on all comparisons (Table 5.4).

In terms of level of agreement, automated volume estimates underestimated all hippocampal volumes, 25.9% ($SD = 7.7$) for the left, 26.1% ($SD = 6.3$) for the right, and 26.0% ($SD = 6.5$) for the total hippocampal volumes. The Bland-Altman plots with the respective regression lines were not indicative of directional biases (Figure 5.5).

Figure 5.5

Bland-Altman Plots of Agreement Between Manual and Automated Hippocampal Volumes in All Samples



Note. Bland-Altman plot of left, right and total hippocampal volumes estimated from both clinical and control samples, separated into different scanner strengths. The 1.5T scans ($n = 100$) were obtained from the Symphony and Avanto scanners, while 3T scans ($n = 70$) were obtained from the Skyra scanner. The percentage difference plot (manual – automated estimation) was plotted against the averaged manual and automated volumes. Solid lines represent the mean percentage difference volumes between methods, while dashed lines represent the upper and lower 95% limits of agreement. Regression line is represented by the broken line flanked by its 95% confidence interval.

Discussion

In this study, the performance of the new FreeSurfer (v.6.0) hippocampal subfields protocol was investigated in a clinical population, using scans acquired on routine clinical scan parameters. Firstly, whether scans acquired using different magnetic field strength would result in different hippocampal volume estimates were evaluated. The results found that while manual estimations based on different scanner strengths were similar, FreeSurfer automated algorithm generated hippocampal volume estimates on the 1.5T scans that were 1.37% to 1.89% smaller than the 3T scans. The findings of Potvin and his colleagues have demonstrated that after age, estimated total intracranial volume and gender, magnetic field strength accounts for most the variance found in FreeSurfer segmentation volumes (Potvin, Mouiha, Dieumegarde, & Duchesne, 2016; Potvin, Mouiha, Dieumegarde, Duchesne, et al., 2016).

The findings from this study suggest that this volume variability due to different magnetic field strengths was not a trivial difference, especially considering the average annual atrophy rates that have been reported in the literature earlier. For example, Schmidt et al. (2018) reported annual atrophy rates in a group of 55-70 years old adults ($n = 364$) to be 0.34% and 0.43% - 0.49%, as estimated by manual methods and FreeSurfer respectively. Similarly, Fraser et al. (2018) reported similar annual atrophy rates of 0.30% using manual methods and 0.51% using FreeSurfer, in their sample of 190 healthy older adults between the ages of 60 to 64. These results suggest that changes in scanners with different magnetic field strengths would likely introduce noise into the segmentation volumes and undermine the sensitivity of detection of the rate of hippocampal atrophy over time. Given the results, it appears vital that scanners with the same field strength should be used, particularly in longitudinal follow-ups. However, scanners do get replaced with newer and more powerful magnets, and scanner software is upgraded occur over time. A practical approach to minimise

estimation differences would be to use computational methods to correct for scanner-dependent variabilities (e.g. Mirzaalian et al., 2016), but such correction methods will require further research development and independent validations beyond the scope of this study.

The validation of the segmentation of FreeSurfer subfield protocol against manual measurements across the clinical and control samples, as well as across different scanners, was also undertaken as part of this study. The 1.5T scans from the clinical samples were obtained from either a Symphony and Avanto scanner, and 3T scans from the clinical samples were obtained using a Skyra scanner, while the control sample received scans on both the Symphony and the Skyra scanners. Absolute agreement between methods was poor on all comparisons, which was not a surprising finding, as previous research has consistently reported systematic differences between manual and FreeSurfer hippocampal volume estimates (Akudjedu et al., 2018; Cherbuin et al., 2009; Morey et al., 2009; Sanchez-Benavides et al., 2010; Schoemaker et al., 2016). These systematic differences varied not just according to the version of FreeSurfer or the automated segmentation modules used, but also differed according to the manual protocol adopted for segmentation (Schmidt et al., 2018; Wisse et al., 2017; Yushkevich, Amaral, et al., 2015).

Given the limitations of estimating absolute agreement between measures, consistency measures would serve as a better estimate of reliability based on relative agreement between the two methods. In this study, consistency measures for all the scans from the different groups reflected moderate to high levels of consistency, with the worst consistency single-measure ICC being .77 to the best being .91, and most of the consistency single-measure ICCs falling around .80 - .90 range. When the scans from the clinical and control samples were combined to form a group made up of 1.5T scans ($n = 100$) and another group made up of 3T scans ($n = 70$), consistency single-measure ICCs were also good, ranging from .81 to .87. These results are similar to, but mostly better than the ICCs reported

in other studies that used the standard segmentation protocol on older FreeSurfer versions (Akudjedu et al., 2018; Schoemaker et al., 2016; Tae et al., 2008; Wenger et al., 2014). The relatively higher consistency measures in this current study could be evidence suggestive of an improvement in the newer FreeSurfer (v.6.0) software, coupled with an enhanced hippocampal segmentation algorithm. This finding would be in line with the conclusions from previous studies that had compared different FreeSurfer versions and hippocampal segmentation protocols and reported improvement with newer versions of the software (Clerx et al., 2015; Schmidt et al., 2018).

Schmidt et al. (2018) also used the newest version of FreeSurfer (v.6.0) to examine the hippocampal subfields module. However, they found poor to moderate ICCs when they compared the automated estimations in the values that excluded the hippocampal tail, while the inter-rater ICCs were much more accurate in this study. One possible reason for the higher levels of consistency in this study is that the study population only included people above the age of 50 years. On the other hand, in Schmidt's et al. (2018) study, while much larger at a total sample size of 664, included a much wider age range of 38 to 84 years old, 171 of whom were younger than 55 years old. Therefore, it is not inconceivable that the inclusion of the younger participants might have introduced unintended biases in their estimations.

An age-deferential segmentation bias was seen in the study by Wenger et al. (2014), who reported that the longitudinal processing scheme found in FreeSurfer (v.5.3) overestimated the hippocampal volumes of their 20-30 year-old participants ($n = 44$) to a larger degree than the hippocampal volumes of their 60 – 70 year-old participants ($n = 47$). Nevertheless, it is also possible that an age-related bias is particularly salient with the hippocampal subfields module because this new atlas was designed based on scans from elderly subjects. Some other differences, albeit likely comparatively minor, were that

Schmidt's et al. (2018) study used older scans obtained from 2001 to 2006 and were acquired with slice thickness of 1.6mm, while scans included in this study were sourced post-2006 and acquired at a slightly finer slice thickness of 1.4mm. These differences in scan quality and scan resolution could have resulted in different biases being introduced during manual segmentations and during the re-sampling of scans into isotropic resolutions for FreeSurfer processing.

Although most studies have reported inflated automated hippocampal volumes based on estimations by FreeSurfer, in contrast to previous findings, the results from this study showed that automatic hippocampal volumes generated by the hippocampal subfields protocol were smaller than manual hippocampal volumes (Akudjedu et al., 2018; Cherbuin et al., 2009; Morey et al., 2009; Sanchez-Benavides et al., 2010; Schoemaker et al., 2016); these results are comparable to the findings in Schmidt et al. (2018). Even though the standard hippocampal segmentation pipeline produced increased volumes in their study, the whole hippocampal subfield volumes that excluded the hippocampal tail were smaller than manual volumes, in which their manual protocol excluded the hippocampal tail.

The volumes obtained from clinical scans in this study had slightly larger deviations around the mean compared to the control sample that used the same scanners, and this variability could be attributable to the clinical scans being older acquisitions and to variabilities in the hippocampal structures that may be more common in a clinical population. However, the deviations around the average difference did not appear to be excessively large, with the clinical scans having standard deviations that were approximately 1 – 2% wider than the controls. Reassuringly, there was very little difference in the degree of underestimation between scans obtained from clinical participants and control participants employing automated volumetry, with most numbers varying from 25.1% to 27.5%. The only notable difference was in the scans obtained on the Avanto scanner, in which there were larger

underestimations, with percentages in the low 30s. The smaller average hippocampal volumes of the scans acquired on the Avanto scanner could account for this observation, as it would result in the same absolute volumetric difference generating a larger percentage value.

Another encouraging finding in terms of agreement between the methods was that overall, scans from 1.5T and 3T yielded similar degrees of discrepancies, and similar 95% confidence intervals. The confidence intervals from the 3T scans was slightly narrower than that of 1.5T, which could be expected since segmentation accuracy is expected to increase with better scan resolution, but on the whole, there was evidence that FreeSurfer (v6.0) hippocampal subfields segmentation protocol performs similarly regardless of scanner field strength.

A possible limitation of this study is that the hippocampal tail was excluded in both the manual and automated hippocampal volume estimations. Iglesias et al. (2015) highlighted that the neuroanatomical literature surrounding the hippocampal tail is lacking, which makes reliable automated annotations in this section of the hippocampus difficult. Manually, it can also be difficult to visually discriminate the hippocampal tail from the adjacent retro-splenial cortex (Morris, Paxinos, & Petrides, 2000). This challenging discrimination between the hippocampal tail and adjacent cortex is especially the case in poor quality scans, occurring mostly due to movement artefacts. Hence, systematically excluding the hippocampal tail would likely yield more consistency between the automated and manual estimations.

Another possible limitation of this study is that the evaluation of the hippocampal segmentations was based only on T1 scans. With the new hippocampal subfield module, there is an option to include an additional T2 scan for processing, which would potentially improve hippocampal boundary discrimination (Iglesias et al., 2015). Nonetheless, it should be noted that T2 scans are more sensitive to motion artefact because of their longer pulse sequence times and might be less suitable for effective application in a clinical population

(Habas, 2018). Patients are prone to movement during scans and re-scans are not always feasible in hospital settings in which there are usually considerable time constraints on MRI scanner use due to high demand.

There were two other scan-related limitations in this study. Firstly, the scans were not acquired using isotropic resolutions. Regardless of the original scan acquisition resolution, the scans were re-sampled to match the dimensions required for manual segmentation and re-sampled for processing in FreeSurfer. Secondly, the clinical scans were acquired with varying scan parameters over several years and there would have been several scanner upgrades over that time period. However, in contrast, control scans were taken on the same scanners using the same standard scanning parameters. When compared against the clinical scans acquired over a lengthy time period, the ICCs and levels of agreement were similar between the control and the clinical samples (Table 5.4; Figure 5.3 and Figure 5.4). Even though the clinical scans from 1.5T Avanto scanner did not have a control group for direct comparison the older clinical scans obtained on the Symphony showed very good reliability and performed similarly to the Symphony scans from the control sample. With previous research showing that scanner manufacturer has negligible impact on FreeSurfer hippocampal volume estimations, it appears reasonable to infer that scans acquired on different 1.5T models from the same scanner manufacturer (Siemens) would result in similar scan quality (Potvin et al., 2016). While some might argue that the variations in scan resolutions and scanning parameters would have negative impacts on the segmentation results, the positive findings across all the scans strongly suggest that FreeSurfer hippocampal subfield segmentation module is robust against such variables.

It is recognised that a multi-modal diagnostic system - beyond solely cognitive and behavioural measures - is essential for the early detection of AD. Extensive research has pointed towards the identification of pathophysiological changes on structural MRI to being

one of the most powerful tools currently available to improve the diagnostic process (Alberdi et al., 2016; Bauer et al., 2018; Belleville et al., 2014). FreeSurfer can offer a highly efficient method for generating quantitative information to support clinical diagnosis if streamlined into part of a routine diagnostic workup. Therefore, the potential translational diagnostic utility of FreeSurfer hippocampal segmentation was investigated by testing its performance on “real world” clinical scans and compared it against a healthy control group.

Conservatively, it is recommended by Portney and Watkins (2013) that ICC values should be more than .90 for clinical measurements. Although ICC values in this study were below this criterion, FreeSurfer estimates performed similarly on both control scans and routinely acquired diagnostic MRI brain scans. By and large, FreeSurfer segmentation demonstrated good reliability without any manual intervention required during processing, which highlights its potential diagnostic utility in a clinical context. Overall, despite the differences in absolute volumes in manual and automated estimations, the degree in which these differences in volumetric assessments is acceptable is ultimately a decision based on clinical criterion and the ways in which these measurements would be applied (Bland & Altman, 1999). Clinically, if diagnostic accuracy is not compromised by measurement variability, these systematic methodological differences could be deemed acceptable.

The findings from this study do not suggest complete interchangeability between the methods but do lend support to the efforts of prospective investigations to continually narrow measurement variability and translate FreeSurfer into a clinically relevant diagnostic tool. Although segmentations of scans acquired on different field strengths did produce slightly different absolute values, from a cross-sectional point of view, the automated segmentation performed similarly on the validity measures. However, such discrepancies would be important to address from a longitudinal point of view. On-going and future normative and methodological studies could aim to provide correctional methods to minimise these

differences and henceforth improve the ability for FreeSurfer to accurately quantify intra-subject rates of volumetric atrophy over time.

In summary, in this sample of older adults, hippocampal volume estimations generated from the hippocampal subfields package in FreeSurfer (v.6.0) showed good reliability and systematic correlation with manual measurement in scans obtained from both research participants and memory clinic patients, under a variety of differing scan conditions. This pattern of findings demonstrates the robustness of the FreeSurfer hippocampal segmentation algorithms in dealing with variable hippocampal structures, magnetic field strength and scan parameters, hence suggesting a strong potential for its translation into a clinical diagnostic tool.

References

- Akudjedu, T. N., Nabulsi, L., Makelyte, M., Scanlon, C., Hehir, S., Casey, H., . . . Cannon, D. M. (2018). A comparative study of segmentation techniques for the quantification of brain subcortical volume. *Brain Imaging and Behavior*, 12(6), 1678-1695.
doi:10.1007/s11682-018-9835-y
- Alberdi, A., Aztiria, A., & Basarab, A. (2016). On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial Intelligence in Medicine*, 71, 1-29.
doi:10.1016/j.artmed.2016.06.003
- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to 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), 270-279. doi:10.1016/j.jalz.2011.03.008
- Atiya, M., Hyman, B. T., Albert, M. S., & Killiany, R. (2003). Structural magnetic resonance imaging in established and prodromal Alzheimer disease: a review. *Alzheimer Disease and Associated Disorders*, 17(3), 177-195. doi:10.1097/00002093-200307000-00010
- Bauer, C. M., Cabral, H. J., & Killiany, R. J. (2018). Multimodal discrimination between normal aging, mild cognitive impairment and Alzheimer's disease and prediction of cognitive decline. *Diagnostics*, 8(1), 20. doi:10.3390/diagnostics8010014
- Bayram, E., Caldwell, J. Z. K., & Banks, S. J. (2018). Current understanding of magnetic resonance imaging biomarkers and memory in Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 4, 395-413.
doi:10.1016/j.trci.2018.04.007

- Belleville, S., Fouquet, C., Duchesne, S., Collins, D. L., & Hudon, C. (2014). Detecting early preclinical Alzheimer's disease via cognition, neuropsychiatry, and neuroimaging: qualitative review and recommendations for testing. *Journal of Alzheimer's Disease*, 42 Suppl 4, S375-382. doi:10.3233/JAD-141470
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135-160. doi:10.1191/096228099673819272
- Braak, H., & Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging*, 16(3), 271-278. doi:10.1016/0197-4580(95)00021-6
- Braak, H., & Del Tredici, K. (2011). The pathological process underlying Alzheimer's disease in individuals under thirty. *Acta Neuropathologica*, 121(2), 171-181. doi:10.1007/s00401-010-0789-4
- Braak, H., & Del Tredici, K. (2015a). *Neuroanatomy and pathology of sporadic Alzheimer's Disease*: Springer.
- Braak, H., & Del Tredici, K. (2015b). The preclinical phase of the pathological process underlying sporadic Alzheimer's disease. *Brain*, 138(Pt 10), 2814-2833. doi:10.1093/brain/awv236
- Braak, H., Thal, D. R., Ghebremedhin, E., & Del Tredici, K. (2011). Stages of the pathologic process in Alzheimer disease: Age categories From 1 to 100 years. *Journal of Neuropathology & Experimental Neurology*, 70(11), 960. doi:10.1097/NEN.0b013e318232a379
- Calderon-Garciduenas, A. L., & Duyckaerts, C. (2017). Alzheimer disease. *Handbook of Clinical Neurology*, 145, 325-337. doi:10.1016/b978-0-12-802395-2.00023-7
- Cherbuin, N., Anstey, K. J., Reglade-Meslin, C., & Sachdev, P. S. (2009). In vivo hippocampal measurement and memory: A comparison of manual tracing and

- automated segmentation in a large community-based sample. *PLoS One*, 4(4), e5265.
doi:10.1371/journal.pone.0005265
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 7. doi:10.1037/1040-3590.6.4.284
- Clerx, L., Gronenschild, E. H. B. M., Echavarri, C., FransVerhey, Aalten, P., & Jacobs, H. I. L. (2015). Can FreeSurfer compete with manual volumetric measurements in Alzheimer's disease? *Current Alzheimer Research*, 12(4), 358-367.
doi:10.2174/1567205012666150324174813
- Clerx, L., van Rossum, I. A., Burns, L., Knol, D. L., Scheltens, P., Verhey, F., . . . Visser, P. J. (2013). Measurements of medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment. *Neurobiology of Aging*, 34(8), 2003-2013. doi:10.1016/j.neurobiolaging.2013.02.002
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.): L. Erlbaum Associates.
- Cook, M. J., Fish, D. R., Shorvon, S. D., Straughan, K., & Stevens, J. M. (1992). Hippocampal volumetric and morphometric studies in frontal and temporal and temporal-lobe epilepsy. *Brain*, 115, 1001-1015. doi:10.1093/brain/115.4.1001
- Cover, K. S., van Schijndel, R. A., Bosco, P., Damangir, S., Redolfi, A., & Alzheimer's Disease Neuroimaging, I. (2018). Can measuring hippocampal atrophy with a fully automatic method be substantially less noisy than manual segmentation over both 1 and 3 years? *Psychiatry Research: Neuroimaging*, 280, 39-47.
doi:10.1016/j.psychres.2018.06.011
- Cover, K. S., van Schijndel, R. A., Versteeg, A., Leung, K. K., Mulder, E. R., Jong, R. A., . . . Alzheimer's Disease Neuroimaging Initiative, n. (2016). Reproducibility of

- hippocampal atrophy rates measured with manual, FreeSurfer, AdaBoost, FSL/FIRST and the MAPS-HBSI methods in Alzheimer's disease. *Psychiatry Research: Neuroimaging*, 252, 26-35. doi:10.1016/j.pscychresns.2016.04.006
- Despotovic, I., Goossens, B., & Philips, W. (2015). MRI segmentation of the human brain: challenges, methods, and applications. *Computational and Mathematical Methods in Medicine*, 2015, 450341. doi:10.1155/2015/450341
- Doody, R. (2017). Developing disease-modifying treatments in Alzheimer's disease - A perspective from Roche and Genentech. *J Prev Alzheimers Dis*, 4(4), 264-272. doi:10.14283/jpad.2017.40
- Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., Dekosky, S. T., Barberger-Gateau, P., . . . Scheltens, P. (2010). Revising the definition of Alzheimer's disease: A new lexicon. *The Lancet Neurology*, 9(11), 1118-1127. doi:10.1016/s1474-4422(10)70223-4
- Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., . . . Washington Dc, U. S. A. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 12(3), 292-323. doi:10.1016/j.jalz.2016.02.002
- Fiest, K. M., Roberts, J. I., Maxwell, C. J., Hogan, D. B., Smith, E. E., Frolkis, A., . . . Jette, N. (2016). The prevalence and incidence of dementia due to Alzheimer's disease: A systematic review and meta-analysis. *The Canadian Journal of Neurological Sciences*, 43 Suppl 1, S51-82. doi:10.1017/cjn.2016.36
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341-355. doi:10.1016/s0896-6273(02)00569-x

- Fraser, M. A., Shaw, M. E., Anstey, K. J., & Cherbuin, N. (2018). Longitudinal assessment of hippocampal atrophy in midlife and early old age: Contrasting manual tracing and semi-automated segmentation (FreeSurfer). *Brain Topography*, 31(6), 949-962. doi:10.1007/s10548-018-0659-2
- Frisoni, G. B., Boccardi, M., Barkhof, F., Blennow, K., Cappa, S., Chiotis, K., . . . Winblad, B. (2017). Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *The Lancet Neurology*, 16(8), 661-676. doi:10.1016/S1474-4422(17)30159-X
- Habas, C. (2018). *The neuroimaging of brain diseases : Structural and functional advances*: Springer.
- Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., . . . Alzheimer's Disease Neuroimaging, I. (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. doi:10.1016/j.neuroimage.2015.04.042
- Jack, C. R., Jr. (2011). Alliance for aging research AD biomarkers work group: structural MRI. *Neurobiology of Aging*, 32 Suppl 1, S48-57. doi:10.1016/j.neurobiolaging.2011.09.011
- Jack, C. R., Jr. (2012). Alzheimer disease: new concepts on its neurobiology and the clinical role imaging will play. *Radiology*, 263(2), 344-361. doi:10.1148/radiol.12110433
- Jack, C. R., Jr., Albert, M. S., Knopman, D. S., McKhann, G. M., Sperling, R. A., Carrillo, M. C., . . . Phelps, C. H. (2011). Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 257-262. doi:10.1016/j.jalz.2011.03.004

- Jack, C. R., Jr., Barkhof, F., Bernstein, M. A., Cantillon, M., Cole, P. E., Decarli, C., . . . Foster, N. L. (2011). Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. *Alzheimer's & Dementia*, 7(4), 474-485 e474. doi:10.1016/j.jalz.2011.04.007
- Kantarci, K., & Jack, C. R., Jr. (2004). Quantitative magnetic resonance techniques as surrogate markers of Alzheimer's disease. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics*, 1(2), 196-205. doi:10.1602/neurorx.1.2.196
- Lane, C. A., Hardy, J., & Schott, J. M. (2018). Alzheimer's disease. *European Journal of Neurology*, 25(1), 59-70. doi:10.1111/ene.13439
- Lehmann, M., Douiri, A., Kim, L. G., Modat, M., Chan, D., Ourselin, S., . . . Fox, N. C. (2010). Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. *Neuroimage*, 49(3), 2264-2274. doi:10.1016/j.neuroimage.2009.10.056
- Mak, E., Gabel, S., Mirette, H., Su, L., Williams, G. B., Waldman, A., . . . O'Brien, J. (2017). Structural neuroimaging in preclinical dementia: From microstructural deficits and grey matter atrophy to macroscale connectomic changes. *Ageing Research Reviews*, 35, 250-264. doi:10.1016/j.arr.2016.10.001
- McEvoy, L. K., & Brewer, J. B. (2010). Quantitative structural MRI for early detection of Alzheimer's disease. *Expert Review of Neurotherapeutics*, 10(11), 1675-1688. doi:10.1586/ern.10.162
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to 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), 263-269. doi:10.1016/j.jalz.2011.03.005
- Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., . . . Rathi, Y. (2016). Inter-site and inter-scanner diffusion MRI data harmonization. *Neuroimage*, 135, 311-323. doi:10.1016/j.neuroimage.2016.04.041
- Miyagawa, T., & Iwatsubo, T. (2017). Toward preclinical trials for Alzheimer's disease. *Brain and Nerve*, 69(7), 711-722. doi:10.11477/mf.1416200812
- Morey, R. A., Petty, C. M., Xu, Y., Hayes, J. P., Wagner, H. R., II, Lewis, D. V., . . . McCarthy, G. (2009). A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage*, 45(3), 855-866. doi:10.1016/j.neuroimage.2008.12.033
- Morris, R., Paxinos, G., & Petrides, M. (2000). Architectonic analysis of the human retrosplenial cortex. *The Journal of comparative neurology*, 421(1), 14-28. doi:10.1002/(sici)1096-9861(20000522)421:1<14::aid-cne2>3.0.co;2-s
- Mueller, S. G., Schuff, N., Yaffe, K., Madison, C., Miller, B., & Weiner, M. W. (2010). Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease. *Human Brain Mapping*, 31(9), 1339-1347. doi:10.1002/hbm.20934
- Mulder, E. R., de Jong, R. A., Knol, D. L., van Schijndel, R. A., Cover, K. S., Visser, P. J., . . . Alzheimer's Disease Neuroimaging, I. (2014). Hippocampal volume change measurement: Quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST. *Neuroimage*, 92, 169-181. doi:10.1016/j.neuroimage.2014.01.058
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment - Clinical characterization and outcome. *Archives of Neurology*, 56(3), 303-308. doi:10.1001/archneur.56.3.303

- Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., . . . Frisoni, G. B. (2016). Brain atrophy in Alzheimer's disease and aging. *Ageing Research Reviews*, 30, 25-48. doi:10.1016/j.arr.2016.01.002
- Portney, L. G., & Watkins, M. P. (2013). *Foundations of clinical research: Applications to practice* (3rd ed.). Harlow: Pearson.
- Potvin, O., Mouiha, A., Dieumegarde, L., & Duchesne, S. (2016). FreeSurfer subcortical normative data. *Data Brief*, 9, 732-736. doi:10.1016/j.dib.2016.10.001
- Potvin, O., Mouiha, A., Dieumegarde, L., Duchesne, S., & Alzheimer's Disease Neuroimaging, I. (2016). Normative data for subcortical regional volumes over the lifetime of the adult human brain. *Neuroimage*, 137, 9-20. doi:10.1016/j.neuroimage.2016.05.016
- Ramos Bernardes da Silva Filho, S., Oliveira Barbosa, J. H., Rondinoni, C., Dos Santos, A. C., Garrido Salmon, C. E., da Costa Lima, N. K., . . . Moriguti, J. C. (2017). Neurodegeneration profile of Alzheimer's patients: A brain morphometry study. *NeuroImage: Clinical*, 15, 15-24. doi:10.1016/j.nicl.2017.04.001
- Reynolds, D. S. (2019). A short perspective on the long road to effective treatments for Alzheimer's disease. *British Journal Of Pharmacology*, 176(18), 3636-3648. doi:10.1111/bph.14581
- Sanchez-Benavides, G., Gomez-Anson, B., Sainz, A., Vives, Y., Delfino, M., & Pena-Casanova, J. (2010). Manual validation of FreeSurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer Disease subjects. *Psychiatry Research*, 181(3), 219-225. doi:10.1016/j.psychresns.2009.10.011
- Schmidt, M. F., Storrs, J. M., Freeman, K. B., Jack, C. R., Jr., Turner, S. T., Griswold, M. E., & Mosley, T. H., Jr. (2018). A comparison of manual tracing and FreeSurfer for

- estimating hippocampal volume over the adult lifespan. *Human Brain Mapping*, 39(6), 2500-2513. doi:10.1002/hbm.24017
- Schoemaker, D., Buss, C., Head, K., Sandman, C. A., Davis, E. P., Chakravarty, M. M., . . . Pruessner, J. C. (2016). Hippocampus and amygdala volumes from magnetic resonance images in children: Assessing accuracy of FreeSurfer and FSL against manual segmentation. *Neuroimage*, 129, 1-14. doi:10.1016/j.neuroimage.2016.01.038
- Schroder, J., & Pantel, J. (2016). Neuroimaging of hippocampal atrophy in early recognition of Alzheimer's disease - A critical appraisal after two decades of research. *Psychiatry Research: Neuroimaging*, 247, 71-78. doi:10.1016/j.psychresns.2015.08.014
- Serrano-Pozo, A., Frosch, M. P., Masliah, E., & Hyman, B. T. (2011). Neuropathological alterations in Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, 1(1), a006189. doi:10.1101/cshperspect.a006189
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations - uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428. doi:10.1037/0033-2909.86.2.420
- Sperling, R., & Johnson, K. (2013). Biomarkers of Alzheimer disease: Current and future applications to diagnostic criteria. *Continuum*, 19(2 Dementia), 325-338. doi:10.1212/01.CON.0000429181.60095.99
- 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), 569-581. doi:10.1007/s00234-008-0383-9
- Tan, C. C., Yu, J. T., & Tan, L. (2014). Biomarkers for preclinical Alzheimer's disease. *Journal of Alzheimer's Disease*, 42(4), 1051-1069. doi:10.3233/JAD-140843
- Teipel, S. J., Ewers, M., Wolf, S., Jessen, F., Kolsch, H., Arlt, S., . . . Hampel, H. (2010). Multicentre variability of MRI-based medial temporal lobe volumetry in Alzheimer's

- disease. *Psychiatry Research*, 182(3), 244-250.
doi:10.1016/j.psychresns.2010.03.003
- Teipel, S. J., Kurth, J., Krause, B., & Grothe, M. J. (2015). The relative importance of imaging markers for the prediction of Alzheimer's disease dementia in mild cognitive impairment — Beyond classical regression. *NeuroImage: Clinical*, 8, 583-593.
doi:doi.org/10.1016/j.nicl.2015.05.006
- Trzepacz, P. T., Yu, P., Sun, J., Schuh, K., Case, M., Witte, M. M., . . . Alzheimer's Disease Neuroimaging, I. (2014). Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to Alzheimer's dementia. *Neurobiology of Aging*, 35(1), 143-151. doi:10.1016/j.neurobiolaging.2013.06.018
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., . . . Alzheimer's Disease Neuroimaging, I. (2015). 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's & Dementia*, 11(6), e1-120. doi:10.1016/j.jalz.2014.11.001
- Wenger, E., Martensson, J., Noack, H., Bodammer, N. C., Kuehn, S., Schaefer, S., . . . Lovden, M. (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. doi:10.1002/hbm.22473
- Wisse, L. E. M., Daugherty, A. M., Olsen, R. K., Berron, D., Carr, V. A., Stark, C. E. L., . . . Hippocampal Subfields, G. (2017). A harmonized segmentation protocol for hippocampal and parahippocampal subregions: Why do we need one and what are the key goals? *Hippocampus*, 27(1), 3-11. doi:10.1002/hipo.22671
- Yushkevich, P. A., Amaral, R. S., Augustinack, J. C., Bender, A. R., Bernstein, J. D., Boccardi, M., . . . Hippocampal Subfields, G. (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.

doi:10.1016/j.neuroimage.2015.01.004

Chapter 6

Medial Temporal Volumetric Correlates to Memory Performance and Diagnosis of Mild Cognitive Impairment and Alzheimer's Disease

Through Chapters 1 to 5, repeated emphasis had been placed on how disproportionate structural atrophy observed in the medial temporal structures in the early phase of symptomatic Alzheimer's disease would make these structures good neuroimaging biomarkers for Alzheimer's disease detection. Using the data obtained from a subset of the overall larger sample pool, this chapter addressed the discriminatory ability of these different structures between different clinical samples and healthy older adults, hypothesising that these structural volumes would be able to differentiate between the different groups. This chapter also explored the relationship between the medial temporal structural volumes and memory function.

Abstract

Changes in medial temporal structures, specifically the hippocampus, entorhinal cortex and transentorhinal cortex, can potentially serve morphologically as surrogate biomarkers for the identification of Alzheimer's disease (AD). The relationship between the volumetric measures of these three structures were investigated, estimated using both manual and automated (FreeSurfer) methods, and memory performance. The diagnostic discriminant ability volumetric measures could add over comprehensive memory assessment was also investigated. The sample consisted of 100 participants including 14 with AD, 13 with mild cognitive impairment, 33 with subjective memory concerns and 40 healthy controls without memory complaints. All subjects underwent T1-weighted 1.5T or 3T MRI and a comprehensive memory assessment using the Weschler Memory Scale III/IV. The left hippocampal volume was the best predictor of verbal and overall memory performance, while the right entorhinal volume was the best predictor of visual memory performance. Although the left and right hippocampal volumes improved overall prediction of diagnosis, they were unable to provide additional discriminatory ability above that already provided by comprehensive cognitive testing. There was evidence of lateralisation and asymmetry effects between the left and right hippocampus in the different diagnostic groups, but these differences had little discriminant ability. Both manual and automated estimations produced similar findings. The results suggest that the diagnostic utility of visual memory should be further explored, and that given the significant role the medial temporal structures, further investigations into the efficacy of using a combination of other biomarkers on top of volumetric measures to improve diagnostic classification should be considered.

Keywords: Alzheimer's disease, mild cognitive impairment, medial temporal lobe, hippocampus, entorhinal, transentorhinal, memory, volume, FreeSurfer

Medial Temporal Volumetric Correlates to Memory Performance and Diagnosis of Mild Cognitive Impairment and Alzheimer's Disease

In the context of progressively ageing populations across many countries, the global number of people living with dementia is expected to approximately double every twenty years (Alzheimer's Association, 2018). The diagnosis of dementia is an important and sometimes complex process and in Australia, it may take approximately three years from onset of symptoms to receive a firm diagnosis of dementia (Nair, Mansfield, & Waller, 2016).

Alzheimer's disease (AD) is the most common subtype of dementia, and as specific disease modifying agents are developed for this and other conditions, it will become increasingly important to accurately identify the disease in its early stage, typically as mild cognitive impairment (MCI). The diagnostic guideline for the clinical stages of AD was laid out by the National Institute of Neurological and Communicative Disorder and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) in 2011 (McKhann et al., 2011). The clinical criteria for AD include an insidious onset and decline in cognitive function that interferes with daily functioning, in which this cognitive impairment is not explained by other medical conditions. The cognitive impairment is established through history taking and objective cognitive assessment, and cognitive impairment should impact two or more domains, and those domains could include or not include memory dysfunction (McKhann et al., 2011).

The term MCI was introduced as a clinical construct to capture the earliest symptomatic features of dementia, including AD (Petersen et al., 2009). The main distinguishing clinical criteria of MCI from dementia is independence of daily functions despite the presence of cognitive impairments (Albert et al., 2011; Petersen et al., 1999; Winblad et al., 2004). Early clinical definition of MCI was focussed on MCI as a condition

preceding progression into AD, but later refinements of the MCI condition noted that not all people with MCI evolve into AD (Petersen et al., 1999; Winblad et al., 2004). Later iterations of the MCI clinical diagnosis provided further classifications of MCI, in which MCI could present with amnesic and non-amnesic cognitive impairments, and the impairment could occur in single or multiple cognitive domains (Winblad et al., 2004). In the newest clinical criteria focussed on MCI due to AD, which are detailed in Albert et al. (2011), although the authors highlighted that memory dysfunction is the most common cognitive impairment in MCI due to AD, there was no discussion on the subtypes of MCI. Instead, Albert et al. (2011) emphasised the necessity to rule out any conditions or underlying aetiologies that could be the primary cause for the cognitive impairment seen in MCI. The process of ruling out other causes in turns increases the likelihood that the aetiology of cognitive impairment observed in an MCI diagnosis is due AD (Albert et al., 2011).

The new 2011 clinical criteria for MCI and AD remain largely similar to the previous published criteria (Albert et al., 2011; Petersen et al., 1999; Winblad et al., 2004). One of the main notable difference is the inclusion of anatomical, physiological and biochemical biomarkers for diagnosis to reflect the pathophysiological features of the underlying AD aetiology (Jack, Albert, et al., 2011). However, these biomarkers serve a complementary role while the fulfilment of the clinical criteria remain paramount for a clinical diagnosis of MCI and AD (Jack, Albert, et al., 2011).

Although the use of these biomarkers remain conservative, studies that have been published since the 2011 clinical criteria continue to reinforce the use of these biomarkers as proxies for AD-related pathological changes and will continue to provide a framework for AD characterisation, particularly in the initial phase of the AD development (Jack et al., 2018). The incorporation of biomarkers aids in an early accurate diagnosis of AD and MCI, in which timely detection optimises symptomatic intervention and planning of care prior to

further devastating neurodegeneration (Sperling et al., 2011). Furthermore, when pharmaceutical interventions become available, they are likely to be most effective during the early clinical stages of the disease (Sperling et al., 2011).

Medial Temporal Lobe and Alzheimer's Disease

The medial temporal lobe (MTL) is made up of a set of highly interconnected brain regions that are crucial for the successful formation of short- and long-term declarative memory. The MTL is mainly comprised by the hippocampus, entorhinal cortex and its adjacent perirhinal cortex (Squire & Zola-Morgan, 1991). The MTL is reputed to be one of the earliest vulnerable sites afflicted by AD pathology and this pathological process clinically manifests as insidious short-term memory impairment early in the disease. One of the major pathological hallmarks of AD is intracellular neurofibrillary tangles (NFTs) that are hyperphosphorylated and misfolded aggregates of abnormal tau protein. Proliferation of NFTs follows a highly stereotypical spatiotemporal pattern of progression, first appearing within specific MTL structures before extending into the neocortex (Braak et al., 2006; Braak & Braak, 1991, 1995; Braak et al., 1993). As the deposition of abnormal AD-related tau protein progresses in a highly predictable manner, it allows for the topographical staging of pathological NFTs from stage I to VI (Braak & Braak, 1991).

In the earliest NFT stage I, the bulk of these abnormal hyperphosphorylated tau protein is detectable in the perirhinal cortex, with little to no involvement of the entorhinal cortex. More specifically, the abnormal tau proteins are found in the medial portion of the perirhinal cortex, which Braak and Braak (1985) coined the “transentorhinal cortex”. This term describes its transitional role between the adjacent isocortex and the entorhinal cortex, in which majority of the inputs from the neocortex would converge onto the transentorhinal cortex before being guided through to the entorhinal cortex (Braak & Del Tredici, 2015a).

In NFT stage II, proceeding from the transentorhinal cortex, the NFTs encroach onto the entorhinal cortex (Braak et al., 2006; Braak & Braak, 1991). Together with the transentorhinal cortex, the entorhinal cortex serves as a neuronal bridge between the hippocampus and the neocortex. The entorhinal cortex provides majority of the input to the hippocampus via the perforant pathway, following which, the output from the hippocampus is back-projected into the entorhinal cortex before being transferred into the neocortex (Braak & Del Tredici, 2015a; Lavenex & Amaral, 2000).

In NFT stage III, the abnormal hyperphosphorylated tau proteins becomes more severe in the transentorhinal and entorhinal cortices and extends into the neighbouring structures. This proliferation of the NFTs includes invasion into the adjoining temporal neocortex as well as significant involvement of the hippocampus (Braak et al., 2006; Braak & Braak, 1991). While the abnormal tau pathology continues its unrelenting progression into the rest of the neocortex during NFT stages IV – VI, the projection cells within the transentorhinal cortex, entorhinal cortices and hippocampus become increasingly destroyed.

The extent of the pervasiveness of the NFTs is correlated to the classical early symptoms of memory inefficiencies observed in people with AD (Braak & Del Tredici, 2015b; Hyman, Van Hoesen, & Damasio, 1990). Typically, patients in NFT stage I – II are cognitively asymptomatic (Braak & Del Tredici, 2015a, 2015b). However, cognitive symptoms are increasingly likely to begin manifesting from NFT stage III onwards (Braak & Del Tredici, 2015a, 2015b). The unremitting invasion of the NFTs increasingly disrupts the neural connectivity between the transentorhinal cortex, entorhinal cortex and hippocampus, which would eventually lead to a disconnection between the hippocampus and the neocortex (Braak, Braak, Yilmazer, & Bohl, 1996; Delbeuck, Van der Linden, & Collette, 2003).

Medial Temporal Lobe Structures and Memory Function

The early foundations of neuropsychological inference that were based on studies of lesion localisation demonstrated that different regions of the MTL could be differentially involved in memory formation. One such difference was the material-specific lateralisation of function, with lesional studies demonstrating a degree of dissociation between the left and right hemispheres with verbal and non-verbal memory function. Resection of the left MTL in patients with intractable epilepsy resulted in impaired verbal memory but relatively intact non-verbal memory, with the inverse being true in patients who received right MTL resections (Gleißner, Helmstaedter, & Elger, 1998; Jones-Gotman, 1986; Jones-Gotman et al., 1997; Squire & Butters, 1992). However, the same studies also noted that patients with right MTL resections had a less profound material-specific memory deficit than that seen in patients with left MTL resections (Gleißner et al., 1998; Jones-Gotman, 1986; Jones-Gotman et al., 1997; Squire & Butters, 1992).

Non-invasive neuroimaging studies have also observed similar patterns of material-specific patterns of lateralisation. Functional studies using magnetic resonance imaging have shown preferential involvement of mainly the left MTL structures, particularly the left hippocampus, in the processing of verbal stimuli (Dalton, Hornberger, & Piguet, 2016; Rosazza et al., 2009). On the other hand, non-verbal and spatial processing did not appear to be strictly lateralised or localised to the right hippocampus. Studies have seen more widespread activations in the right MTL, including the right parahippocampal and perirhinal regions, with a tendency for concurrent left hippocampal and parahippocampal activation during more challenging visual tasks (Baumann, Chan, & Mattingley, 2010; Dalton et al., 2016; Hartley, Maguire, Spiers, & Burgess, 2003; Maguire et al., 1998). Some volumetric studies have also seen evidence of a left-right material-specific lateralisation of hippocampal function although some researchers found results that pointed towards a lack of material-

specific lateralisation (Bonner-Jackson et al., 2015; K. H. Chen, Chuah, Sim, & Chee, 2010; Ezzati et al., 2016; Vyhnaček et al., 2014).

Observations of MTL atrophy in neurological conditions associated with memory deficits, such as amnesic MCI diagnosed on the basis of likely underlying neuropathological AD and dementia of the AD type, allows one to postulate a direct correlation between MTL structural volumes and memory function (Bottino et al., 2002; Jack et al., 1997; Mizuno, Wakai, Takeda, & Sobue, 2000). Numerous studies have demonstrated a direct association of MTL atrophy with decline in memory function, in particular atrophy of the hippocampus and entorhinal cortex (see Weiner et al., 2015 for a selective review).

Unfortunately, the relationship between hippocampal size and memory ability may not be as straightforward as much of the literature may suggest. A meta-analysis by Van Petten (2004) showed surprisingly weak empirical support for a direct relationship between hippocampal volume and declarative memory in older adults without apparent cognitive impairment, which was likely an effect of the large inter-individual variability in both hippocampal volume and memory abilities in this older population. Perhaps even more disconcertingly, after adjusting for age-related decline, Van Petten (2004) did not find any relationship between hippocampal volume and memory performance. The finding by Van Petten (2004) was echoed in a study by Apostolova et al. (2010) who reported that there was no association between memory and hippocampal volumes in cognitively normal older adults; however, the authors did find associations in the sample with MCI and AD, suggesting that the relationship between hippocampal volume and memory function may be driven by memory consolidation strategies that are negatively impacted by AD pathology. Similar effects of diagnosis-dependent association between memory and entorhinal and perirhinal cortex was found in another study, in which memory performance correlated with

both the entorhinal and perirhinal cortical thickness measures only in the AD group but not in the healthy controls (Dickerson et al., 2009).

The lack of direct correlation between volume and memory could be associated with the observation that while the initiation of tau protein pathology is associated with neuronal loss, a lot of the neurons continue to survive and function, albeit at a more restricted manner (Stokin & Goldstein, 2006; Wirths & Bayer, 2010). Cognitive function is therefore not determined independently by the number of neurons present, but also impacted by the number of defective neurons (Braak & Del Tredici, 2015b). When the increasingly massive number of surviving but defective neurons surpasses a given threshold in the presence of pathology, cognitive dysfunction becomes apparent (Braak & Del Tredici, 2015b). In addition, functional studies have shown that compared to younger adults, older adults had greater bilateral neural connectivity throughout the brain during demanding cognitive tasks, which had been proposed to serve as a compensatory mechanism for deficits within the MTL (Deng et al., 2020; Dolcos, Rice, & Cabeza, 2002).

Role of Neuroimaging in the Diagnosis of Alzheimer's Disease

Undoubtedly, the development of various neuroimaging acquisition and analysis techniques represent some of the greatest advances in detecting structural and pathological brain changes *in vivo*. The role of neuroimaging targeted at detecting pathophysiological mechanisms specific to AD has been evolving over the years. For example, the use of positron emission tomography (PET) to measure the accumulation and deposition of amyloid-beta in amyloid plaques, amyloid-beta being another molecular pathological hallmark of AD, is now commonplace (X. Q. Chen & Mobley, 2019; Chi et al., 2018). However, amyloid burden per se is not a good predictor of cognitive performance and does not correlate well with the severity or duration of dementia. (Giannakopoulos et al., 2003; Wirth et al., 2013).

Contrary to the lack of correspondence between amyloid plaques and cognitive decline, the relationship between density and distribution of NFTs with severity of cognitive impairment is well-established (Braak & Braak, 1997; Bussiere et al., 2003; Giannakopoulos et al., 2003; Maass et al., 2017; Maass et al., 2018; Mortimer, Gosche, Riley, Markesbery, & Snowden, 2004; Nelson et al., 2012; Theofilas et al., 2018). The formation of NFTs have been found to correlate highly with synapse loss, and both the density of NFTs and degree of synapse loss are associated with cognitive decline in the course of AD (Gomez-Isla et al., 1997). Although several tau PET studies have illustrated a clear relationship between patterns of tau distribution in the MTL with memory decline (Cho et al., 2016; Wang et al., 2016), tau PET tracers are still in development and the reliability of tau PET is still under investigation (Leuzy et al., 2019; Márquez & Yassa, 2019).

Currently, the mechanistic relationship between NFTs and neuronal cell death remains elusive, but several studies have found associations between severity of NFT lesions with neuronal loss and the degree of neuronal loss could be approximated using volumetric measures (Atiya et al., 2003; Gomez-Isla et al., 1996; Iqbal, Liu, Gong, Alonso Adel, & Grundke-Iqbal, 2009; Theofilas et al., 2018). Studies focused on people with probable AD pathology have repeatedly demonstrated early disproportionate MTL atrophy in a topographical pattern consistent with the deposition of AD-related abnormal tau protein, which eventually led to the identification of MTL atrophy as a useful tool in the early diagnosis of AD (Duara et al., 2015; Duara et al., 2008; Jhoo et al., 2010; Min et al., 2017; Visser, Verhey, Hofman, Scheltens, & Jolles, 2002).

Hippocampal atrophy has been a consistent finding in people with AD and MCI on a neuropathological basis of AD (Hyman, Van Hoesen, Damasio, & Barnes, 1984; Maass et al., 2018; Stoub et al., 2010; Van Hoesen, Hyman, & Damasio, 1991). In addition to hippocampal atrophy, entorhinal and transentorhinal volumes were also reduced in people at

baseline in the earlier stages of the disease and were shown to be good predictors of future cognitive decline (Burggren et al., 2011; Kulason et al., 2018; Pennanen et al., 2004; Reitz et al., 2009; Stoub et al., 2010; Tapiola et al., 2008; Tward et al., 2017; Velayudhan et al., 2013; Venneri et al., 2011; Xie et al., 2018). Another group of interest are patients with subjective memory decline without objective impairments, in which a meta-analysis suggested that this group of patients might be at risk of future degeneration (Mitchell, Beaumont, Ferguson, Yadegarfar, & Stubbs, 2014). Several studies have noted atrophic changes in the hippocampus and entorhinal cortex in this group, which might point towards early signs of AD pathology (Fan et al., 2018; Perrotin et al., 2017; Yue et al., 2018; Zhao et al., 2019).

Compared to other modalities such as assessing for amyloid and tau burden using PET, one of the advantages of using volumetric measures is the ease of access to magnetic resonance imaging (MRI). Structural MRI is commonly included as part of a routine diagnostic workup in patients with cognition concerns. Given the extensive documentation of MTL atrophy that has been shown to follow closely the clinical phases of AD progression, the benefit of incorporating structural biomarkers to advance AD diagnosis appears unequivocal. As such, MTL atrophy has been included as supporting pathophysiological evidence in the National Institute on Ageing as part of the clinical research diagnostic criteria for neurodegeneration associated with probable AD dementia and MCI likely due to AD (Albert et al., 2011; McKhann et al., 2011).

Manual delineation of regions of interest, although tedious and time-consuming, remains the “gold standard” for quantitative volumetric assessment of specific structures (Good et al., 2002; Wilke et al., 2011). However, major advances in structural volumetric analysis has come with the refinements of automated and semi-automated analysis software, one of the most popular being open-source neuroimaging analysis software, FreeSurfer (<http://surfer.nmr.mgh.harvard.edu>). FreeSurfer provides an array of automated neuroimaging

analysis algorithms that can be used to quantify several properties of the human brain, including structural measures like volume and cortical thickness (Fischl, 2012). The automation of neuroimaging analysis provides the potential for standardised analysis on large datasets to generate normative neuroanatomical and neuropathological data (Bigler, 2015). The ease of automated analysis, coupled with the possibility of generating normative data, FreeSurfer presents a great opportunity for the integration of volumetric analysis into the routine diagnostic process (Bigler, 2015).

Lateralisation and Asymmetry of Medial Temporal Structures in Alzheimer's Disease

The examination of lateralisation and asymmetrical hippocampal atrophy between people with or without AD has gained immense interest in recent years. Asymmetrical hippocampal volumes, typically right larger than left, has been found in healthy adults (Fraser et al., 2018; Guadalupe et al., 2017; Schmidt et al., 2018). Some studies have noticed that this asymmetry decreases with severity of AD, suggesting a disproportionate hippocampal volume loss lateralised to the right in people with AD (Ardekani, Hadid, Blessing, & Bachman, 2019; Yue et al., 2018). In contrast, other studies have found the degree of this typical left-right asymmetry increased with AD severity, suggesting a left-lateralised atrophy of the hippocampus (Muller et al., 2005; Sarica et al., 2018; Thompson et al., 2003).

A meta-analysis by Shi, Liu, Zhou, Yu, and Jiang (2009) looked at AD-related hippocampal volumetric studies between the years 1998 – 2007 that were mostly based on manual delineation. Unsurprisingly, they found significant bilateral hippocampal atrophy in MCI and AD and a consistent left-less-than-right asymmetry in all groups - healthy controls, MCI and AD. However, this asymmetry was most significant in the MCI group, with a left hippocampus that is 9.1% smaller than the right, compared to 6.3% in the AD group and 5.8% in the control. In contrast, a recent meta-analysis study by Minkova et al. (2017) reviewed hippocampal asymmetry based on voxel-base morphometry. They found that

studies reported a right-lateralised atrophy in the MCI group. However, this pattern of asymmetry was reversed in people with AD. Although this reversal in atrophic lateralisation might suggest AD-stage specific atrophic effects, the authors highlighted that different MCI and AD diagnostic criteria were adopted across the selected studies, which would have increased the variability in the data (Minkova et al., 2017).

Some studies have also explored asymmetries in the entorhinal cortex in individuals with AD pathology. Comparing the asymmetry of entorhinal cortical thickness between healthy controls and people with AD, Donix et al. (2013) found that healthy control participants had a non-significant thinner left entorhinal cortex than the right (5%), while the degree of asymmetry increased with a significantly thinner left entorhinal cortex than the right in people with AD (10%). In contrast, Long, Jiang, and Zhang (2018) reported a trend of towards zero in the asymmetry index between the left and right entorhinal cortex in people who converted from MCI to AD compared to those who did not. Information on the transentorhinal cortex is unfortunately quite sparse, likely due to difficulties in demarcating its anatomical location. One study that specifically investigated the transentorhinal cortex had noted that there was little asymmetry in the healthy controls but a strong left-sided biased atrophy in the MCI participants (Yushkevich, Pluta, et al., 2015).

The neurobiological basis of lateralisation and asymmetry remains unclear. One study found that rather than a dominant direction of hippocampal atrophy in one of the hemispheres, atrophy could occur on either of the hemispheres and it is the magnitude of asymmetry between the two hemispheres that was associated with AD progression (Wachinger, Salat, Weiner, Reuter, & Alzheimer's Disease Neuroimaging, 2016). Derflinger et al. (2011) argued that brain atrophy was asymmetrical rather than lateralised, and that the observed left-lateralised atrophy likely arose from a selection bias. As memory assessments are commonly language-based, those language-based tests are more likely to detect memory

impairments in patients with left-lateralised atrophy and select those patients with verbal memory difficulties into the studies. Consequently, verbal memory assessments would miss out on people with right-lateralised atrophy who are more likely perform more poorly on visuospatial memory tasks (Derflinger et al., 2011).

Furthermore, histopathological studies have suggested that although there might be some lateralisation in tau-related pathology, there were no evidence of preferential hemispheric vulnerability (Arnold, Hyman, Flory, Damasio, & Van Hoesen, 1991; Braak & Del Tredici, 2015a; Janota & Mountjoy, 1988; Moossy et al., 1989; Stefanits, Budka, & Kovacs, 2012; Wilcock & Esiri, 1987). Considering the inconsistent hippocampal asymmetry findings, limited breadth of research focused on the asymmetries in the entorhinal and transentorhinal cortices, and histopathological evidence that suggest the lack of hemispheric vulnerability, it is questionable whether a left-right asymmetry would serve as a reliable indicator of disease.

Neuropsychological Assessment of Memory Function

One of the common pitfalls of memory research is that the quality of findings is often undermined by short memory tests or the use of only one subtest from longer memory batteries. Abbreviated testing produces poorer signal-to-noise ratios, introduces considerable variability to test scores and therefore reduces sensitivity to detecting cognitive changes. For example, Carstairs, Shores, and Myers (2012) investigated the reliability of one of the most widely-used verbal memory tests, the Rey Auditory Verbal Learning Test, which is a task that requires participants to remember a list of 15 words over several trials. They administered the word-list learning task in an Australian sample of 390 participants aged between 18 – 34 years and found poor test-retest reliability ranging from .26 to .64. A popular non-verbal memory test, the Rey Complex Figure Test, which requires the participant to copy a complex diagram and then draw that diagram from memory after a period of delay,

was also found to have very poor reliability. One study reported poor reliabilities of .47 to .59 in the immediate recall and 30-minute delayed-recall trials respectively and another study reported somewhat better but less than ideal reliabilities of .57 to .77 for the three-minute delayed recall trial (Strauss, Sherman, & Spreen, 2006).

The Weschler Memory Scale (WMS) is a battery of subtests developed by David Wechsler in 1945 to serve as a clinical instrument to measure various domains of memory function. Since its inception, it had undergone rigorous revisions and psychometric development. The first revision, WMS-R, came in 1987 but was accompanied by several norming and standardisation issues through use of a small sample size consisting of only 316 people, having extrapolated scores in certain age groups and having a lack of norms for people above the age of 75 years (Kent, 2013). To address the shortcomings of WMS-R, the next revision, WMS-III, arrived on the market just ten years after WMS-R. WMS-III expanded its normative sample size to 1250 individuals, included age norms from 16 to 89 years of age, and the reported reliability coefficients for the main index scores ranged from .74 - .93 (Kreutzer, DeLuca, & Caplan, 2018). The most recent revision, WMS-IV (2009), further improved its standardisation process over its predecessor, developing norms on a slightly larger sample size of 1400 individuals. The reported reliability coefficients for the main memory index scores, which included auditory, visual, immediate and delayed memory indices, were excellent, ranging from .91 to .97 (Wechsler, 2009).

Interestingly, despite advocating the potential clinical utility of immediate and delayed memory indices in the WMS, attempts to separate immediate and delayed memory functioning in the same confirmatory factor analysis had consistently resulted in errors because of the high correlation between the immediate and delayed measures (Price, Tulskey, Millis, & Weiss, 2002; Tulskey & Price, 2003). However, confirmatory factor analyses have constantly found only three factors in the WMS-III and WMS-IV. These factors were (1)

auditory memory (immediate and delayed), (2) visual memory (immediate and delayed), and (3) working memory (Holdnack, Xiaobin, Larrabee, Millis, & Salthouse, 2011; Miller, Davidson, Schindler, & Messier, 2013; Price et al., 2002). Given these findings, focus should be placed on using both the WMS immediate and delayed indices as these indices are measuring the same underlying single memory construct, while also capturing measures from both the auditory and visual memory subtests.

Aims of Current Study

To overcome the limitations of abbreviated testing, the memory index scores obtained on the WMS were used as measures of auditory, visual and general memory functioning. To increase the patient sample size, measures from both WMS-III and WMS-IV were included. Results interpretation is unlikely to be impacted by the different versions of WMS, as despite the major changes made to content and structure from WMS-III to WMS-IV, similar constructs were measured in both versions (Weschler, 2009). Furthermore, the effect sizes of the mean differences between the relevant index scores, namely the auditory, visual, immediate and delayed indices found in the WMS-III and WMS-IV, were all negligible, with measures of effect size ranging from .01 - .15 (Weschler, 2009).

Although manual delineation of brain structures remains the gold standard, the constant evolution of automated tools to measure MRI brain structural volumes suggests that automated volume estimation may now have the accuracy of manual estimation. Therefore, both manual and automated methods were utilised for volume estimations of the putative MTL structures acknowledged to display AD vulnerability, namely the hippocampus, entorhinal cortex and transentorhinal cortex.

There are several objectives to the present study: (1) examine the volumetric relationship of MTL structures with memory function, (2) investigate the additional value

that quantitative MTL measures could bring to differential diagnosis, and (3) explore the role of laterality and asymmetry in different diagnostic samples.

Method

Study Population

Clinical sample. The clinical sample ($n = 60$) was drawn retrospectively from an archival database of patients who were referred to the memory clinic in St Vincent's Hospital Melbourne between 2006 – 2018. All patients received a routine neurological assessment by a neurologist at the memory clinic, underwent a brain MRI scan either on a 1.5T or 3T scanner and a comprehensive neuropsychological assessment that included the full administration of the complete WAIS-IV/III and WMS-III/IV. Cognitive assessment took place within 15 months of the MRI scans (mean = 2 months after the MRI assessment, $SD = 5$).

Although the criteria for MCI and AD have been updated across the years of retrospective patient recruitment, with the inclusion of relevant CSF and neuroimaging biomarkers as supportive evidence for AD in the latest research criteria, the clinical diagnostic criteria have remained similar. In the current study, patients with AD were classified based on clinical diagnosis guidelines that were basically the same between the original 1984 and revised 2011 NINCDS-ADRDA guidelines (McKhann et al., 1984; McKhann et al., 2011). The diagnosis of AD requires a loss of independence in daily function that is a result of disproportionate objective cognitive impairment of more than 1.5 standard deviations in two or more domains, and the impairment occurs in the absence of other medical conditions that could account for the cognitive impairment.

Similarly, in the current study, patients with MCI were classified based on clinical diagnostic criteria that were almost identical between the 1999 Petersen's criteria and 2011 MCI guideline proposed by NINCDS-ADRDA (Albert et al., 2011; Petersen et al., 1999).

Effectively, the diagnosis of MCI is distinguished from dementia based on essentially normal functional capacity in activities of daily life despite the presence of disproportionate objective cognitive impairment of more than 1.5 standard deviations in one or more domains. In accordance to the newest MCI clinical criteria detailed in Albert et al. (2011), the present study did not distinguish between the different subtypes of MCI. However, as recommended by the newest clinical criteria, patients with MCI diagnoses that were likely due to a non-AD pathology were excluded in this study. In addition, the clinical diagnostic criteria for MCI and AD are also in line respectively with the current diagnostic guidelines for mild neurocognitive disorder and major neurocognitive disorder that are laid out in the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013).

All patients included in this study were diagnosed as having MCI ($n = 13$) or AD ($n = 14$) according to recommended diagnostic criteria (Albert et al., 2011; McKhann et al., 1984; McKhann et al., 2011; Petersen et al., 1999). Majority of the patients diagnosed with MCI had an amnesic presentation, with only one patient with a non-amnesic presentation. In line with the 2011 MCI clinical diagnostic criteria (Albert et al., 2011), memory dysfunction can be absent in MCI due to AD and there were no other conditions or underlying aetiologies to explain the cognitive impairment, and the patient was therefore included in this study.

All participants who did not meet the diagnostic criteria, were considered to have subjective memory concerns (SMC; $n = 33$). Patients with a prior or current history of mild-to-moderate anxiety or depression were retained given the “real world” naturally increased prevalence of mood-related difficulties in this patient group. General exclusionary criteria for the clinical sample were age below 50 years and a history of an acquired head injury with persisting cognitive deficits. Patient demographics and cognitive characteristics could be found in Table 6.1.

Control sample. A group of healthy control participants (HC; $n = 40$) was recruited for this study between the years 2017 and 2018. All control participants underwent a brain MRI scan on a 3T scanner and a comprehensive neuropsychological assessment that included the complete administration of the WAIS-IV and WMS-IV.

General exclusionary criteria for the control sample were being below 50 years of age, having contraindications to undertaking an MRI, having significant current medical or neurological illness that would have impact on cognition, and having a history of an acquired head injury with persisting cognitive deficits. All control participants included in this study had no unexpected or unusual findings on their cognitive assessments and MRI scans. Control participants with a prior or current history of mild-to-moderate anxiety or depression were retained. Demographic and cognitive characteristics of the HC sample can be found in Table 6.1. This study was approved by the St. Vincent's Hospital (Melbourne) Human Research Ethics Committee (HREC-A 057/15).

Neuropsychometric Measures

Memory measures. The current study used three standard memory scores from either WMS-III or WMS-IV, depending on the WMS version administered. *Auditory memory* was either the average of the auditory immediate index and auditory delayed index in the WMS-III, or the auditory memory index in the WMS-IV. *Visual memory* was either the average of the visual immediate index and visual delayed index in the WMS-III, or the visual memory index in the WMS-IV. *General memory* was either the average of the immediate memory index and general memory index in the WMS-III, or the average of the immediate memory index and delayed memory index in the WMS-IV.

Brain MRI Acquisition Parameters

All T1-weighted scans were acquired coronally along the hippocampal axis with 3D-sequenced MPRAGE. The 1.5T scans were acquired on a Siemens Symphony scanner with

Table 6.1*Demographics and Neuropsychological Characteristics of the Clinical and Control Samples*

	AD (n = 14)	MCI (n = 13)	SMC (n = 33)	HC (n = 40)	ANOVA or Chi-Square			
					<i>F</i> or χ^2	<i>p</i> -value	Effect size ^b	Post-hoc
Age	70.3 ± 8.7	70.9 ± 8.3	65.6 ± 9.5	67.7 ± 5.4	2.06	.11	0.17	n/a
Female	21.4%	30.8%	60.6%	75.0%	16.34	.001	0.40	AD < expected; HC > expected
Education (years)	12.3 ± 4.1	12.8 ± 3.2	13.1 ± 4.2	16.0 ± 2.7	6.48	<.001	0.06	HC > AD, MCI, SMC
3T scanner ^a	42.9%	64.5%	48.5%	100%	29.77	<.001	0.55	SMC < expected; HC > expected
WAIS-IV/WMS-IV ^a	71.4%	69.2%	75.8%	100%	13.04	.005	0.36	HC > expected
General Memory	66 ± 13.2	81 ± 13.3	100 ± 16.4	116 ± 10.8	55.36	<.001	0.63	AD < MCI, SMC, HC; MCI < SMC, HC; SMC < HC
Auditory Memory	65 ± 14.8	81 ± 15.7	101 ± 17.3	117 ± 12.4	49.30	<.001	0.61	AD < MCI, SMC, HC; MCI < SMC, HC; SMC < HC
Visual Memory	72 ± 13.0	86 ± 13.0	98 ± 13.9	110 ± 9.3	38.18	<.001	0.37	AD < MCI, SMC, HC; MCI < SMC, HC; SMC < HC
VCI ^c	88 ± 17.3	101 ± 12.6	109 ± 16.5	119 ± 12.1	15.35	<.001	0.36	AD < SMC, HC; MCI < HC
PRI ^c	82 ± 20.3	96 ± 15.2	104 ± 19.1	115 ± 11.6	14.10	<.001	0.33	AD < SMC, HC; MCI < HC
WMI	85 ± 16.9	96 ± 8.1	103 ± 15.3	115 ± 11.9	17.39	<.001	0.38	AD < MCI, SMC, HC; MCI < HC; SMC < HC
PSI	78 ± 17.1	87 ± 10.5	99 ± 13.8	115 ± 10.1	36.43	<.001	0.55	AD < SMC, HC; MCI < SMC, HC; SMC < HC
FSIQ	80 ± 18.8	95 ± 12.3	105 ± 17.7	120 ± 11.4	25.65	<.001	0.47	AD < SMC, HC; MCI < HC; SMC < HC

Note. This table includes the statistical comparisons between the clinical and healthy controls. Total sample size is $n = 100$. AD: Alzheimer's disease, MCI: mild cognitive impairment, SMC: subjective memory concerns, HC: healthy controls, VCI: Verbal Comprehension Index, PRI: Perceptual Reasoning Index, WMI: Working Memory Index, PSI: Processing Speed Index, FSIQ: Full Scale Intelligence Quotient.

^a Participants were assessed using either WAIS-IV with WMS-IV, or WAIS-III with WMS-III. As all scans and memory assessment in the HC sample was on the 3T scanner, WAIS-IV and WMS-IV, chi-square tests were re-ran using only the memory clinic population (SMC, MCI and AD). There were no significant frequency differences between these three samples in the type of scanner used, $\chi^2(2, N = 70) = 1.01$, $p = .604$, Cramer's $V = .13$, and the version of cognitive and memory batteries used, $\chi^2(2, N = 70) = 0.24$, $p = .888$, Cramer's $V = .06$.

^b Effect sizes reported are eta-square for the ANOVA comparisons and the Cramer's V for chi-square comparisons

^c VCI and POI (Perceptual Organisation Index) were reported for participants who were assessed using WAIS-III as they had the highest correlation with VCI and PRI in WAIS-IV. $r = .91$, $r = .84$ respectively (Wechsler, 2008).

the scan parameters as follows: CP head array coil, TR = 1530, TE = 3.41, FOV = 156mm * 250mm, flip angle = 8°, voxel size 1 * 1 * 1.4mm³, matrix 256 * 160. The 3T scans were acquired on a Siemens Skyra scanner with scan parameters as follows: 64-channel head coil, TR = 2200, TE = 2.03, FOV = 230 * 230, flip angle = 8°, voxel size = 0.9 * 0.9 * 1mm³, matrix 256 * 256.

There were occasional deviations from these exact parameters that resulted in some differences in total slice numbers, but in-plane resolution did not change. Both clinical and control scans were acquired using the standard clinical protocol routinely employed through St Vincent's Hospital memory clinic. An additional, consecutive T1-weighted MPRAGE was also acquired on the 3T scanner for the control group with the same scan parameters.

Volumetric measures. A total of six structures were included in this study: left and right hippocampus, left and right entorhinal cortex, and left and right transentorhinal cortex. Volumetric estimations of these six structures were obtained via manual and automated methods. Details of the manual and automated segmentations can be found in the next section.

Lateralisation of volumes was calculated by subtracting the right volume from the left volume for each structure to obtain a lateralisation measure based on raw volumes. To correct for overall variability in sizes between different subjects, a lateralisation index was calculated (Basso et al., 2006; Derflinger et al., 2011), using the following: $(\text{left} - \text{right}) / (\text{left} + \text{right}) * 100$. In both the lateralisation measure and lateralisation index, a more negative value indicates a smaller size (HC) or greater atrophy in the left volume. As mathematically, the lateralisation measure and index would cancel out the effects if the frequency of left and right lateralisation were similarly present in the samples, the asymmetry measure and index were calculated using the absolute value of the respective lateralisation measure and index to

capture the asymmetry in either hemisphere (Guadalupe et al., 2014; Sarica et al., 2018). In this case, the larger the value, the larger the asymmetry between the two hemispheres.

Volumetric Procedures

Automated segmentation. FreeSurfer (v6.0) was employed for the fully automated processing and segmentation of scans. FreeSurfer is one of the most extensively utilised automated brain image analysis software and technical details can be found in select publications (Fischl & Dale, 2000; Fischl et al., 2002; Fischl et al., 2004). Automated segmentations of scans were completed on a Linux workstation over approximately six to eight hours analysis per scan on one core.

The FreeSurfer cortical reconstruction process was performed on raw DICOM images using the “recon-all” command. Recon-all was executed on two T1 scans in the healthy control group as it was expected to produce a more accurate demarcation of the structure boundaries. Recon-all was executed on a single T1 scan in the clinical group as only a single T1 scan is available as part of the standard clinical neuroimaging protocol.

To generate the values of the hippocampal subfields, an additional flag, “-hippocampal-subfields-T1”, was appended after the recon-all command (Iglesias et al., 2015). The developers of this hippocampal subfield algorithm recommended this tool as a better estimate of whole hippocampal volumes compared to the volumes generated by the standard subcortical segmentation protocol and therefore this subfield protocol was used in this study for validation. The volume statistics for the subfield protocol were extracted directly from an automatically generated text file. To more closely approximate the automated volumes to the manual protocol used in this study, the hippocampal tail was subtracted from the whole hippocampal volume to generate the final whole hippocampal estimates used in this study.

Volume statistics for the entorhinal cortex and transentorhinal cortex were based on the outputs generated by cytoarchitecturally-defined probabilistic labels, a process that is

included as part of the standard pipeline (Augustinack et al., 2013; Fischl et al., 2009). As part of the standard pipeline, anatomical statistics for the thresholded cytoarchitecturally-defined structures were generated by default from their annotation files and their unique values could be found in the automatically-generated stats file under their corresponding names, “entorhinal cortex” and “perirhinal cortex”, the “perirhinal cortex” being analogous to the transentorhinal cortex.

Worth noting was that, around the time of writing, a systematic mislabelling of the structures generated in the FreeSurfer stats files was detected, the details of this mislabelling reported in a letter to the editors of the Journal of Alzheimer’s diseases (Appendix A in Chapter 3; Fung et al., 2019b). In short, the “entorhinal cortex” was labelled as the more lateral structure and the “perirhinal cortex” was labelled as the more medial structure, although the reverse is the more accurate representation of their typical anatomical location. Therefore, to work around the labelling error, the values labelled under “perirhinal cortex” were used in the analysis of the entorhinal segmentation while the values labelled under “entorhinal cortex” were used in the analysis of the transentorhinal segmentation.

Manual segmentation. Volumes of the hippocampus, the entorhinal cortex and the transentorhinal cortex were determined by manual delineation on the coronal T1-weighted slices using the ROI measure on Analyze 12.0 (Brain Imaging Resource, Mayo Foundation, Rochester, MN, USA, version 12 released 2017). To prepare for manual tracing, raw DICOM files were converted into volumes of 0.3*0.3*1.4mm voxel size using cubic spline and saved into Analyze format. The x and y dimensions were converted to 0.3 voxels to improve the resolution for manual tracing in both 1.5T and 3T scans. The slice thickness for 3T scans was converted into 1.4mm for comparability with the 1.5T scans. The original slice thickness of 1.4mm was preserved in 1.5T scans.

The hemisphere in which manual segmentation begun was counterbalanced to avoid possible bias. The hippocampus was traced according to the protocol described in Cook et al. (1992). In brief, the posterior limit of the hippocampus was the longest length of the unbroken fornix and the anterior limit was the last slice where the hippocampus could be discerned. The outer border of the hippocampus was determined by the alveus. The entorhinal and transentorhinal cortices were traced together on the same slices according to the protocol described in Kivisaari et al. (2013). According to the protocol by Kivisaari et al. (2013), the anterior border of the entorhinal and transentorhinal cortices were defined to be 2mm posterior to the first anterior slice of visible limen insulae while the posterior border was defined to be 1mm posterior to the last slice of visible apex of the limbic gyrus. Due to differences in slice thickness, the anterior limit was chosen to be one slice caudal to the first anterior slice in which the white matter of the limen insulae is visible, and the posterior limit was chosen to be one slice caudal to the last slice where the intra-limbic gyrus is visible. This was done to avoid underestimation of the structures.

Manual segmentation of the hippocampus, entorhinal and transentorhinal cortices was completed by a single investigator who was blinded to the subject's diagnosis and cognitive results. All manual delineation of the healthy control scans was only performed on the first 3T scan obtained. Ten scans based on different scanner strengths, 1.5T scanner ($n = 10$, mean age = 71.5 ± 8.9 years) and 3T scanner ($n = 10$, mean age = 69.4 ± 6.9 years), were randomly selected from the clinical group to assess for intra-rater reliability using intraclass correlation coefficient (ICC), two-way mixed, absolute agreement single-measure (Cicchetti, 1994; McGraw & Wong, 1996; Portney & Watkins, 2013; Shrout & Fleiss, 1979).

Statistical Analysis

Statistical analysis was conducted using SPSS (version 25). Pearson correlation was used to investigate the associations between the manual and automated volume estimations.

Paired-sample *t*-test was used to assess for significant differences in volumes between the left and right structures. Comparisons between the four groups for differences in volumes, lateralisation and asymmetry were performed using ANOVA with Tukey post hoc.

The relationships between different MTL structures and the WMS index scores were examined using stepwise linear regression, using six MTL structures as predictors and the WMS index scores as the dependent variable. Three WMS index scores were separately investigated in this analysis, *auditory memory*, *visual memory* and *general memory*. Control variables included in all the linear regression analyses were gender and membership in the clinical or control group. All categorical variables, which included gender, membership in the clinical or control group, were dummy coded before entry into the regression. Other control variables included in the initial analyses were age and years of education, but neither were significant contributors in the final models. The squared value of the semi-partial correlations was used as a measure of effect size (Cohen, Cohen, West, & Aiken, 2013).

Multinomial logistic regression was used to investigate the predictive power that MTL volumetric measures provide on top of comprehensive cognitive assessment, as assessed by *general memory*, in predicting membership in the AD, MCI, SMC or HC group. Only *general memory* was used as a measure of memory performance as the literature have shown that the domain of general memory, which includes both immediate and delayed indices, captured both auditory and visual memory (Holdnack et al., 2011; Price et al., 2002). Initially, a less stringent criterion of $p < .15$ was used to select predictors from the full model to include into a reduced model as predictors were likely to be excluded by another predictor (Hosmer, Lemeshow, & Sturdivant, 2013). A more traditional stringent criterion of $p \leq .05$ was employed for inclusion of predictors in the final culled model. The “goodness of fit” of the regression model was assessed by Pearson chi-square test and the odds ratio was used to

assess the association between the predictor and diagnostic classification (Hosmer et al., 2013). The odds ratio was also used as a measure of effect size.

Results

Demographic and Cognitive Characteristics

As shown in Table 6.1, there were no significant differences in age between the four samples: AD, MCI, SMC and HC ($p = .11$). There were significantly more male participants in the AD sample and significantly more female participants in the HC sample ($p = .001$). The HC sample had also received significantly more years of education than the other three clinical groups ($p < .001$). As expected, there were significantly more participants (100%) who underwent a structural MRI using the 3T scanner and undertook cognitive assessment using WAIS-IV and WMS-IV in the HC sample compared to those in the clinical samples ($p < .05$), but there were no differences in the scanners and WAIS/WMS version used between the three clinical groups ($p > .05$; Table 6.1).

On all three measures of memory function, the AD sample scored significantly less than MCI, SMC and HC, while MCI sample scored significantly less than SMC and HC, and SMC scored significantly less than HC ($p < .001$). Details of analyses on demographics and cognitive performance are found in Table 6.1.

Inter- and Intra-rater Correlations

On the 1.5T scans, excellent single measure intra-rater single-measure ICCs of .98 (95% CI: .93, .99) and .99 (95% CI: .96, .99) was achieved on left and right hippocampal estimates respectively. Similar excellent results were found on the 3T scans, achieving single measure intra-rater single-measure ICCs of .99 (95% CI: .95, .99) and .99 (95% CI: .97, .99) on the left and right hippocampus respectively.

In the 1.5T scans, intra-rater reliability found for manual entorhinal segmentation was moderate to excellent on the left, single-measure ICC = .90 (95% CI: .66, .97), and good to

excellent on the right, single-measure ICC = .95 (95% CI: .80, .99). Intra-rater reliability for manual transentorhinal segmentation was good to excellent on the left, single-measure ICC = .96 (95% CI: .84, .99) and excellent on the right, single-measure ICC = .98 (95% CI: .93, .99). Single-measure ICC values found in the 3T scans were good to excellent on the left entorhinal manual estimations, ICC = .97 (95% CI: .88, .99), and excellent on the right entorhinal manual estimations, ICC = .98 (95% CI: .90, .99). Intra-rater reliability for manual transentorhinal segmentation was excellent on the left, single-measure ICC = .98 (95% CI: .90, .99), and good to excellent on the right, single-measure ICC = .97 (95% CI: .87, .99).

Volume Differences Across Groups

Manual estimations. Table 6.2 shows significant correlations between all the MTL structures as measured with manual delineation via the Analyze software, except in the right transentorhinal cortex, for which left and right hippocampal volumes were not correlated, $r = .18$, $p = .067$ and $r = .18$, $p = .080$ respectively. The left transentorhinal cortex was weakly correlated with both the left hippocampal volume, $r = .24$, $p = .017$, and the right hippocampal volume, $r = .23$, $p = .023$. However, correlations between the volumes found in each of the left and right hippocampus, left and right entorhinal cortex, and left and right transentorhinal cortex ranged from strong to very strong, $r = .67 - .89$, $p < .01$ (Table 6.2).

Figure 6.1 shows the distribution of the MTL volumes within each participant sample and the Tukey post-hoc analyses of the manual estimates for the three MTL structures between the four samples. Further quantitative details are shown in Table 6.3. An ANOVA revealed significant differences in the left hippocampal volumes between the four samples with a large effect size and Tukey post-hoc showed that the AD sample had significantly smaller volumes than SMC ($p = .003$) and HC ($p < .001$), and MCI presented with significantly smaller volumes than SMC ($p = .040$) and HC ($p = .004$). Similarly, ANOVA showed significant differences in the right hippocampal volume across the samples with a

large effect size. Tukey post-hoc test showed the AD sample had significantly smaller right hippocampal volumes than SMC ($p < .001$) and HC ($p < .001$; Figure 6.1).

There were no significant differences between the samples in terms of the left entorhinal, left transentorhinal and right transentorhinal volumes ($p > .05$; Table 6.3). However, an ANOVA demonstrated significant differences in the right entorhinal volumes between the four samples with a medium to large effect size, and at post-hoc the right entorhinal volumes in the AD sample was found to be significantly smaller than SMC ($p = .002$) and HC ($p = .004$; Figure 6.1).

FreeSurfer estimations. Table 6.2 shows the correlations between all the MTL structural volumes that was estimated using FreeSurfer (v 6.0). There were significant correlations between all the structures, with the weakest correlation being between the left entorhinal and the right transentorhinal automated volume estimates, $r = .22$, $p = .027$, and the strongest correlation being between the left and right hippocampal automated volume estimates, $r = .85$, $p < .01$. Correlations between the left and right volumes for the entorhinal cortex and transentorhinal cortex were strong, $r = .61$, $p < .01$, and moderate, $r = .46$, $p < .01$ respectively (Table 6.2).

Figure 6.1 shows the distribution of the MTL volumes within each sample and the Tukey post-hoc analyses of the three MTL volume estimates between all four samples. The ANOVA analyses, as detailed in Table 6.3, revealed significant differences between samples with large effect sizes for both the left and right hippocampal automated volumes. Tukey post-hoc analyses showed that both AD and MCI samples had significantly smaller left hippocampus than SMC ($p = .001$, $p = .014$ respectively) and HC ($p < .001$, $p = .001$ respectively; Figure 6.1). Tukey post-hoc test the AD sample had significantly smaller right hippocampal volumes than SMC ($p < .001$) and HC ($p < .001$; Figure 6.1).

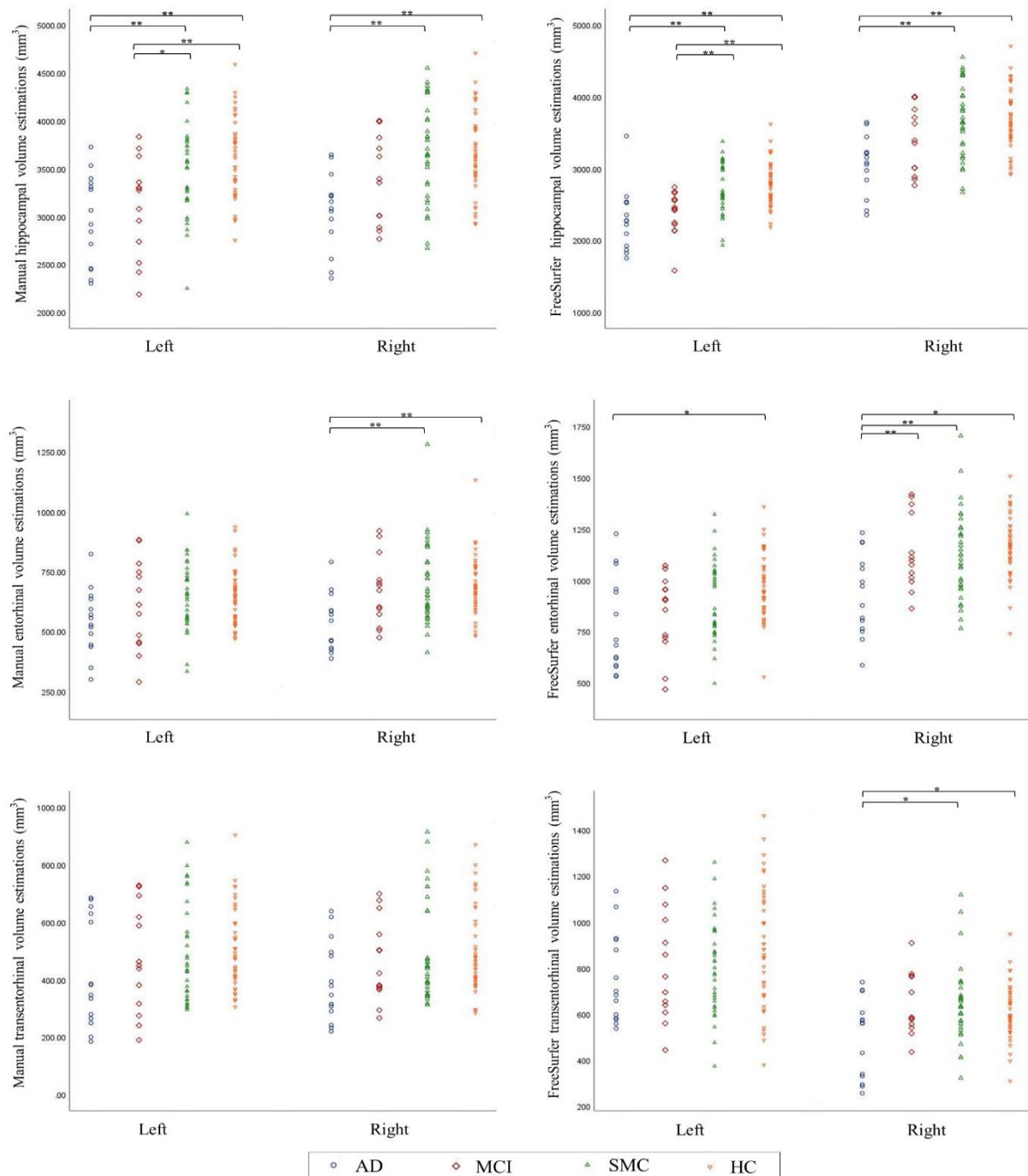
Table 6.2*Correlations Between the Different Structural Volumes Assessed by Manual Delineation and FreeSurfer*

	Manual					FreeSurfer					
	Hippocampus	Entorhinal		Transentorhinal		Hippocampus	Entorhinal		Transentorhinal		
	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
Manual											
Left hippocampus	.89**	.39**	.43**	.24*	.18	.87**	.84**	.64**	.49**	.39**	.28**
Right hippocampus		.35**	.46**	.23*	.18	.75**	.89**	.53**	.55**	.35**	.41**
Left entorhinal			.67**	.59**	.41**	.41**	.38**	.39**	.32**	.13	.21*
Right entorhinal				.63**	.64**	.44**	.50**	.30**	.37**	.02	.23*
Left transentorhinal					.83**	.24*	.25*	.18	.07	-.31**	-.13
Right transentorhinal						.21*	.24*	.13	.09	-.30**	-.17
FreeSurfer											
Left hippocampus							.85**	.63**	.47**	.38**	.30**
Right hippocampus								.57**	.57**	.39**	.41**
Left entorhinal									.61**	.55**	.22*
Right entorhinal										.48**	.56**
Left transentorhinal											.46**

Note. Total sample size is $N = 100$ and all volumes were measured in mm^3 . * $p < .05$, ** $p < .01$

Figure 6.1

Scatterplot and Tukey Post-hoc Comparisons of Medial Temporal Structural Volumes



Note. The figure shows the scatterplot and Tukey post-hoc comparisons of the manual and FreeSurfer estimates of hippocampal, entorhinal and transentorhinal structural volumes (mm³), separated by hemispheres and diagnosis. AD: Alzheimer's disease, MCI: mild cognitive impairment, SMC: subjective memory concerns, HC: healthy controls. * $p < .05$, ** $p < .01$

There were also significant differences between the automated measurement of volumes of the left and right entorhinal cortices between the samples (Table 6.3). Post-hoc tests demonstrated significantly smaller left entorhinal cortex in the AD sample when compared to HC ($p = .012$), while right entorhinal cortical volume in the AD sample was significantly smaller than MCI, SMC and HC samples ($p = .021$, $p = .006$, $p < .001$ respectively; Fig. 6.1). In terms of the FreeSurfer-estimated transentorhinal cortical volumes, there were no significant differences in the left hemispheric volumes between the groups (Table 6.3). However, with the right transentorhinal volume in the AD sample was significantly smaller than the SMC ($p = .049$) and HC samples ($p = .015$; Figure 6.1).

Left-Right Hemispheric Differences

Manual estimations. The patterns of asymmetry between the left and right hemispheres for the three MTL structures and the respective t -test results are reported for the different samples and are shown in Figure 6.2. Paired-sample t -test showed that both MCI and SMC presented with significantly larger right hippocampal volumes than the left, while this pattern of right-larger-than-left trended towards significance in HC. The right entorhinal volumes were found to be larger than the left in SMC and HC. There were no significant differences between the left and right transentorhinal volumes in any of the samples.

The lateralisation between the left and right hippocampus within each sample differed significantly from each other when analysed using ANOVA with post-hoc test. As shown in the boxplots in Figure 6.3, lateralisation was more negative (left < right) in MCI than that found in AD and HC, for both the lateralisation measure that used raw volumes ($p = .034$, $p = .003$ respectively) and the lateralisation index ($p = .034$, $p = .001$ respectively). No differences between samples in laterality was found in the entorhinal and transentorhinal cortical volumes (Figure 6.3).

Table 6.3

Average Raw Volumes of Hippocampus, Entorhinal and Transentorhinal Cortices Between the Clinical and Healthy Control Samples

Volume (mm ³)	AD (<i>n</i> = 14)	MCI (<i>n</i> = 13)	SMC (<i>n</i> = 33)	HC (<i>n</i> = 40)	ANOVA		
					<i>F</i> -value	η ²	<i>p</i> -value
Manual							
Hippocampus							
Left	2975 ± 471	3098 ± 512	3507 ± 493	3613 ± 414	9.13	0.22	<.001
Right	3055 ± 402	3415 ± 470	3676 ± 504	3668 ± 436	7.47	0.19	<.001
Entorhinal							
Left	541 ± 136	612 ± 188	646 ± 140	646 ± 118	2.31	0.06	.081
Right	532 ± 118	669 ± 145	699 ± 170	688 ± 123	5.01	0.13	.003
Transentorhinal							
Left	419 ± 188	469 ± 186	485 ± 170	510 ± 138	1.13	0.03	.342
Right	393 ± 140	466 ± 143	490 ± 172	498 ± 147	1.73	0.05	.166
FreeSurfer							
Hippocampus							
Left	2302 ± 437	2372 ± 312	2717 ± 352	2793 ± 304	10.45	0.25	<.001
Right	2307 ± 286	2588 ± 294	2784 ± 337	2846 ± 298	11.71	0.27	<.001
Entorhinal							
Left	786 ± 233	834 ± 192	904 ± 188	967 ± 161	4.05	0.11	.009
Right	929 ± 200	1137 ± 184	1125 ± 210	1164 ± 150	5.84	0.15	.001
Transentorhinal							
Left	755 ± 196	817 ± 250	801 ± 201	893 ± 263	1.61	0.05	.193
Right	496 ± 169	637 ± 132	642 ± 164	617 ± 123	3.52	0.10	.018

The measures of asymmetry were also assessed (Figure 6.4). Post-hoc testing showed that the asymmetry in raw volumes between the left and right hippocampus was larger in the MCI than in the HC sample ($p = .012$). Both the asymmetry indices in AD and MCI were larger than HC ($p = .023$, $p = .003$ respectively). The degree of asymmetry between the left and right transentorhinal cortex in MCI was larger than both SMC and HC in raw volume asymmetry measure ($p = .032$, $p = .049$ respectively), as well as in their asymmetry index ($p = .004$, $p = .003$ respectively). No asymmetry between left and right entorhinal cortex was observed across the cohorts.

FreeSurfer estimations. As shown in Figure 6.2, the right hippocampus was significantly larger than the left hippocampus in MCI, SMC and HC. The right entorhinal cortex was found to be larger than the left entorhinal cortex in all four samples and the left transentorhinal cortex was found to be larger than the right transentorhinal cortex in all four samples.

For the ANOVA results, as shown in Figure 6.5, the lateralised raw volume difference between the left and right hippocampal volumes (left smaller than right) were significantly different at post-hoc testing between AD and MCI ($p = .040$). After accounting for total hippocampal volume sizes, the lateralisation index in MCI was significantly more negative than AD ($p = .047$) and HC ($p = .029$). The ANOVA results did not show significant differences between the samples for lateralised automated entorhinal and transentorhinal volumes (Figure 6.5).

As shown Figure 6.6, the ANOVA results assessing asymmetry between the left and right hippocampal volumes showed significant differences between the samples. The asymmetry measure in raw hippocampal volumes in AD was larger than that found in SMC ($p = .028$) and HC ($p = .047$), and MCI was found to be larger than SMC ($p = .010$) and HC ($p = .016$). Similar findings were demonstrated in the hippocampal asymmetry index, in

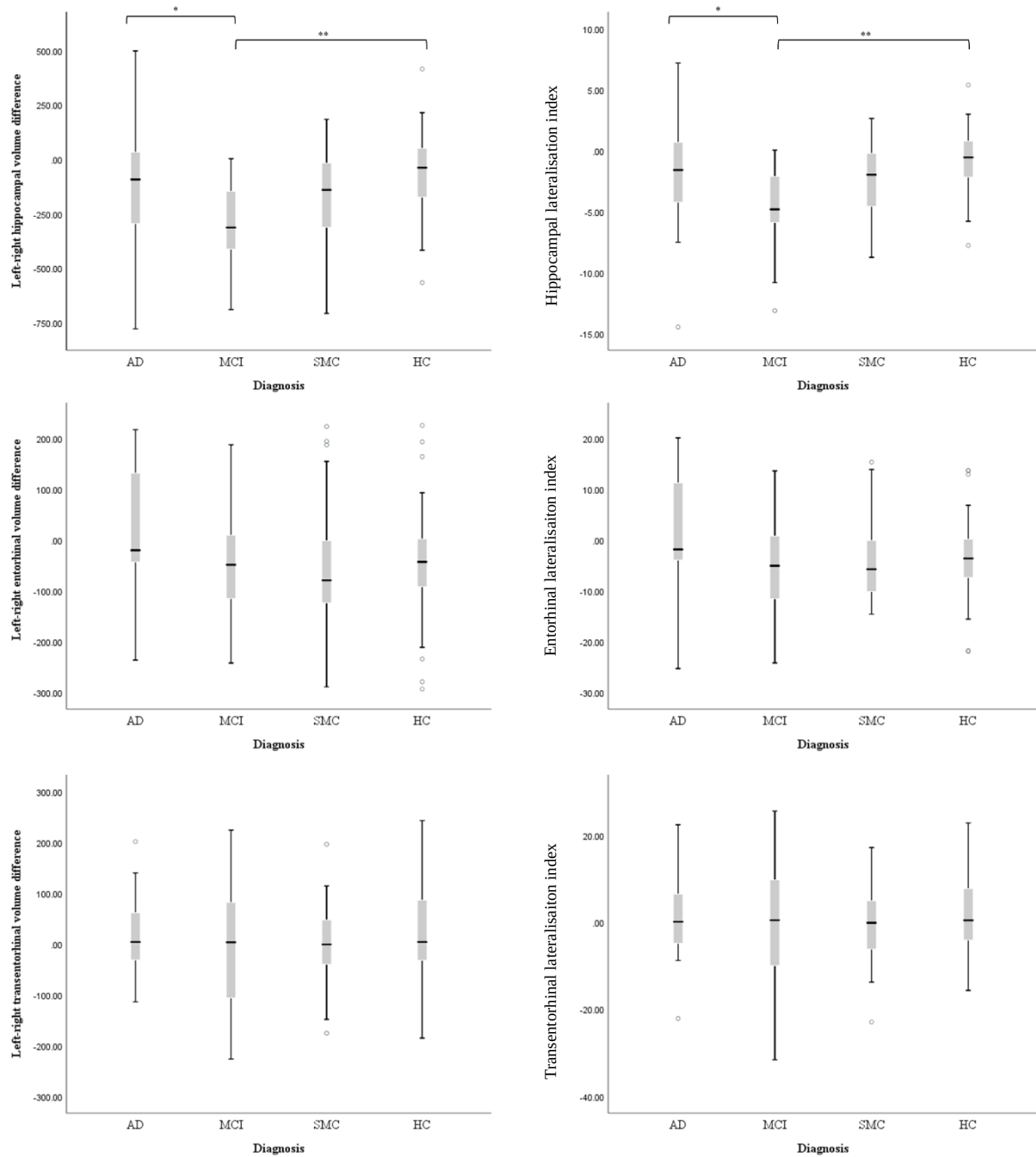
Figure 6.2*Volumetric Differences Between Left and Right Medial Temporal Lobe Volumes*

	AD (n = 14)	MCI (n = 13)	SMC (n = 33)	HC (n = 40)
Manual				
Hippocampus	 $M = -79.9, SD = 346$ $p = .40, d = 0.23$	 $M = -316, SD = 226$ $p < .001, d = 1.40$	 $M = -169, SD = 221$ $p < .001, d = 0.76$	 $M = -55.3, SD = 182$ $p = .06, d = 0.30$
Entorhinal	 $M = 8.96, SD = 129$ $p = .80, d = 0.07$	 $M = -56.6, SD = 116$ $p = .10, d = 0.49$	 $M = -52.7, SD = 121$ $p = .018, d = 0.43$	 $M = -41.2, SD = 113$ $p = .027, d = 0.36$
Transentorhinal	 $M = 26.1, SD = 85$ $p = .27, d = 0.31$	 $M = 2.98, SD = 138$ $p = .94, d = 0.02$	 $M = -5.46, SD = 81.7$ $p = .70, d = 0.07$	 $M = 11.8, SD = 88.7$ $p = .41, d = 0.13$
FreeSurfer				
Hippocampus	 $M = -5.41, SD = 364$ $p = .96, d = 0.01$	 $M = -216, SD = 239$ $p = .007, d = 0.91$	 $M = 67.2, SD = 155$ $p = .02, d = 0.43$	 $M = -52.6, SD = 137$ $p = .02, d = 0.38$
Entorhinal	 $M = -142, SD = 220$ $p = .03, d = 0.65$	 $M = -303, SD = 157$ $p < .001, d = 1.92$	 $M = -221, SD = 168$ $p < .001, d = 1.31$	 $M = -196, SD = 157$ $p < .001, d = 1.25$
Transentorhinal	 $M = 259, SD = 182$ $p < .001, d = 1.4$	 $M = 180, SD = 188$ $p = .005, d = 0.96$	 $M = 159, SD = 207$ $p < .001, d = 0.77$	 $M = 276, SD = 227$ $p < .001, d = 1.22$

Note. This figure shows a schematic representation of differences between the left and right volumes as estimated using either manual delineation or FreeSurfer v6.0. A larger circle diameter represents a significantly larger volume in the respective hemisphere as evaluated with paired-sample *t*-test (left – right volume, $p < .05$). Effect size is indicated below each pair, calculated using Cohen's $d = \text{mean} / \text{standard deviation}$ (Cohen, 1988). L: Left hemisphere, R: Right hemisphere, AD: Alzheimer's disease, MCI: mild cognitive impairment, SMC: subject memory concerns, HC: healthy controls.

Figure 6.3

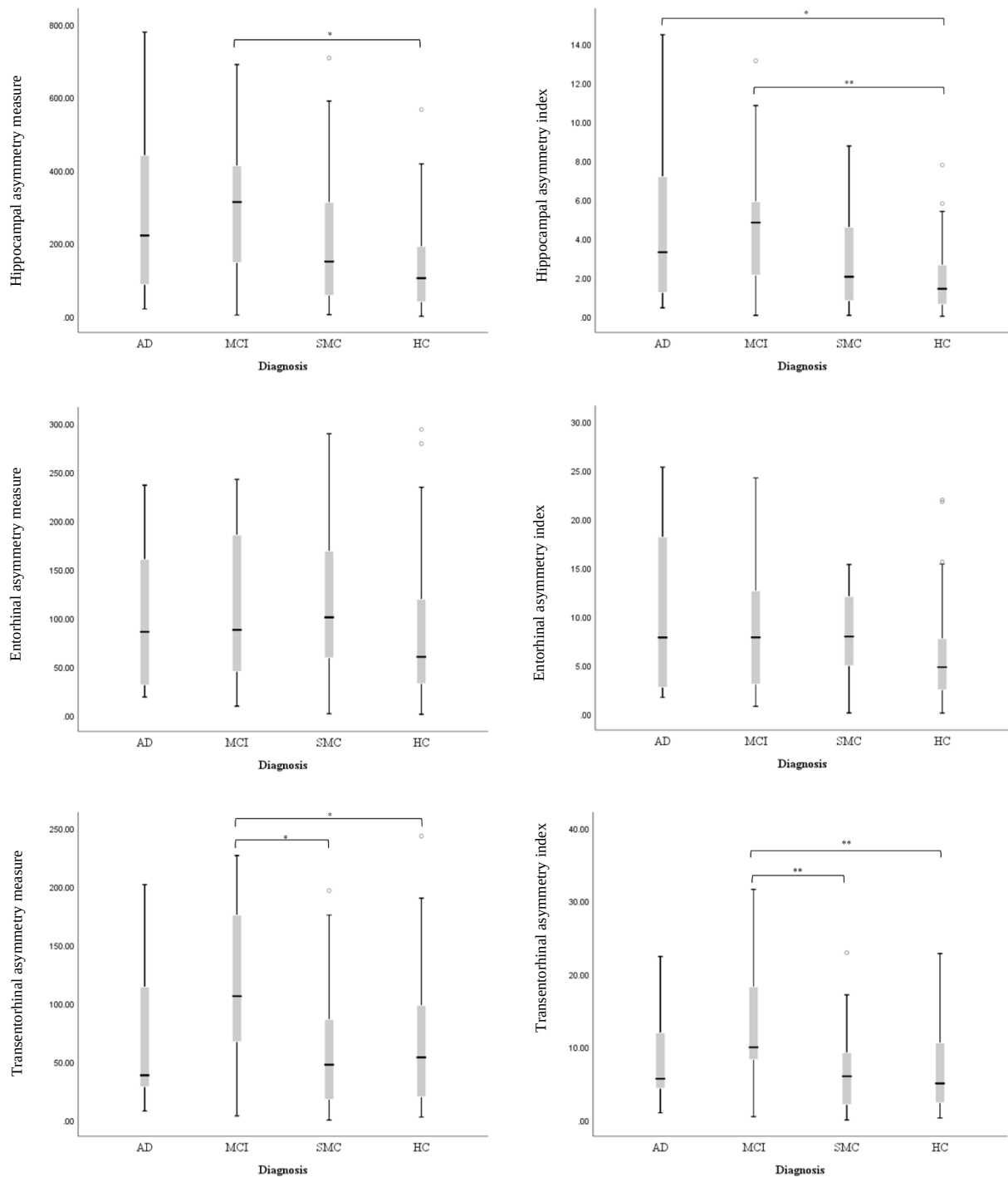
Boxplots of Lateralised Differences Between Left and Right Manual Volume Estimations



Note. Boxplots of manual estimations of the lateralised difference in raw volumes (mm³) between the left and right hemispheres (left – right volume) and the lateralisation indices of the structures. Lateralisation index was calculated using (left – right) / (left + right) x 100. Increasing negative value for the lateralised measure (raw volume) and lateralisation index indicates a smaller left hippocampus compared to the right. AD: Alzheimer’s disease, MCI: mild cognitive impairment, SMC: subject memory concerns, HC: healthy controls. * $p < .05$, ** $p < .01$

Figure 6.4

Boxplots of Asymmetry Between Left and Right Manual Volume Estimations



Note. Boxplots of manual estimations of asymmetry in raw volumes (mm^3) between the left and right hemispheres and the asymmetry indices of the structures. Asymmetry measure was calculated using the absolute difference between the left and right volumes. Asymmetry index was calculated using the absolute value of $(\text{left} - \text{right}) / (\text{left} + \text{right}) * 100$. The larger the values, the larger the asymmetry between the two hemispheres. AD: Alzheimer's disease, MCI: mild cognitive impairment, SMC: subject memory concerns, HC: healthy controls. * $p < .05$, ** $p < .01$

which the asymmetry index in AD was larger than that seen in SMC ($p = .008$) and HC ($p = .008$), and asymmetry index in MCI was larger than that seen in SMC ($p = .002$) and HC ($p = .002$). No significant asymmetrical differences between the left and right entorhinal volumes were found across the groups for both raw volumes and index measures (Figure 6.6).

Although no significant asymmetry differences in raw volumes were found between the left and right transentorhinal cortex, the transentorhinal asymmetry index in AD was found to be significantly larger than that found in MCI ($p = .018$) and SMC ($p = .022$).

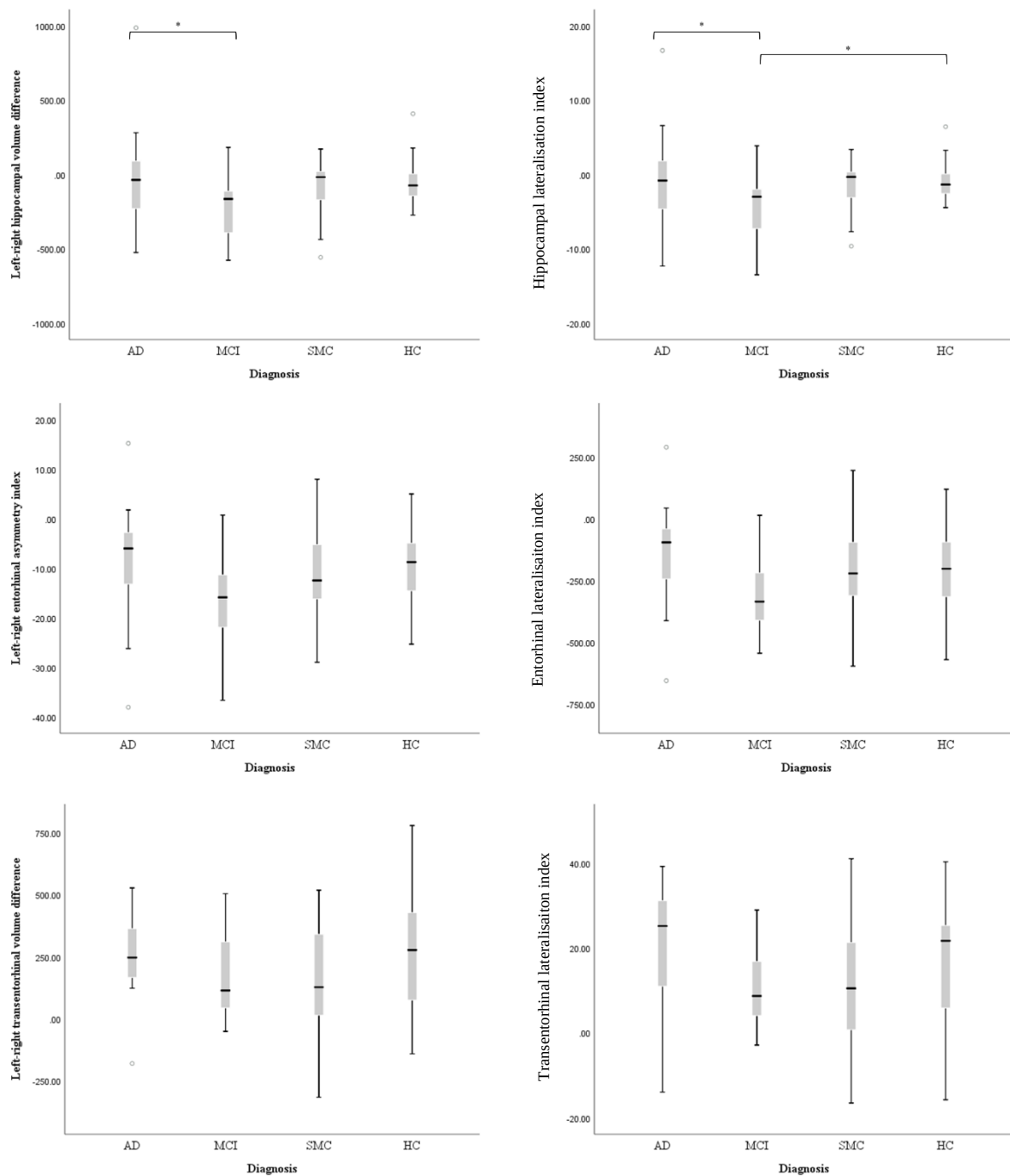
Relationship Between Medial Temporal Structures and Memory Performance

Auditory memory. Auditory memory performance was used as a dependent variable in a stepwise linear regression with either all manual or FreeSurfer MTL volume estimations as predictors entered concurrently with the control variables mentioned earlier. In both manual and FreeSurfer volume estimations, the left hippocampus was found to be the best predictor of auditory memory performance, $F(3, 96) = 31.18, p < .001, R^2 = .49$ and $F(3, 96) = 32.51, p < .001, R^2 = .50$ respectively. Manual and FreeSurfer left hippocampal estimates accounted for 8% and 9% of the variance respectively. In both regressions, the only other independent variables that were included in the final stepwise regression model was diagnostic category, and gender. Details of the results are found in Table 6.4.

Visual memory. Stepwise linear regression found that only the right manual estimation of entorhinal volume was a significant predictor of visual memory performance, $F(2, 97) = 28.94, p < .001, R^2 = .37$, accounting for 5% of the variance in visual memory. Similarly, only right FreeSurfer estimation of entorhinal volume was a significant predictor of visual memory, $F(2, 97) = 30.14, p < .001, R^2 = .38$. and accounted for 6% of the variance. In both regressions, the other independent variable that was included in the final stepwise regression model was diagnostic category. Details of the final model results are found in Table 6.5.

Figure 6.5

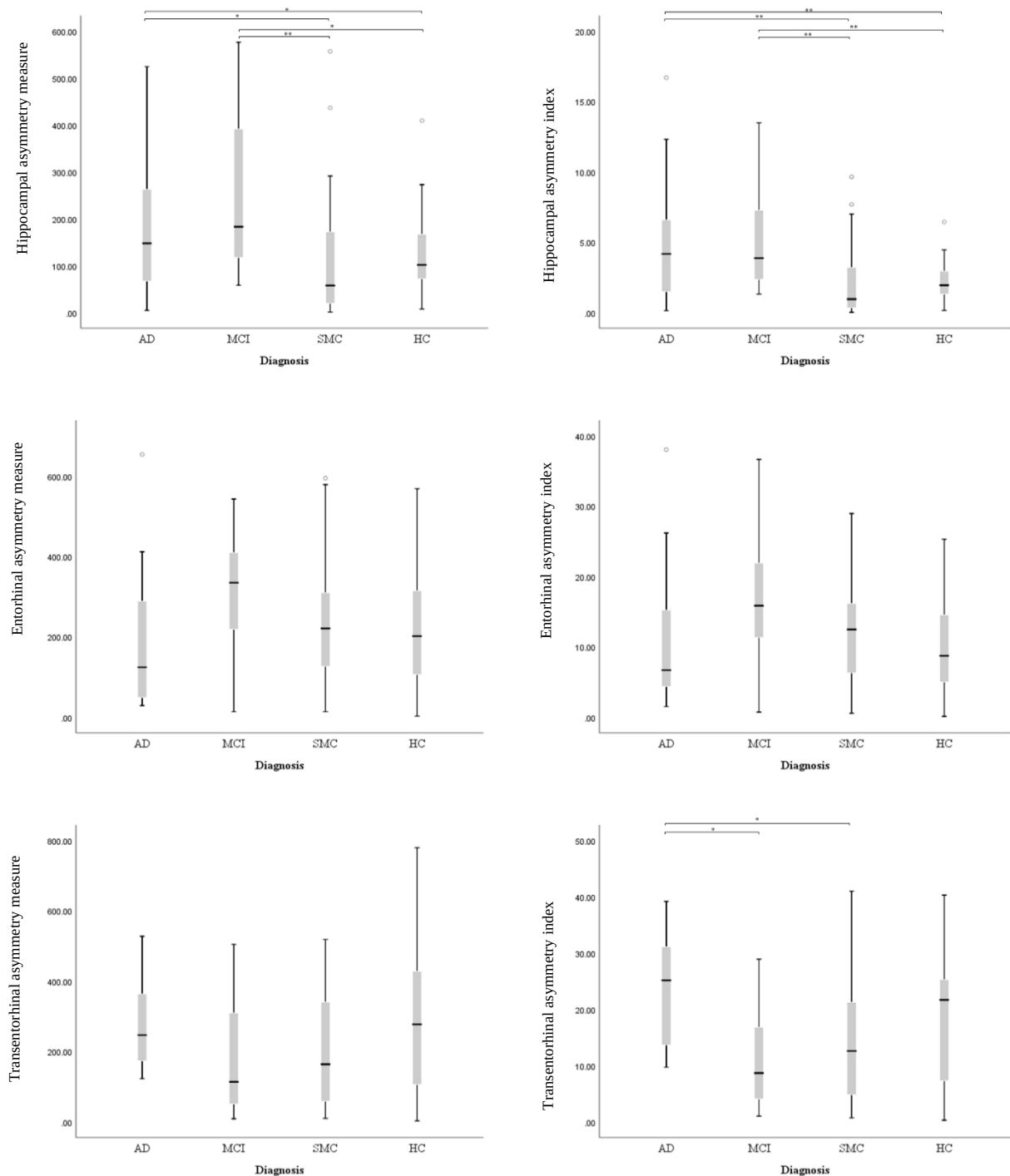
Boxplots of Lateralised Differences Between Left and Right FreeSurfer Volume Estimations



Note. Boxplots of FreeSurfer estimations of lateralised difference in raw volumes (mm^3) between the left and right hemispheres (left – right volume) and the lateralisation indices of the structures. Lateralisation index was calculated using $(\text{left} - \text{right}) / (\text{left} + \text{right}) \times 100$. Increasing negative value for the lateralised measure (raw volume) and lateralisation index indicates a smaller left hippocampus compared to the right. AD: Alzheimer's disease, MCI: mild cognitive impairment, SMC: subject memory concerns, HC: healthy controls. * $p < .05$

Figure 6.6

Boxplots of Asymmetry Between Left and Right FreeSurfer Volume Estimations



Note. Boxplots of FreeSurfer estimations of asymmetry in raw volumes (mm^3) between the left and right hemispheres and the asymmetry indices of the structures. Asymmetry measure was calculated using the absolute difference between the left and right volumes. Asymmetry index was calculated using the absolute value of $(\text{left} - \text{right}) / (\text{left} + \text{right}) * 100$. The larger the values, the larger the asymmetry between the two hemispheres. AD: Alzheimer's disease, MCI: mild cognitive impairment, SMC: subject memory concerns, HC: healthy controls. * $p < .05$, ** $p < .01$

General memory. After accounting for the other independent variables, diagnostic category and gender, only the left hippocampal volume was a significant predictor of general memory performance. This result was found in both methods of volume estimations, manual estimations, $F(3, 96) = 32.76, p < .001, R^2 = .51$ and FreeSurfer estimations, $F(3, 96) = 34.47, p < .001, R^2 = .52$. Left hippocampal manual estimation accounted for 6% of the variance and left hippocampal FreeSurfer estimation accounted for 8% of the variance. Results of the final regression models are found in Table 6.6.

Table 6.4

Stepwise Multiple Regression Analysis Examining the Relationship between Auditory Memory and MTL Volumes

	Unstandardised		β	<i>t</i> -value	<i>p</i> -value	Correlations		<i>sr</i> ²
	<i>B</i>	<i>SE (B)</i>				Partial	Semi-partial	
Manual^a								
Control/Clinical sample	20.18	3.807	0.42	5.30	<.001	.48	.39	.15
Gender	13.28	3.591	0.28	3.70	<.001	.35	.27	.07
Left hippocampus	0.01	0.003	0.29	3.79	.002	.36	.28	.08
FreeSurfer^b								
Control/Clinical sample	19.57	3.79	0.41	5.16	<.001	.47	.37	.14
Gender	13.77	3.56	0.29	3.87	<.001	.37	.28	.08
Left hippocampus	0.02	0.005	0.31	4.09	<.001	.38	.29	.09

Note. Dependent variable is auditory memory index. The square of the semi-partial correlation is indicated by *sr*², which serves as a measure of the effect size of the predictor variable after controlling for the other variables in the model. Total sample size is *N* = 100.

^a All variables were entered simultaneously and analysed in a stepwise manner. $F(3, 96) = 31.18, p < .001, R^2 = .49$

^b All variables were entered simultaneously and analysed in a stepwise manner. $F(3, 96) = 32.51, p < .001, R^2 = .50$

Table 6.5

Stepwise Multiple Regression Analysis Examining the Relationship between Visual Memory and MTL Volumes

	<u>Unstandardised</u>		β	<i>t</i> -value	<i>p</i> -value	<u>Correlations</u>		<i>sr</i> ²
	<i>B</i>	<i>SE (B)</i>				Partial	Semi-partial	
Manual^a								
Control/Clinical Sample	19.17	2.85	0.54	6.72	<.001	.56	.54	.29
Right entorhinal	0.03	0.01	0.23	2.80	.037	.27	.22	.05
FreeSurfer^b								
Control/Clinical Sample	18.25	2.87	0.52	6.35	<.001	.54	.54	.29
Right entorhinal	0.02	0.01	0.25	3.07	.017	.36	.24	.06

Note. Dependent variable is visual memory. The square of the semi-partial correlation is indicated by *sr*², which serves as a measure of the effect size of the predictor variable after controlling for the other variables in the model. Total sample size is *N* = 100

^a All variables were entered simultaneously and analysed in a stepwise manner. $F(2, 97) = 28.94$, $p < .001$, $R^2 = .37$

^b All variables were entered simultaneously and analysed in a stepwise manner. $F(2, 97) = 30.14$, $p < .001$, $R^2 = .38$

Table 6.6

Stepwise Multiple Regression Analysis Examining the Relationship between General Memory and MTL Volumes

	<u>Unstandardised</u>		β	<i>t</i> -value	<i>p</i> -value	<u>Correlations</u>		<i>sr</i> ²	
	<i>B</i>	<i>SE (B)</i>				Partial	Semi-partial		
Manual^a									
Control/Clinical sample	20.66	3.52	0.46	5.87	<.001	.51	.42	.18	
Gender	11.48	3.32	0.26	3.46	.001	.33	.25	.06	
Left hippocampus	0.01	0.00	0.27	3.55	.004	.34	.25	.06	
FreeSurfer^b									
Control/Clinical sample	20.02	3.50	0.45	5.73	<.001	.50	.41	.16	
Gender	11.93	3.28	0.27	3.63	<.001	.35	.26	.07	
Left hippocampus	0.02	0.00	0.29	3.93	.001	.37	.28	.08	

Note. Dependent variable is *General Memory*. The square of the semi-partial correlation is indicated by *sr*², which serves as a measure of the effect size of the predictor variable after controlling for the other variables in the model. Total sample size is *N* = 100.

^a All variables were entered simultaneously and analysed in a stepwise manner. $F(3, 96) = 32.76$, $p < .001$, $R^2 = .51$

^b All variables were entered simultaneously and analysed in a stepwise manner. $F(3, 96) = 34.47$, $p < .001$, $R^2 = .52$

Relationship Between Volumetric Measures and Diagnosis

Multinomial regression was used to model the relationship between the individual medial temporal structures and diagnostic classification in AD, MCI, SMC or HC sample. The model that included only general memory as the predictor was showed that general memory was a significant predictor of diagnosis, $\chi^2(3, N = 100) = 191.83$, Nagelkerke $R^2 = .65$, $p < .001$ (Table B1). Coefficient estimates for diagnostic classifications were all significant ($p < .001$; Table B1) and overall classification accuracy for this model was at 71.4% for AD, 15.4% for MCI, 54.5% for SMC and 85% for HC, achieving an overall classification rate of 64%. Scanner type was not included as a predictor as it would act as a proxy to predicting classification into HC because all HC participants were scanned on the 3T scanner. This was evident from the lack of unique contribution of scanner type to the model when multinomial regression was only conducted with the clinical samples, AD, MCI and SMC ($p = .444$; Table B2 for details).

Manual estimations. Table 6.7 shows the final model for manual measurements that included only predictors with significant, unique contributions as recommended by Hosmer et al. (2013). The final regression model consisted of general memory, and the left and right hippocampal volumes, $\chi^2(9, N = 100) = 109.43$, Nagelkerke $R^2 = .72$, $p < .001$, and overall classification accuracy was 72%, with specific classification accuracy of AD at 78.6%, MCI at 30.8%, SMC at 63.6% and HC at 90%. As seen in Table 6.7, in this model, both general memory scores and the left hippocampal volume had significant coefficients for comparing MCI and SMC against HC. However, the effect size as indicated by the odds ratio, showed that the left hippocampal volume contribution to diagnostic classification was trivial, and did not survive corrections for multiple comparison. See Table B3 for details of the model prior to the removal of the non-significant predictors, which included the general memory scores and all the manually estimated MTL volumes.

As the left and right hippocampal volumes were significant predictors in the above final logistic regression model, the unique contribution of the total volume, lateralisation raw volume, lateralisation index, asymmetry raw volume and asymmetry index of the hippocampal volumes to predicting diagnostic classification was further examined. Although the lateralisation raw volume measure (left < right) and lateralisation index (left < right) were significant predictors ($p = .003$, $p = .007$ respectively), none of these variables had significant regression coefficients after correction for multiple comparisons. A summary of these findings is presented in Table 6.8. Details of the different prediction models are presented in Tables B5 – B9.

FreeSurfer estimations. The final model that included the automated FreeSurfer volume estimates provided results similar to that reported in the last section. As shown in Table 6.7, the final model consisted of general memory, and the left and right hippocampal volumes, $\chi^2(9, N = 100) = 106.44$, Nagelkerke $R^2 = .71$, $p < .001$, with an overall classification of 63%. Specific classification accuracies of AD, MCI, SMC and HC were at 85.7%, 23.1%, 48.5% and 80% respectively. As shown in Table 6.7, only general memory and the right hippocampal volume had significant coefficients. However, the effect size as indicated by the odds ratio, showed that the contribution of the right hippocampal volume to diagnostic classification was trivial and did not survive correction for multiple comparisons. See Table B4 for the model prior to the removal of non-significant predictors, which included the general memory scores and all the left and right MTL volumes as estimated using FreeSurfer.

As the left and right hippocampal volumes were significant predictors in the above final logistic regression model, the unique contribution of the total volume, lateralisation raw volume, lateralisation index, asymmetry raw volume and asymmetry index of the hippocampus to predicting diagnostic classification was further examined. Although the

Table 6.7

Multinomial Logistic Regression Analysis Examining the Relationship Between the General Memory, Volumetric Measures and Classification in the Clinical or Healthy Control Sample in Final Model

Predictor	Predictor Model			Coefficient Estimates ^c						
	χ^2	df	p-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value	
Manual ^a										
General Memory	72.625	3	<.001	AD	-0.302	0.739	0.649	±	0.842	<.001
				MCI	-0.160	0.853	0.791	±	0.918	<.001
				SMC	-0.088	0.916	0.874	±	0.959	<.001
Left hippocampus	12.850	3	.005	AD	0.011	1.011	0.961	±	1.063	.676
				MCI	-0.044	0.957	0.919	±	0.996	.030
				SMC	-0.029	0.971	0.943	±	1.000	.050
Right hippocampus	15.084	3	.002	AD	-0.038	0.963	0.911	±	1.018	.188
				MCI	0.033	1.034	0.995	±	1.074	.089
				SMC	0.026	1.026	0.998	±	1.055	.067
FreeSurfer ^b										
General Memory	66.49	3	<.001	AD	-0.294	0.746	0.657	±	0.846	<.001
				MCI	-0.157	0.855	0.795	±	0.920	< .001
				SMC	-0.085	0.919	0.880	±	0.959	< .001
Left hippocampus	8.78	3	.032	AD	0.019	1.019	0.975	±	1.064	.410
				MCI	-0.042	0.959	0.918	±	1.002	.061
				SMC	-0.008	0.992	0.968	±	1.023	.611
Right hippocampus	10.16	3	.017	AD	-0.062	0.940	0.886	±	0.997	.040
				MCI	0.016	1.016	0.971	±	1.063	.484

Note. AD: Alzheimer's disease, MCI: mild cognitive impairment, SMC: subjective memory impairment, HC: healthy controls.

^a Logistic Regression model was statistically significant, $\chi^2(9, N = 100) = 109.43$, Nagelkerke $R^2 = .72$, $p < .001$; Goodness of fit was not significant

^b Logistic regression model was statistically significant, $\chi^2(9, N = 100) = 106.44$, Nagelkerke $R^2 = .71$, $p < .001$; Goodness of fit was not significant.

^c The reference group is the HC. Each unit increase for *General Memory* is 1 standard score. Each unit increase for each MTL structure is 10 mm³. Significant p -values are bolded. The p -values presented are uncorrected for multiple comparisons. Total sample size is $N = 100$. β : Regression Coefficient, Odds ratio: exponential of the regression coefficient.

difference in raw volume between the left and right hippocampus ($p = .028$), lateralisation index ($p = .049$) and asymmetry index ($p = .043$) all provided significant unique contributions, none of those three predictors had significant coefficient estimates. A summary of the findings for these FreeSurfer hippocampal measures is presented in Table 6.8. See Tables B10 – B14 for full details on each prediction model.

Discussion

Histopathological studies have provided consistent evidence that the medial temporal structures are one of the earliest sites afflicted with AD abnormal tau pathology and correlating with the early cardinal signs of memory disruptions in people with AD. As reviewed above, this affliction disrupts the neuronal connectivity among the three main structures critical to memory formation in a highly stereotypical manner, following a pathway that starts in the transentorhinal cortex, followed by the entorhinal cortex and the hippocampus (Braak et al., 2006; Braak & Braak, 1995). The stereotypical nature of propagation allows the postulation that early volumetric changes to these structures could provide more timely diagnostic information. Specifically, this study aimed to investigate the relationship of these three medial temporal structures, the hippocampus, entorhinal cortex and transentorhinal cortex, to memory function, and the clinical utility of these volumes in detecting AD and MCI due to AD.

Medial Temporal Structural Correlates to Memory Performance

As mentioned above, most studies supported an important role for the left MTL structures in verbal memory abilities, although the role of the right MTL structures appear less defined (Bonner-Jackson et al., 2015; K. H. Chen et al., 2010; Ezzati et al., 2016; Gleißner et al., 1998; Jones-Gotman, 1986; Jones-Gotman et al., 1997; Squire & Butters, 1992). This study investigated the associations between memory performance and the

Table 6.8

Summary of Coefficient Estimates of the Sum, Difference Volume and Asymmetry Index of the Left and Right Hippocampal Volumes for Prediction of Classification in AD, MCI, SMC or HC Sample

Predictors	Predictor Model		Coefficient Estimates					Classification Accuracy	
	χ^2	<i>p</i> -value	Group	β	Odds Ratio	Odds ratio 95% CI	<i>p</i> -value	Group	Overall
Manual									
Sum	2.65	.449	AD	-0.001	0.999	0.998 $\pm$ 1.000	.148	78.6%	62.0%
			MCI	-0.001	0.999	0.999 $\pm$ 1.000	.296	15.4%	
			SMC	0.000	1.000	0.999 $\pm$ 1.001	.781	48.5%	
			HC					82.5%	
Lateralisation (raw volume)	13.65	.003	AD	0.020	1.021	0.973 $\pm$ 1.070	.398	78.6%	69.0%
			MCI	-0.037	0.963	0.928 $\pm$ 1.000	.047	30.8%	
			SMC	-0.027	0.974	0.948 $\pm$ 1.001	.055	54.5%	
			HC					90.0%	
Asymmetry (raw volume)	3.361	.339	AD	0.038	1.039	0.983 $\pm$ 1.098	.174	71.4%	61.0%
			MCI	0.042	1.043	0.995 $\pm$ 1.094	.079	23.1%	
			SMC	0.021	1.021	0.984 $\pm$ 1.059	.266	39.4%	
			HC					87.5%	
Lateralisation index	12.22	.007	AD	0.053	1.004	0.784 $\pm$ 1.417	.174	78.6%	70.0%
			MCI	-0.256	0.774	0.604 $\pm$ 0.991	.042	30.8%	
			SMC	-0.187	0.830	0.685 $\pm$ 1.004	.055	57.6%	
			HC					90.0%	
Asymmetry index	4.19	.242	AD	0.293	1.340	0.929 $\pm$ 1.934	.117	71.4%	62.0%
			MCI	0.330	1.392	0.997 $\pm$ 1.942	.052	15.4%	
			SMC	0.162	1.176	0.904 $\pm$ 1.529	.227	45.5%	
			HC					87.5%	
FreeSurfer									
Sum	4.48	.214	AD	-0.016	0.984	0.966 $\pm$ 1.002	.075	71.4%	60.0%
			MCI	-0.013	0.987	0.973 $\pm$ 1.001	.074	23.1%	
			SMC	-0.004	0.996	0.987 $\pm$ 1.006	.432	45.5%	
			HC					80.0%	
Lateralisation (raw volume)	9.10	.028	AD	-0.016	1.027	0.985 $\pm$ 1.071	.212	71.4%	63.0%
			MCI	-0.013	0.968	0.928 $\pm$ 1.009	.122	30.8%	
			SMC	-0.007	0.993	0.964 $\pm$ 1.023	.645	45.5%	
			HC					85.0%	
Asymmetry (raw volume)	7.51	.060	AD	0.038	1.039	0.998 $\pm$ 1.010	.228	71.4%	64.0%
			MCI	0.033	10.033	0.998 $\pm$ 1.009	.226	23.1%	
			SMC	-0.022	0.979	0.993 $\pm$ 1.003	.380	51.5%	
			HC					85.0%	
Lateralisation index	7.88	.049	AD	0.101	1.106	0.884 $\pm$ 1.384	.378	64.3%	64.0%
			MCI	-0.159	0.853	0.692 $\pm$ 1.052	.136	30.8%	
			SMC	-0.040	0.960	0.819 $\pm$ 1.126	.619	48.5%	
			HC					87.5%	
Asymmetry index	8.16	.043	AD	0.220	1.246	0.897 $\pm$ 1.731	.190	71.4%	65.0%
			MCI	0.231	1.260	0.948 $\pm$ 1.675	.111	30.8%	
			SMC	-0.075	0.927	0.715 $\pm$ 1.204	.571	51.5%	
			HC					85.0%	

Note. Each predictor is added onto the initial model that had *General Memory* as a predictor. Sum: left + right hippocampal volume, Lateralisation (raw volume): left– right hippocampal volume, Asymmetry (raw volume): absolute of lateralisation (raw volume), Lateralisation index: (left – right hippocampal volume)/sum of left and right hippocampal volume*100, Asymmetry index: absolute of lateralisation index. Reference group for coefficient estimates is healthy controls, and a unit change in difference and absolute difference is 10mm³, while a unit change in lateralisation index and asymmetry index is 1%. Significant findings are bolded. The *p*-values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

hippocampus, entorhinal and transentorhinal cortices on a sample of healthy participants and a heterogeneous memory clinic population. The results largely concurred with past findings, suggesting that the left hippocampal volume, in both manual and automated estimates, were most strongly associated with the scores on the WMS auditory memory index and general memory. Conversely, performance on the WMS visual memory tasks were more strongly associated with both manual and automated estimates of the right entorhinal volume.

Considering the spatiotemporal trajectory of AD-related tau propagation, this association between the right entorhinal volume and visual memory score suggests that visual memory symptoms could emerge as a clinical deficit earlier, with the disruption of the right entorhinal cortex. In contrast, verbal and general memory dysfunction may be symptoms that appear only later in the AD process, when the hippocampus becomes involved. This finding is in line with some studies that have seen stronger correlation of visual memory with changes in the MTL and observed the presence of visuospatial memory inefficiencies before the diagnosis of AD (Bonner-Jackson et al., 2015; Kawas et al., 2003; Kessels, Rijken, Joosten-Weyn Banningh, Van Schuylenborgh, & Olde Rikkert, 2010). Although language-based assessments are frequently used as a sole measure of memory in many studies (e.g. Apostolova et al., 2010; Sarica et al., 2018), the findings from this study, coupled with previous research (Bonner-Jackson et al., 2015; Kawas et al., 2003; Kessels et al., 2010), suggest that non-verbal measures could provide additional diagnostic value for early detection such as in people manifesting MCI due to AD.

Given this finding, more significant differences in right entorhinal volume between the MCI and HC samples might be expected, which was not the case in this study. However, there could be several reasons driving this absence of difference between the MCI and HC samples. Explanations include self-selection bias and a less lateralised visuospatial function,

both of which would be explored in more detail in the later part of this discussion that focuses on volume lateralisation and asymmetry.

Contribution of Volumetric Measures to Improving Diagnosis

Results from the group comparisons consistently showed that the AD and MCI samples had significantly smaller left hippocampal volumes than SMC and HC samples. The differences in the right hippocampus was less dramatic, with only the AD sample having a significantly smaller hippocampus than the SMC and HC samples, with no significant difference found from the MCI sample. The entorhinal measures were not as consistent across both methods of measurement, but overall, the group comparisons suggested that the right entorhinal volumes in the AD sample was smaller than that found in SMC and HC. Generally, these findings were consistent with past research and provided additional support that the volumetric MTL structures would serve as good surrogate pathophysiological evidence for AD (Weiner et al., 2015).

Several studies have demonstrated that patients with subjective memory impairment had reduced hippocampal and entorhinal volumes, which suggested that they could form part of the at-risk preclinical AD spectrum (Fan et al., 2018; Perrotin et al., 2017; Yue et al., 2018; Zhao et al., 2019). However, this study did not find any statistical volumetric differences between the SMC and HC samples. This lack of statistical difference may be due to the conservative sample size of $n = 33$ and $n = 40$ respectively, which limited the power of this sample to less than 55% to observe a small true effect size, as calculated by a power analysis software G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). A small effect size is considered to be barely noticeable (Cohen, 1988), and considering the clinical practicality of detecting differences between different diagnostic groups, a large true effect size is desirable. Therefore, in all likelihood, there are no clinically discernible volumetric differences between

these two samples, as this current sample size had more than 80% power to observe a medium to large effect size (Faul et al., 2007).

The volumetric similarities between these two samples could highlight two points. Firstly, subjective memory concerns not detectable on detailed objective testing are likely not driven by pathology, but by affective symptoms such as depression and anxiety (Auning et al., 2015; Buckley et al., 2013). Secondly, a lot of reported memory assessments in research defined subjective cognitive decline using abbreviated cognitive screening instruments known to have poor sensitivity and specificity. Together, the subtlety of detecting memory deficits in subjective memory complainers using objective memory assessments emphasises the need for future investigations to use comprehensive neuropsychological assessments in tandem with other biomarkers to build a more complete understanding of the diagnostic and prognostic utility of subjective memory complaints.

The current study also investigated the improvement in classification that the inclusion of quantitative MTL structures could add to differentiation of AD, MCI, SMC or HC. Overall classification accuracy that included only the general memory scores was 64%. The inclusion of manual estimations of the left and right hippocampal volumes in addition to the general memory scores improved classification accuracy to 72%. Automated measures of the left and right hippocampal volumes on the other hand, although shown to be significant contributors in the prediction model, provided little to no overall improvement to overall classification accuracy. However, this reduction in classification accuracy came mainly from the misclassification of the SMC and HC samples, which as mentioned earlier, were likely to be volumetrically more similar to each other than not. In both manual and automated methods, inclusion of the hippocampal measures made improvements in the classification of AD and MCI, although classification of MCI remained modest at 30.8% and 23.1% for the manual and automated measurements, respectively. The ability of the left and right

hippocampal volumes in predicting specific diagnostic groups was surprisingly poor, with neither hippocampal volume surviving significance testing after controlling for multiple comparisons

Although the relationship between hippocampal volumes and diagnosis did not appear to be as strong as predicted from the literature, the discriminatory ability of hippocampal volumes per se would have been complicated by the weak relationship found between hippocampal volume and memory function in healthy older adults (Van Petten, 2004). Other studies have also indicated that the relationship between medial temporal volumes and memory performance may be influenced mainly by people with disease (Apostolova et al., 2010; Dickerson et al., 2009). The results from this study indicated the presence of a disease-dependent differential relationship between memory and volume, as diagnostic classification in either the clinical or healthy control sample provided significant variance to memory performance. These findings support the possibility that older adults without AD are less reliant on specific MTL structures as they are able to more effectively recruit other parahippocampal networks for memory processing (Grady, 2012).

The complex nature of memory formation could account for the heterogeneity of memory performance observed in people with amnesic MCI due to AD, in which there might be early AD-related structural changes but their memory processing network is not as severely disrupted as people in the later stages of AD. Furthermore, it has been shown that neurons do not necessarily die but can continue to survive and function in a defective state even in the presence of tau-pathology (Stokin & Goldstein, 2006; Wirths & Bayer, 2010). Therefore, this could suggest that the degree of atrophy in people with the disease could be disproportionate to the severity of memory function, especially early on in the disease, during which there is relatively less neuronal loss but a massive number of dysfunctional neurons (Braak & Del Tredici, 2015b). Therefore, while the evaluation of structural volumes could

indicate the presence or absence of pathological abnormalities, there is a need to concurrently examine the individual's memory performance.

The transentorhinal and entorhinal cortices, which are two of the earliest structures affected in AD, serve as major gateways to mediating neocortical connections to the hippocampus (Braak & Del Tredici, 2015a). Given the pertinent involvements of those structures in memory formation, several studies have investigated and found strong associations between memory performance and the entorhinal and transentorhinal cortices, which provided discriminatory utility between people with and without AD (Kulason et al., 2018; Wolk et al., 2017; Yushkevich, Pluta, et al., 2015). In contrast to such findings, aside from the right entorhinal cortex, the statistical models employed in this study were unable to find stronger roles for these two structures in memory formation or diagnostic classification. One limiting factor is the wide range of entorhinal volumes and transentorhinal volumes across both manual and automated measures. The entorhinal volumes ranged from a difference of 13% -31% about the mean and the transentorhinal volumes had an even wider range of 21% to 45% about the mean across the entire samples. This variability would have attenuated the ability to detect meaningful differences in volumes across different diagnostic groups and their relation to memory and diagnostic groups.

Perhaps a counterargument for these findings - or rather the lack of significant findings - could be that most studies used entorhinal and transentorhinal cortical thickness instead of volumes to detect changes. Nevertheless, it has been found that performance in diagnostic separability was comparable between thickness and volumetric measures, and volume-based measures was generally more reliable across different automated segmentation platforms (Schwarz et al., 2016). Moreover, it is important to consider that early structural changes in these structures in isolation might not necessarily equate to AD-related neurodegeneration. Cortical thinning in the medial temporal regions is not an atypical finding

in clinically normal people and was observed even in people with very low risk of developing AD (Fjell et al., 2014).

A large autopsy study that examined the brains of 2,332 people between the ages of 1 – 100 years showed that approximately 70% of people in their 60s and 70s ($n = 822$) had lesions consistent with Braak NFT stage I-II, and approximately 50% – 20% of people in their 80s and 90s ($n = 1054$) were classified NFT stage I-II (Braak et al., 2011). Therefore, pathological changes in the entorhinal and transentorhinal cortices appear to be very prevalent phenomena in the older “normal” population, and not specific to AD on its own, and arguably only a proportion would progress onto the symptomatic NFT stages III and beyond.

Lateralisation and Asymmetry of the Hippocampus

One common finding across the literature is that the right hippocampus is typically larger than the left hippocampus in healthy adults (Fraser et al., 2018; Guadalupe et al., 2017; Schmidt et al., 2018), and the current results concurred with those previous researches. An interesting point that was observed in the literature regarding this left-right hippocampal asymmetry was that the degree of lateralisation or asymmetry differed in people with or without AD or when they are in different stages of AD (Ardekani et al., 2019; Muller et al., 2005; Sarica et al., 2018; Shi et al., 2009; Yue et al., 2018).

Even though this study observed a pattern of lateralisation that was consistent with the findings of the meta-analysis by Shi et al. (2009), in which the most significant left-lateralised atrophy was seen in the MCI sample, this finding is likely to be a product of self-selection bias due to the differential symptoms associated with the different hemispheres. In accord with much of the literature and the current results, the left hippocampal volume was predictive of verbal and overall memory functioning while the right MTL structures were more predictive of visual memory. Considering that daily functioning is largely driven by the

exchange of language-based information, individuals with more significant atrophy in the left hippocampus are more likely to detect changes in their verbal memory and present themselves to a memory clinic. Conversely, individuals with changes in their right MTL structures, which could include the right hippocampus, entorhinal and transentorhinal cortices, and the associated deficits in visuospatial memory, would be likely to present themselves at a later point in time.

Another line of reasoning could be that visuospatial function is not as lateralised as verbal functions. Functional MRI studies have demonstrated that although verbal function was more localised to the left hemisphere, visuospatial function tended to involve both the left and right hemispheres (Baumann et al., 2010; Dalton et al., 2016; Hartley et al., 2003; Maguire et al., 1998; Rosazza et al., 2009). Similarly, lesional studies have shown comparatively less profound visuospatial memory deficits in right MTL lesions in contrast to verbal memory impairments seen in left MTL lesions (Gleißner et al., 1998; Jones-Gotman, 1986; Jones-Gotman et al., 1997; Squire & Butters, 1992). Furthermore, the findings from functional network-based analyses have demonstrated that the left hemisphere has a greater preference for within hemisphere interactions, whereas the right hemisphere has interactions that are more bilateral (Doron, Bassett, & Gazzaniga, 2012; Gotts et al., 2013). All the results from these studies suggest that when early atrophy occurs primarily on the right MTL, some of the right hemispheric functions could be compensated by an uncompromised or less compromised left MTL, while the reverse would be less likely to occur.

Although a left-lateralised pattern of atrophy in AD has been observed in some studies (Thompson et al., 2003), this pattern of atrophy was not seen in the AD sample in this study. This lack of left-lateralised atrophy in the AD sample is not particularly unexpected, as histological investigations have provided limited evidence for preferential AD involvement of either hemispheres. However, those histopathological studies did find some differences in

NFT densities between hemispheres, albeit with extremely varied degrees of asymmetry (Arnold et al., 1991; Braak & Del Tredici, 2015a; Janota & Mountjoy, 1988; Moossy et al., 1989; Wilcock & Esiri, 1987). An NFT staging study on 39 samples by Stefanits et al. (2012) reported that 15% of the samples demonstrated intra-individual hemispheric differences in Braak staging, with five cases showing one score difference and one case showing two score difference. Comparing between the samples in the present study, the AD and MCI samples tended to have more asymmetrical hippocampal volumes than the healthy controls albeit the wide range of asymmetry and considerable overlap with the typical degree of asymmetry found in healthy controls limited the usefulness of asymmetry measures in differentiating between diagnostic groups. Putting it altogether, asymmetry between hippocampal volumes provide grounds for suspicion of pathology. However, it is necessary to bear in mind that pathologically significant asymmetry is an inconsistent phenomenon, and therefore asymmetry measures may prove to have little practical utility given the degree of variability.

Limitations and Strengths of Study

One of the main limitations in this study is that it was a cross-sectional investigation. Given the non-direct relationship between hippocampal volume, and possibly other MTL structures, longitudinal studies that determine the rates of volume change in these structures may offer much greater capacity for differentiating the different groups. Another associated limitation of a cross-sectional design was the ability to draw definite conclusions on whether patients in the MCI or SMC samples would eventually progress onto AD. Furthermore, the AD and MCI samples sizes were arguably quite small, and the routine clinical nature of this utilised study data limits the direct incorporation of additional amyloid-beta or tau biomarker measures, a procedure more expected when recruitment is into research studies, to increase confidence of an individual's diagnosis. Nonetheless, these groups of patients were very well-characterised via a routine but comprehensive clinical workup, which included a neurological

examination, neuroimaging, and comprehensive cognitive assessment. Moreover, both the precision of manual tracing and the convenience of automated procedures were used to create a reasonably thorough view of the potential contribution of volumetric measures in a real-world clinical diagnostic process.

A potential confounding variable that was not fully adjusted for was total intracranial volume (TIV). Adjustments for TIV is not a straightforward process, with no universal agreement on the methods of volume adjustments for TIV (Schwarz et al., 2016). Furthermore, estimated TIV produced by FreeSurfer has been found to have significant positive correlations with total brain volume (Klasson, Olsson, Eckerström, Malmgren, & Wallin, 2018). The significant correlation suggests that TIV in the AD sample may have been underestimated, since global atrophy would have occurred by that stage and may reduce the robustness of the MTL volumes. In contrast, some studies have found little to no advantages in TIV normalisation for the classification performance of AD and other diagnostic groups (Voevodskaya et al., 2014; Zhou et al., 2014). Hence, given the ambiguous findings of past research, the surrogate controlling for head size using gender as a control variable should have mitigated the formal lack of normalisation for TIV on the interpretation of the findings.

Another potential confounder was that patient classification was formulated in part from the memory performance and qualitative evaluation of the neuroimaging scans. In that regard, the use of the cognitive scores and structural measures as predictive variables are likely to have created a circular diagnostic process. However, of reassurance against this potential intrinsic bias was that the classification prior to comprehensive cognitive testing and neuroimaging was highly similar to the diagnosis after the full clinical workup (Kendall's tau-b, $t_b = .96$, $p < .01$). Although not presented in this study, the statistical analysis using the prior diagnosis led to similar results and interpretation. Therefore, the classification of patients in this study was based on the diagnosis following the full workup.

Conclusion and Future Directions

In summary, this study provided similar results when utilising manual and automated measures of the three selected MTL structures: the hippocampus, the entorhinal cortex and the transentorhinal cortex. Verbal and general memory performance was found to be associated with the left hippocampal volume while visual memory dysfunction might provide earlier hints of pathological changes given its association with the right entorhinal cortex. Although the left and right hippocampal volumes have unique and separate contributions to the classification of the different diagnostic samples, they had limited contribution to diagnosis beyond that which comprehensive memory testing provided. Furthermore, hippocampal asymmetry provided limited additional capability in discriminating AD and MCI samples from those people without AD pathology. By and large, volumetric measures of the medial structures offered little classification utility above that of cognitive testing.

In this study, the use of the entorhinal and transentorhinal cortical volumes, regardless of the methods of volume estimation, appeared to have limited utility in isolation as surrogate markers of brain pathology in cross-sectional studies. However, considering the crucial involvement of these structures early in the AD process, it would be of great interest for future investigations to consider the efficacy of a combination of other biomarkers in addition to volumetric measures. Future research should also place emphasis on longitudinal volumetric changes, given the weak linear relationship between MTL volumes and memory performance and the heterogeneity of a clinical population

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to 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), 270-279. doi:10.1016/j.jalz.2011.03.008
- Alzheimer's Association. (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 14(3), 367-425. doi:10.1016/j.jalz.2018.02.001
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Apostolova, L. G., Morra, J. H., Green, A. E., Hwang, K. S., Avedissian, C., Woo, E., . . . Thompson, P. M. (2010). Automated 3D mapping of baseline and 12-month associations between three verbal memory measures and hippocampal atrophy in 490 ADNI subjects. *Neuroimage*, 51(1), 488-499. doi:10.1016/j.neuroimage.2009.12.125
- Ardekani, B. A., Hadid, S. A., Blessing, E., & Bachman, A. H. (2019). Sexual dimorphism and hemispheric asymmetry of hippocampal volumetric integrity in normal aging and Alzheimer disease. *American Journal of Neuroradiology*, 40(2), 276-282. doi:10.3174/ajnr.A5943
- Arnold, S. E., Hyman, B. T., Flory, J., Damasio, A. R., & Van Hoesen, G. W. (1991). The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cerebral Cortex*, 1(1), 103-116. doi:10.1093/cercor/1.1.103
- Atiya, M., Hyman, B. T., Albert, M. S., & Killiany, R. (2003). Structural magnetic resonance imaging in established and prodromal Alzheimer disease: a review. *Alzheimer*

Disease and Associated Disorders, 17(3), 177-195. doi:10.1097/00002093-200307000-00010

Augustinack, J. C., Huber, K. E., Stevens, A. A., Roy, M., Frosch, M. P., van der Kouwe, A. J., . . . Alzheimer's Disease Neuroimaging, I. (2013). Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. *Neuroimage*, 64, 32-42. doi:10.1016/j.neuroimage.2012.08.071

Auning, E., Selnes, P., Grambaite, R., Saltyte Benth, J., Haram, A., Lovli Stav, A., . . . Aarsland, D. (2015). Neurobiological correlates of depressive symptoms in people with subjective and mild cognitive impairment. *Acta Psychiatrica Scandinavica*, 131(2), 139-147. doi:10.1111/acps.12352

Basso, M., Yang, J., Warren, L., MacAvoy, M. G., Varma, P., Bronen, R. A., & van Dyck, C. H. (2006). Volumetry of amygdala and hippocampus and memory performance in Alzheimer's disease. *Psychiatry Research*, 146(3), 251-261. doi:10.1016/j.psychresns.2006.01.007

Baumann, O., Chan, E., & Mattingley, J. B. (2010). Dissociable neural circuits for encoding and retrieval of object locations during active navigation in humans. *Neuroimage*, 49(3), 2816-2825. doi:10.1016/j.neuroimage.2009.10.021

Bigler, E. D. (2015). Structural image analysis of the brain in neuropsychology using magnetic resonance imaging (MRI) techniques. *Neuropsychology Review*, 25(3), 224-249. doi:10.1007/s11065-015-9290-0

Bonner-Jackson, A., Mahmoud, S., Miller, J., & Banks, S. J. (2015). Verbal and non-verbal memory and hippocampal volumes in a memory clinic population. *Alzheimer's Research and Therapy*, 7(1), 61. doi:10.1186/s13195-015-0147-9

Bottino, C. M., Castro, C. C., Gomes, R. L., Buchpiguel, C. A., Marchetti, R. L., & Neto, M. R. (2002). Volumetric MRI measurements can differentiate Alzheimer's disease, mild

- cognitive impairment, and normal aging. *International Psychogeriatrics*, 14(1), 59-72. doi:10.1017/s1041610202008281
- Braak, H., Alafuzoff, I., Arzberger, T., Kretschmar, H. A., & Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathologica*, 112(4), 389-404. doi:10.1007/s00401-006-0127-z
- Braak, H., & Braak, E. (1985). On areas of transition between entorhinal allocortex and temporal isocortex in the human brain. Normal morphology and lamina-specific pathology in Alzheimer's disease. *Acta Neuropathologica*, 68(4), 325-332. doi:10.1007/bf00690836
- Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4), 239-259. doi:10.1007/bf00308809
- Braak, H., & Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging*, 16(3), 271-278. doi:10.1016/0197-4580(95)00021-6
- Braak, H., & Braak, E. (1997). Frequency of stages of Alzheimer-related lesions in different age categories. *Neurobiology of Aging*, 18(4), 351-357. doi:10.1016/s0197-4580(97)00056-0
- Braak, H., Braak, E., & Bohl, J. (1993). Staging of Alzheimer-related cortical destruction. *European Neurology*, 33(6), 403-408. doi:10.1159/000116984
- Braak, H., Braak, E., Yilmazer, D., & Bohl, J. (1996). Functional anatomy of human hippocampal formation and related structures. *Journal of Child Neurology*, 11(4), 265-275. doi:10.1177/088307389601100402
- Braak, H., & Del Tredici, K. (2015a). *Neuroanatomy and pathology of sporadic Alzheimer's Disease*: Springer.

- Braak, H., & Del Tredici, K. (2015b). The preclinical phase of the pathological process underlying sporadic Alzheimer's disease. *Brain*, 138(Pt 10), 2814-2833.
doi:10.1093/brain/awv236
- Braak, H., Thal, D. R., Ghebremedhin, E., & Del Tredici, K. (2011). Stages of the pathologic process in Alzheimer disease: Age categories From 1 to 100 years. *Journal of Neuropathology & Experimental Neurology*, 70(11), 960.
doi:10.1097/NEN.0b013e318232a379
- Buckley, R., Saling, M. M., Ames, D., Rowe, C. C., Lautenschlager, N. T., Macaulay, S. L., . . . Ellis, K. A. (2013). Factors affecting subjective memory complaints in the AIBL aging study: Biomarkers, memory, affect, and age. *International Psychogeriatrics*, 25(8), 1307-1315. doi:10.1017/s1041610213000665
- Burggren, A. C., Renner, B., Jones, M., Donix, M., Suthana, N. A., Martin-Harris, L., . . . Bookheimer, S. Y. (2011). Thickness in entorhinal and subicular cortex predicts episodic memory decline in mild cognitive impairment. *International Journal of Alzheimer's Disease*, 2011, 956053. doi:10.4061/2011/956053
- Bussiere, T., Giannakopoulos, P., Bouras, C., Perl, D. P., Morrison, J. H., & Hof, P. R. (2003). Progressive degeneration of nonphosphorylated neurofilament protein-enriched pyramidal neurons predicts cognitive impairment in Alzheimer's disease: stereologic analysis of prefrontal cortex area 9. *The Journal of comparative neurology*, 463(3), 281-302. doi:10.1002/cne.10760
- Carstairs, J. R., Shores, E. A., & Myors, B. (2012). Australian norms and retest data for the Rey Auditory and Verbal Learning Test. *Australian Psychologist*, 47(4), 191-197.
doi:10.1111/j.1742-9544.2012.00086.x

- Chen, K. H., Chuah, L. Y., Sim, S. K., & Chee, M. W. (2010). Hippocampal region-specific contributions to memory performance in normal elderly. *Brain and Cognition*, 72(3), 400-407. doi:10.1016/j.bandc.2009.11.007
- Chen, X. Q., & Mobley, W. C. (2019). Alzheimer disease pathogenesis: Insights from molecular and cellular biology studies of oligomeric Abeta and tau species. *Frontiers in Neuroscience*, 13, 659. doi:10.3389/fnins.2019.00659
- Chi, H., Chang, H.-Y., & Sang, T.-K. (2018). Neuronal cell death mechanisms in major neurodegenerative diseases. *International Journal of Molecular Sciences*, 19(10), 3082. doi:10.3390/ijms19103082
- Cho, H., Choi, J. Y., Hwang, M. S., Lee, J. H., Kim, Y. J., Lee, H. M., . . . Lee, M. S. (2016). Tau PET in Alzheimer disease and mild cognitive impairment. *Neurology*, 87(4), 375-383. doi:10.1212/wnl.0000000000002892
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 7. doi:10.1037/1040-3590.6.4.284
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.): L. Erlbaum Associates.
- Cohen, J., Cohen, P., West, S., & Aiken, L. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.): Routledge.
- Cook, M. J., Fish, D. R., Shorvon, S. D., Straughan, K., & Stevens, J. M. (1992). Hippocampal volumetric and morphometric studies in frontal and temporal and temporal-lobe epilepsy. *Brain*, 115, 1001-1015. doi:10.1093/brain/115.4.1001
- Dalton, M. A., Hornberger, M., & Piguet, O. (2016). Material specific lateralization of medial temporal lobe function: An fMRI investigation. *Human Brain Mapping*, 37(3), 933-941. doi:10.1002/hbm.23077

- Delbeuck, X., Van der Linden, M., & Collette, F. (2003). Alzheimer's disease as a disconnection syndrome? *Neuropsychology Review*, 13(2), 79-92.
doi:10.1023/a:1023832305702
- Deng, L., Stanley, M. L., Monge, Z. A., Wing, E. A., Geib, B. R., Davis, S. W., & Cabeza, R. (2020). Age-related compensatory reconfiguration of PFC connections during Episodic memory retrieval. *Cerebral Cortex*. doi:10.1093/cercor/bhaa192
- Derflinger, S., Sorg, C., Gaser, C., Myers, N., Arsic, M., Kurz, A., . . . Muhlau, M. (2011). Grey-matter atrophy in Alzheimer's disease is asymmetric but not lateralized. *Journal of Alzheimer's Disease*, 25(2), 347-357. doi:10.3233/jad-2011-110041
- Dickerson, B. C., Feczko, E., Augustinack, J. C., Pacheco, J., Morris, J. C., Fischl, B., & Buckner, R. L. (2009). Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of Aging*, 30(3), 432-440. doi:10.1016/j.neurobiolaging.2007.07.022
- Dolcos, F., Rice, H. J., & Cabeza, R. (2002). Hemispheric asymmetry and aging: right hemisphere decline or asymmetry reduction. *Neuroscience & Biobehavioral Reviews*, 26(7), 819-825. doi:10.1016/s0149-7634(02)00068-4
- Donix, M., Burggren, A. C., Scharf, M., Marschner, K., Suthana, N. A., Siddarth, P., . . . Bookheimer, S. Y. (2013). APOE associated hemispheric asymmetry of entorhinal cortical thickness in aging and Alzheimer's disease. *Psychiatry Research*, 214(3), 212-220. doi:10.1016/j.psychresns.2013.09.006
- Doron, K. W., Bassett, D. S., & Gazzaniga, M. S. (2012). Dynamic network structure of interhemispheric coordination. *Proceedings of the National Academy of Sciences of the United States of America*, 109(46), 18661-18668. doi:10.1073/pnas.1216402109
- Duara, R., Barker, W., Loewenstein, D., Greig, M. T., Rodriguez, R., Goryawala, M., . . . Adjouadi, M. (2015). Insights into cognitive aging and Alzheimer's disease using

- amyloid PET and structural MRI scans. *Clinical and Translational Imaging*, 3(1), 65-74. doi:10.1007/s40336-015-0110-6
- Duara, R., Loewenstein, D. A., Potter, E., Appel, J., Greig, M. T., Urs, R., . . . Potter, H. (2008). Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease. *Neurology*, 71(24), 1986-1992. doi:10.1212/01.wnl.0000336925.79704.9f
- Ezzati, A., Katz, M. J., Zammit, A. R., Lipton, M. L., Zimmerman, M. E., Sliwinski, M. J., & Lipton, R. B. (2016). Differential association of left and right hippocampal volumes with verbal episodic and spatial memory in older adults. *Neuropsychologia*, 93(Pt B), 380-385. doi:10.1016/j.neuropsychologia.2016.08.016
- Fan, L. Y., Lai, Y. M., Chen, T. F., Hsu, Y. C., Chen, P. Y., Huang, K. Z., . . . Chiu, M. J. (2018). Diminution of context association memory structure in subjects with subjective cognitive decline. *Human Brain Mapping*, 39(6), 2549-2562. doi:10.1002/hbm.24022
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. doi:10.3758/bf03193146
- Fischl, B. (2012). FreeSurfer. *Neuroimage*, 62(2), 774-781. doi:10.1016/j.neuroimage.2012.01.021
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, 97(20), 11050-11055. doi:10.1073/pnas.200033797
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341-355. doi:10.1016/s0896-6273(02)00569-x

- Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Segonne, F., Quinn, B. T., & Dale, A. M. (2004). Sequence-independent segmentation of magnetic resonance images. *Neuroimage*, 23, S69-S84. doi:10.1016/j.neuroimage.2004.07.016
- Fischl, B., Stevens, A. A., Rajendran, N., Yeo, B. T., Greve, D. N., Van Leemput, K., . . . Augustinack, J. C. (2009). Predicting the location of entorhinal cortex from MRI. *Neuroimage*, 47(1), 8-17. doi:10.1016/j.neuroimage.2009.04.033
- Fjell, A. M., McEvoy, L., Holland, D., Dale, A. M., Walhovd, K. B., & Alzheimer's Disease Neuroimaging, I. (2014). What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Progress in Neurobiology*, 117, 20-40. doi:10.1016/j.pneurobio.2014.02.004
- Fraser, M. A., Shaw, M. E., Anstey, K. J., & Cherbuin, N. (2018). Longitudinal assessment of hippocampal atrophy in midlife and early old age: Contrasting manual tracing and semi-automated segmentation (FreeSurfer). *Brain Topography*, 31(6), 949-962. doi:10.1007/s10548-018-0659-2
- Fung, Y. L., Ng, K. E. T., Vogrin, S. J., Meade, C., Ngo, M., Collins, S. J., & Bowden, S. C. (2019, March 14). Labelling of Statistical Output of Cytoarchitecture-Defined Variables in FreeSurfer Version 6. Retrieved from <https://www.j-alz.com/content/labelling-statistical-output-cytoarchitecture>
- Giannakopoulos, P., Herrmann, F., Md, M. P. H., Bussiere, T., Bouras, C., Kovari, E., . . . Hof, P. (2003). Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology*, 60(9), 1495-1500. doi:10.1212/01.WNL.0000063311.58879.01
- Gleißner, U., Helmstaedter, C., & Elger, C. E. (1998). Right hippocampal contribution to visual memory: a presurgical and postsurgical study in patients with temporal lobe

- epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 65(5), 665-669.
doi:10.1136/jnnp.65.5.665
- Gomez-Isla, T., Hollister, R., West, H., Mui, S., Growdon, J. H., Petersen, R. C., . . . Hyman, B. T. (1997). Neuronal loss correlates with but exceeds neurofibrillary tangles in Alzheimer's disease. *Annals of Neurology*, 41(1), 17-24. doi:10.1002/ana.410410106
- Gomez-Isla, T., Price, J. L., McKeel, D. W., Jr., Morris, J. C., Growdon, J. H., & Hyman, B. T. (1996). Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *Journal of Neuroscience*, 16(14), 4491-4500.
doi:10.1523/JNEUROSCI.16-14-04491.1996
- Good, C. D., Scahill, R. I., Fox, N. C., Ashburner, J., Friston, K. J., Chan, D., . . . Frackowiak, R. S. J. (2002). Automatic Differentiation of Anatomical Patterns in the Human Brain: Validation with Studies of Degenerative Dementias. *Neuroimage*, 17(1), 29-46. doi:10.1006/nimg.2002.1202
- Gotts, S. J., Jo, H. J., Wallace, G. L., Saad, Z. S., Cox, R. W., & Martin, A. (2013). Two distinct forms of functional lateralization in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 110(36), E3435-E3444. doi:10.1073/pnas.1302581110
- Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Reviews Neuroscience*, 13(7), 491-505. doi:10.1038/nrn3256
- Guadalupe, T., Mathias, S. R., vanErp, T. G. M., Whelan, C. D., Zwiers, M. P., Abe, Y., . . . Francks, C. (2017). Human subcortical brain asymmetries in 15,847 people worldwide reveal effects of age and sex. *Brain Imaging and Behavior*, 11(5), 1497-1514. doi:10.1007/s11682-016-9629-z
- Guadalupe, T., Zwiers, M. P., Teumer, A., Wittfeld, K., Vasquez, A. A., Hoogman, M., . . . Francks, C. (2014). Measurement and genetics of human subcortical and hippocampal

- asymmetries in large datasets. *Human Brain Mapping*, 35(7), 3277.
doi:10.1002/hbm.22401
- Hartley, T., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The well-worn route and the path less traveled: Distinct neural bases of route following and wayfinding in humans. *Neuron*, 37(5), 877-888. doi:10.1016/S0896-6273(03)00095-3
- Holdnack, J. A., Xiaobin, Z., Larrabee, G. J., Millis, S. R., & Salthouse, T. A. (2011). Confirmatory factor analysis of the WAIS-IV/WMS-IV. *Assessment*, 18(2), 178-191.
doi:10.1177/1073191110393106
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed. / David W. Hosmer, Stanley Lemeshow, Rodney X. Sturdivant. ed.): Wiley.
- Hyman, B. T., Van Hoesen, G. W., & Damasio, A. R. (1990). Memory-related neural systems in Alzheimer's disease: an anatomic study. *Neurology*, 40(11), 1721-1730.
doi:10.1212/wnl.40.11.1721
- Hyman, B. T., Van Hoesen, G. W., Damasio, A. R., & Barnes, C. L. (1984). Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. *Science*, 225(4667), 1168-1170. doi:10.1126/science.6474172
- Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., . . . Alzheimer's Disease Neuroimaging, I. (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.
doi:10.1016/j.neuroimage.2015.04.042
- Iqbal, K., Liu, F., Gong, C. X., Alonso Adel, C., & Grundke-Iqbal, I. (2009). Mechanisms of tau-induced neurodegeneration. *Acta Neuropathologica*, 118(1), 53-69.
doi:10.1007/s00401-009-0486-3

- Jack, C. R., Jr., Albert, M. S., Knopman, D. S., McKhann, G. M., Sperling, R. A., Carrillo, M. C., . . . Phelps, C. H. (2011). Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 257-262.
doi:10.1016/j.jalz.2011.03.004
- Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., . . . Sperling, R. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), 535-562.
doi:10.1016/j.jalz.2018.02.018
- Jack, C. R., Jr., Petersen, R. C., Xu, Y. C., Waring, S. C., O'Brien, P. C., Tangalos, E. G., . . . Kokmen, E. (1997). Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*, 49(3), 786-794. doi:10.1212/wnl.49.3.786
- Janota, I., & Mountjoy, C. Q. (1988). Asymmetry of pathology in Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, 51(7), 1011-1012.
doi:10.1136/jnnp.51.7.1011-c
- Jhoo, J. H., Lee, D. Y., Choo, I. H., Seo, E. H., Oh, J. S., Lee, J. S., . . . Woo, J. I. (2010). Discrimination of normal aging, MCI and AD with multimodal imaging measures on the medial temporal lobe. *Psychiatry Research*, 183(3), 237-243.
doi:10.1016/j.psychresns.2010.03.006
- Jones-Gotman, M. (1986). Memory for designs: The hippocampal contribution. *Neuropsychologia*, 24(2), 193-203. doi:10.1016/0028-3932(86)90052-7
- Jones-Gotman, M., Zatorre, R. J., Olivier, A., Andermann, F., Cendes, F., Staunton, H., . . . Wieser, H. G. (1997). Learning and retention of words and designs following excision from medial or lateral temporal-lobe structures. *Neuropsychologia*, 35(7), 963-973.
doi:10.1016/S0028-3932(97)00024-9

- Kawas, C. H., Corrada, M. M., Brookmeyer, R., Morrison, A., Resnick, S. M., Zonderman, A. B., & Arenberg, D. (2003). Visual memory predicts Alzheimer's disease more than a decade before diagnosis. *Neurology*, 60(7), 1089-1093.
doi:10.1212/01.wnl.0000055813.36504.bf
- Kent, P. (2013). The evolution of the Wechsler Memory Scale: A selective review. *Applied Neuropsychology: Adult*, 20(4), 277-291. doi:10.1080/09084282.2012.689267
- Kessels, R. P., Rijken, S., Joosten-Weyn Banningh, L. W., Van Schuylenborgh, V. A. N. E. N., & Olde Rikkert, M. G. (2010). Categorical spatial memory in patients with mild cognitive impairment and Alzheimer dementia: positional versus object-location recall. *Journal of the International Neuropsychological Society*, 16(1), 200-204.
doi:10.1017/s1355617709990944
- Kivisaari, S. L., Probst, A., & Taylor, K. I. (2013). The perirhinal, entorhinal, and parahippocampal cortices and hippocampus: An overview of functional anatomy and protocol for their segmentation in MR Images. In *fMRI : Basics and Clinical Applications* (pp. 239).
- Klasson, N., Olsson, E., Eckerström, C., Malmgren, H., & Wallin, A. (2018). Estimated intracranial volume from FreeSurfer is biased by total brain volume. *European Radiology Experimental*, 2(1), 24. doi:10.1186/s41747-018-0055-4
- Kreutzer, J. S., DeLuca, J., & Caplan, B. (2018). *Encyclopedia of Clinical Neuropsychology* (Second edition. ed.): Springer International Publishing.
- Kulason, S., Tward, D. J., Brown, T., Sicat, C. S., Liu, C. F., Ratnanather, J. T., . . . Alzheimer's Disease Neuroimaging, I. (2018). Cortical thickness atrophy in the transentorhinal cortex in mild cognitive impairment. *NeuroImage: Clinical*, 21, 101617. doi:10.1016/j.nicl.2018.101617

- Lavenex, P., & Amaral, D. G. (2000). Hippocampal-neocortical interaction: a hierarchy of associativity. *Hippocampus*, 10(4), 420-430. doi:10.1002/1098-1063(2000)10:4<420::Aid-hipo8>3.0.Co;2-5
- Leuzy, A., Chiotis, K., Lemoine, L., Gillberg, P.-G., Almkvist, O., Rodriguez-Vieitez, E., & Nordberg, A. (2019). Tau PET imaging in neurodegenerative tauopathies—still a challenge. *Molecular Psychiatry*, 24(8), 1112-1134. doi:10.1038/s41380-018-0342-8
- Long, X., Jiang, C., & Zhang, L. (2018). Morphological biomarker differentiating MCI converters from nonconverters: Longitudinal evidence based on hemispheric asymmetry. *Behavioural Neurology*, 2018, 3954101. doi:10.1155/2018/3954101
- Maass, A., Landau, S., Baker, S. L., Horng, A., Lockhart, S. N., La Joie, R., . . . Jagust, W. J. (2017). Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. *Neuroimage*, 157, 448-463. doi:10.1016/j.neuroimage.2017.05.058
- Maass, A., Lockhart, S. N., Harrison, T. M., Bell, R. K., Mellinger, T., Swinnerton, K., . . . Jagust, W. J. (2018). Entorhinal tau pathology, episodic memory decline, and neurodegeneration in aging. *Journal of Neuroscience*, 38(3), 530-543. doi:10.1523/JNEUROSCI.2028-17.2017
- Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D., & O'Keefe, J. (1998). Knowing Where and Getting There: A Human Navigation Network. *Science*, 280(5365), 921-924. doi:10.1126/science.280.5365.921
- Márquez, F., & Yassa, M. A. (2019). Neuroimaging Biomarkers for Alzheimer's Disease. *Molecular Neurodegeneration*, 14(1), 21-21. doi:10.1186/s13024-019-0325-5
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46. doi:10.1037/1082-989x.1.1.30

- McKhann, G. M., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7), 939-944. doi:10.1212/wnl.34.7.939
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to 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), 263-269. doi:10.1016/j.jalz.2011.03.005
- Miller, D. I., Davidson, P. S. R., Schindler, D., & Messier, C. (2013). Confirmatory factor analysis of the WAIS-IV and WMS-IV in older adults. *Journal of Psychoeducational Assessment*, 31(4), 375-390. doi:10.1177/0734282912467961
- Min, J., Moon, W. J., Jeon, J. Y., Choi, J. W., Moon, Y. S., & Han, S. H. (2017). Diagnostic efficacy of structural MRI in patients with mild-to-moderate Alzheimer disease: Automated volumetric assessment versus visual assessment. *American Journal of Roentgenology*, 208(3), 617-623. doi:10.2214/ajr.16.16894
- Minkova, L., Habich, A., Peter, J., Kaller, C. P., Eickhoff, S. B., & Kloppel, S. (2017). Gray matter asymmetries in aging and neurodegeneration: A review and meta-analysis. *Human Brain Mapping*, 38(12), 5890-5904. doi:10.1002/hbm.23772
- Mitchell, A. J., Beaumont, H., Ferguson, D., Yadegarfar, M., & Stubbs, B. (2014). Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: meta-analysis. *Acta Psychiatrica Scandinavica*, 130(6), 439-451. doi:10.1111/acps.12336
- Mizuno, K., Wakai, M., Takeda, A., & Sobue, G. (2000). Medial temporal atrophy and memory impairment in early stage of Alzheimer's disease: an MRI volumetric and

- memory assessment study. *Journal of the Neurological Sciences*, 173(1), 18-24.
doi:10.1016/s0022-510x(99)00289-0
- Moossy, J., Zubenko, G. S., Martinez, A. J., Rao, G. R., Kopp, U., & Hanin, I. (1989).
Lateralization of brain morphologic and cholinergic abnormalities in Alzheimer's
disease. *Archives of Neurology*, 46(6), 639-642.
doi:10.1001/archneur.1989.00520420059023
- Mortimer, J. A., Gosche, K. M., Riley, K. P., Markesbery, W. R., & Snowdon, D. A. (2004).
Delayed recall, hippocampal volume and Alzheimer neuropathology: Findings from
the Nun Study. *Neurology*, 62(3), 428-432. doi:10.1212/01.wnl.0000106463.66966.65
- Muller, M. J., Greverus, D., Dellani, P. R., Weibrich, C., Wille, P. R., Scheurich, A., . . .
Fellgiebel, A. (2005). Functional implications of hippocampal volume and diffusivity
in mild cognitive impairment. *Neuroimage*, 28(4), 1033-1042.
doi:10.1016/j.neuroimage.2005.06.029
- Nair, B., Mansfield, E., & Waller, A. (2016). A race against time: The dementia epidemic.
Archives of Medicine and Health Sciences, 4(1), 127-134. doi:10.4103/2321-
4848.183363
- Nelson, P. T., Alafuzoff, I., Bigio, E. H., Bouras, C., Braak, H., Cairns, N. J., . . . Beach, T.
G. (2012). Correlation of Alzheimer disease neuropathologic changes with cognitive
status: a review of the literature. *Journal of Neuropathology & Experimental
Neurology*, 71(5), 362-381. doi:10.1097/NEN.0b013e31825018f7
- Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., Hanninen, T., Laakso, M.
P., . . . Soininen, H. (2004). Hippocampus and entorhinal cortex in mild cognitive
impairment and early AD. *Neurobiology of Aging*, 25(3), 303-310.
doi:10.1016/s0197-4580(03)00084-8

- Perrotin, A., La Joie, R., de La Sayette, V., Barre, L., Mezenge, F., Mutlu, J., . . . Chetelat, G. (2017). Subjective cognitive decline in cognitively normal elders from the community or from a memory clinic: Differential affective and imaging correlates. *Alzheimers & Dementia*, 13(5), 550-560. doi:10.1016/j.jalz.2016.08.011
- Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., . . . Jack, C. R., Jr. (2009). Mild cognitive impairment: ten years later. *Archives of Neurology*, 66(12), 1447-1455. doi:10.1001/archneurol.2009.266
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment - Clinical characterization and outcome. *Archives of Neurology*, 56(3), 303-308. doi:10.1001/archneur.56.3.303
- Portney, L. G., & Watkins, M. P. (2013). *Foundations of clinical research: Applications to practice* (3rd ed.). Harlow: Pearson.
- Price, L. R., Tulsky, D., Millis, S., & Weiss, L. (2002). Redefining the factor structure of the Wechsler Memory Scale-III: confirmatory factor analysis with cross-validation. *Journal of Clinical and Experimental Neuropsychology*, 24(5), 574-585. doi:10.1076/jcen.24.5.574.1013
- Reitz, C., Brickman, A. M., Brown, T. R., Manly, J., DeCarli, C., Small, S. A., & Mayeux, R. (2009). Linking hippocampal structure and function to memory performance in an aging population. *Archives of Neurology*, 66(11), 1385-1392. doi:10.1001/archneurol.2009.214
- Rosazza, C., Minati, L., Ghielmetti, F., Maccagnano, E., Erbetta, A., Villani, F., . . . Bruzzone, M. G. (2009). Engagement of the medial temporal lobe in verbal and nonverbal memory: Assessment with functional MR imaging in healthy subjects. *American Journal of Neuroradiology*, 30(6), 1134-1141. doi:10.3174/ajnr.A1518

- Sarica, A., Vasta, R., Novellino, F., Vaccaro, M. G., Cerasa, A., Quattrone, A., & Alzheimer's Disease Neuroimaging, I. (2018). MRI asymmetry index of hippocampal subfields increases through the continuum from the mild cognitive impairment to the Alzheimer's disease. *Frontiers in Neuroscience*, 12, 576. doi:10.3389/fnins.2018.00576
- Schmidt, M. F., Storrs, J. M., Freeman, K. B., Jack, C. R., Jr., Turner, S. T., Griswold, M. E., & Mosley, T. H., Jr. (2018). A comparison of manual tracing and FreeSurfer for estimating hippocampal volume over the adult lifespan. *Human Brain Mapping*, 39(6), 2500-2513. doi:10.1002/hbm.24017
- Schwarz, C. G., Gunter, J. L., Wiste, H. J., Przybelski, S. A., Weigand, S. D., Ward, C. P., . . . Alzheimer's Disease Neuroimaging, I. (2016). A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *NeuroImage: Clinical*, 11, 802-812. doi:10.1016/j.nicl.2016.05.017
- Shi, F., Liu, B., Zhou, Y., Yu, C., & Jiang, T. (2009). Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer's disease: Meta-analyses of MRI studies. *Hippocampus*, 19(11), 1055-1064. doi:doi:10.1002/hipo.20573
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations - uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428. doi:10.1037/0033-2909.86.2.420
- Sperling, R., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., . . . Phelps, C. H. (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. doi:10.1016/j.jalz.2011.03.003
- Squire, L. R., & Butters, N. (1992). *Neuropsychology of memory* (2nd ed. ed.): Guilford Press.

- Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, 253(5026), 1380. doi:10.1126/science.1896849
- Stefanits, H., Budka, H., & Kovacs, G. G. (2012). Asymmetry of neurodegenerative disease-related pathologies: a cautionary note. *Acta Neuropathologica*, 123(3), 449-452. doi:10.1007/s00401-011-0936-6
- Stokin, G. B., & Goldstein, L. S. B. (2006). Axonal Transport and Alzheimer's Disease. *Annual Review of Biochemistry*, 75(1), 607-627. doi:10.1146/annurev.biochem.75.103004.142637
- Stoub, T. R., Rogalski, E. J., Leurgans, S., Bennett, D. A., & deToledo-Morrell, L. (2010). Rate of entorhinal and hippocampal atrophy in incipient and mild AD: Relation to memory function. *Neurobiology of Aging*, 31(7), 1089-1098. doi:10.1016/j.neurobiolaging.2008.08.003
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests : administration, norms, and commentary* (3rd ed. ed.): Oxford University Press.
- Tapiola, T., Pennanen, C., Tapiola, M., Tervo, S., Kivipelto, M., Hänninen, T., . . . Soininen, H. (2008). MRI of hippocampus and entorhinal cortex in mild cognitive impairment: A follow-up study. *Neurobiology of Aging*, 29(1), 31-38. doi:10.1016/j.neurobiolaging.2006.09.007
- Theofilas, P., Ehrenberg, A. J., Nguy, A., Thackrey, J. M., Dunlop, S., Mejia, M. B., . . . Grinberg, L. T. (2018). Probing the correlation of neuronal loss, neurofibrillary tangles, and cell death markers across the Alzheimer's disease Braak stages: a quantitative study in humans. *Neurobiology of Aging*, 61, 1-12. doi:10.1016/j.neurobiolaging.2017.09.007

- Thompson, P. M., Hayashi, K. M., de Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., . . . Toga, A. W. (2003). Dynamics of gray matter loss in Alzheimer's disease. *Journal of Neuroscience*, 23(3), 994-1005. doi:10.1523/JNEUROSCI.23-03-00994.2003
- Tulsky, D. S., & Price, L. R. (2003). The joint WAIS-III and WMS-III factor structure: development and cross-validation of a six-factor model of cognitive functioning. *Psychological Assessment*, 15(2), 149-162. doi:10.1037/1040-3590.15.2.149
- Tward, D. J., Sicat, C. S., Brown, T., Bakker, A., Gallagher, M., Albert, M., & Miller, M. (2017). Entorhinal and transentorhinal atrophy in mild cognitive impairment using longitudinal diffeomorphometry. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 9, 41-50. doi:10.1016/j.dadm.2017.07.005
- Van Hoesen, G. W., Hyman, B. T., & Damasio, A. R. (1991). Entorhinal cortex pathology in Alzheimer's disease. *Hippocampus*, 1(1), 1-8. doi:10.1002/hipo.450010102
- Van Petten, C. (2004). Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. *Neuropsychologia*, 42(10), 1394-1413. doi:10.1016/j.neuropsychologia.2004.04.006
- Velayudhan, L., Proitsi, P., Westman, E., Muehlboeck, J. S., Mecocci, P., Vellas, B., . . . AddNeuroMed, C. (2013). Entorhinal cortex thickness predicts cognitive decline in Alzheimer's disease. *Journal of Alzheimer's Disease*, 33(3), 755-766. doi:10.3233/jad-2012-121408
- Venneri, A., Gorgoglione, G., Toraci, C., Nocetti, L., Panzetti, P., & Nichelli, P. (2011). Combining neuropsychological and structural neuroimaging indicators of conversion to Alzheimer's disease in amnesic mild cognitive impairment. *Current Alzheimer Research*, 8(7), 789-797. doi:10.2174/156720511797633160
- Visser, P. J., Verhey, F. R., Hofman, P. A., Scheltens, P., & Jolles, J. (2002). Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive

- impairment. *Journal of Neurology, Neurosurgery and Psychiatry*, 72(4), 491-497.
doi:10.1136/jnnp.72.4.491
- Voevodskaya, O., Simmons, A., Nordenskjold, R., Kullberg, J., Ahlstrom, H., Lind, L., . . . Alzheimer's Dis, N. (2014). The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Frontiers in Aging Neuroscience*, 6, 264. doi:10.3389/fnagi.2014.00264
- Vyhnalek, M., Nikolai, T., Andel, R., Nedelska, Z., Rubinova, E., Markova, H., . . . Hort, J. (2014). Neuropsychological correlates of hippocampal atrophy in memory testing in nondemented older adults. *Journal of Alzheimer's Disease*, 42 Suppl 3, S81-90.
doi:10.3233/jad-132642
- Wachinger, C., Salat, D. H., Weiner, M., Reuter, M., & Alzheimer's Disease Neuroimaging, I. (2016). Whole-brain analysis reveals increased neuroanatomical asymmetries in dementia for hippocampus and amygdala. *Brain*, 139(Pt 12), 3253-3266.
doi:10.1093/brain/aww243
- Wang, L., Benzinger, T. L., Su, Y., Christensen, J., Friedrichsen, K., Aldea, P., . . . Ances, B. M. (2016). Evaluation of tau imaging in staging Alzheimer disease and revealing interactions between beta-amyloid and tauopathy. *Jama Neurology*, 73(9), 1070-1077.
doi:10.1001/jamaneurol.2016.2078
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., . . . Alzheimer's Disease Neuroimaging, I. (2015). 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's & Dementia*, 11(6), e1-120. doi:10.1016/j.jalz.2014.11.001
- Weschler, D. (2009). *WMS-IV technical and interpretive manual*. San Antonio, Tex: Pearson.
- Wilcock, G. K., & Esiri, M. M. (1987). Asymmetry of pathology in Alzheimer's disease. *J Neurol Neurosurg Psychiatry*, 50(10), 1384-1386. doi:10.1136/jnnp.50.10.1384

- Wilke, M., de Haan, B., Juenger, H., & Karnath, H. O. (2011). Manual, semi-automated, and automated delineation of chronic brain lesions: a comparison of methods. *Neuroimage*, 56(4), 2038-2046. doi:10.1016/j.neuroimage.2011.04.014
- Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., . . . Petersen, R. C. (2004). Mild cognitive impairment – beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, 256(3), 240-246. doi:10.1111/j.1365-2796.2004.01380.x
- Wirth, M., Madison, C. M., Rabinovici, G. D., Oh, H., Landau, S. M., & Jagust, W. J. (2013). Alzheimer's disease neurodegenerative biomarkers are associated with decreased cognitive function but not beta-amyloid in cognitively normal older individuals. *The Journal of Neuroscience*, 33(13), 5553-5563. doi:10.1523/jneurosci.4409-12.2013
- Wirths, O., & Bayer, T. A. (2010). Neuron loss in transgenic mouse models of Alzheimer's disease. *International Journal of Alzheimer's Disease*, 2010. doi:10.4061/2010/723782
- Wolk, D. A., Das, S. R., Mueller, S. G., Weiner, M. W., Yushkevich, P. A., & Alzheimer's Disease Neuroimaging, I. (2017). Medial temporal lobe subregional morphometry using high resolution MRI in Alzheimer's disease. *Neurobiology of Aging*, 49, 204-213. doi:10.1016/j.neurobiolaging.2016.09.011
- Xie, L., Das, S. R., Wisse, L. E. M., Ittyerah, R., Yushkevich, P. A., Wolk, D. A., & Alzheimer's Dis Neuroimaging, I. (2018). Early tau burden correlates with higher rate of atrophy in transentorhinal cortex. *Journal of Alzheimer's Disease*, 62(1), 85-92. doi:10.3233/jad-170945
- Yue, L., Wang, T., Wang, J., Li, G., Wang, J., Li, X., . . . Xiao, S. (2018). Asymmetry of hippocampus and amygdala defect in subjective cognitive decline among the

community dwelling Chinese. *Frontiers in Psychiatry*, 9, 226.

doi:10.3389/fpsy.2018.00226

Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S. L., Gertje, E. C., . . . Wolk, D. A.

(2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment.

Human Brain Mapping, 36(1), 258-287. doi:10.1002/hbm.22627

Zhao, W., Wang, X., Yin, C., He, M., Li, S., & Han, Y. (2019). Trajectories of the

hippocampal subfields atrophy in the Alzheimer's disease: A structural imaging study.

Frontiers in Neuroinformatics, 13, 13. doi:10.3389/fninf.2019.00013

Zhou, Q., Goryawala, M., Cabrerizo, M., Barker, W., Duara, R., & Adjouadi, M. (2014).

Significance of normalization on anatomical MRI measures in predicting Alzheimer's disease. *Scientific World Journal*, 2014, 541802. doi:10.1155/2014/541802

APPENDIX B

Additional Statistical Outputs

Table B1

Multinomial Logistic Regression Analysis Examining the Relationship Between the General Memory and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b					
	χ^2	df	P-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value
General Memory	91.827	3	<.001	AD	-0.268	0.765	0.693	$\pm$ 0.845	<.001
				MCI	-0.173	0.841	0.784	$\pm$ 0.903	<.001
				SMC	-0.084	0.920	0.881	$\pm$ 0.959	<.001

^a $\chi^2(3, N = 100) = 191.83$, Nagelkerke $R^2 = .65$, $p < .001$; Goodness of fit is not significant

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score.

β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B2

Multinomial Logistic Regression Analysis Examining the Relationship Between the General Memory and Scanner Type, With and Without Healthy Control

Predictor	Predictor Model with Healthy Control (N = 100) ^a			Predictor Model without Healthy Control (n = 70) ^b		
	χ^2	df	p-value	χ^2	df	p-value
General Memory	77.90	3	<.001	40.83	2	<.001
Scanner Type	26.08	3	<.001	1.62	2	.444

^a $\chi^2(6, N = 100) = 117.91$, Nagelkerke $R^2 = .75$, $p < .001$; Goodness of fit is not significant.

^b $\chi^2(4, N = 70) = 41.84$, Nagelkerke $R^2 = .58$, $p < .001$; Goodness of fit is not significant

Table B3

Multinomial Logistic Regression Analysis Examining the Relationship Between the General Memory, Manual Volumetric Estimates and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	χ^2	df	P-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value	
General Memory	67.705	3	<.001	AD	-0.348	0.706	0.594	$\pm$ 0.839	<.001	
				MCI	-0.178	0.837	0.770	$\pm$ 0.909	<.001	
				SMC	-0.095	0.909	0.864	$\pm$ 0.956	<.001	
Left hippocampus	12.986	3	.005	AD	0.016	1.016	0.956	$\pm$ 1.080	.612	
				MCI	-0.050	0.951	0.910	$\pm$ 0.993	.023	
				SMC	-0.033	0.968	0.938	$\pm$ 0.998	.035	
Right hippocampus	11.320	3	.01	AD	-0.043	0.958	0.895	$\pm$ 1.026	.219	
				MCI	0.028	1.028	0.986	$\pm$ 1.072	.189	
				SMC	0.028	1.028	0.962	$\pm$ 1.058	.061	
Left entorhinal	1.922	3	.59	AD	-0.076	0.927	0.756	$\pm$ 1.137	.467	
				MCI	0.035	1.036	0.933	$\pm$ 1.149	.509	
				SMC	0.022	1.022	0.962	$\pm$ 1.086	.480	
Right entorhinal	0.247	3	.97	AD	-0.028	0.972	0.793	$\pm$ 1.191	.785	
				MCI	0.011	1.011	0.901	$\pm$ 1.134	.851	
				SMC	0.008	1.008	0.948	$\pm$ 1.073	.791	
Left transentorhinal	5.690	3	.13	AD	0.152	1.164	0.928	$\pm$ 1.460	.188	
				MCI	0.006	1.006	0.887	$\pm$ 1.140	.926	
				SMC	-0.051	0.950	0.878	$\pm$ 1.028	.205	
Right transentorhinal	2.499	3	.48	AD	-0.068	0.934	0.756	$\pm$ 1.154	.526	
				MCI	0.012	1.012	0.894	$\pm$ 1.145	.850	
				SMC	0.043	1.044	0.972	$\pm$ 1.122	.235	

^a $\chi^2(21, N = 100) = 118.73$, Nagelkerke $R^2 = .75$, $p < .001$; Goodness of fit was not significant.

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Each unit increase for each medial temporal structure is 10 mm³. The p -values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B4

Multinomial Logistic Regression Analysis Examining the Relationship Between the General Memory, FreeSurfer Volumetric Estimates and Classification in the Clinical or Healthy Control Sample(N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	χ^2	df	p-value	Group	β	Odds Ratio	Odds ratio 95% CI		p-value	
General Memory	72.075	3	<.001	AD	-0.475	0.622	0.485	± 0.798	<.001	
				MCI	-0.180	0.853	0.770	± 0.907	<.001	
				SMC	-0.089	0.915	0.874	± 0.958	<.001	
Left hippocampus	5.818	3	.12	AD	0.027	1.027	0.957	± 1.103	.461	
				MCI	-0.048	0.953	0.904	± 1.004	.071	
				SMC	-0.004	0.413	0.963	± 1.029	.804	
Right hippocampus	12.037	3	.007	AD	-0.169	0.845	0.734	± 0.973	.019	
				MCI	0.003	1.003	0.949	± 1.059	.928	
				SMC	-0.002	0.996	0.962	± 1.037	.933	
Left entorhinal	1.241	3	.74	AD	0.060	1.062	0.939	± 1.201	.336	
				MCI	0.013	1.014	0.938	± 1.096	.735	
				SMC	-0.003	0.998	0.949	± 1.047	.900	
Right entorhinal	6.262	3	.10	AD	0.218	1.243	1.000	± 1.545	.050	
				MCI	0.047	1.048	0.972	± 1.130	.221	
				SMC	0.003	0.997	0.960	± 1.048	.893	
Left transentorhinal	4.603	3	.20	AD	-0.076	0.927	0.839	± 1.024	.135	
				MCI	-0.020	0.980	0.928	± 1.035	.462	
				SMC	-0.028	0.973	0.942	± 1.005	.097	
Right transentorhinal	7.417	3	.06	AD	-0.154	0.857	0.696	± 1.056	.148	
				MCI	0.013	1.013	0.930	± 1.104	.764	
				SMC	0.041	1.041	0.990	± 1.095	.114	

^a $\chi^2(21, N = 100) = 123.59$, Nagelkerke $R^2 = .77$, $p < .001$; Goodness of fit was not significant.

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Each unit increase for each medial temporal structure is 10 mm³. The p -values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B5

Multinomial Logistic Regression Analysis Examining the Relationship Between General Memory, Sum of Left and Right Hippocampal Volumes Based on Manual Estimates and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	χ^2	df	P-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value	
General Memory	71.37	3	< .001	AD	-0.252	0.777	0.704	±	0.858	< .001
				MCI	-0.162	0.850	0.792	±	0.913	< .001
				SMC	-0.084	0.919	0.881	±	0.959	< .001
Sum of left and right hippocampal volumes	2.65	3	.449	AD	-0.001	0.999	0.998	±	1.000	.148
				MCI	-0.001	0.999	0.999	±	1.000	.296
				SMC	0.000	1.000	0.999	±	1.001	.781

^a $\chi^2(6, N = 100) = 94.48$, Nagelkerke $R^2 = .66$, $p < .001$; Goodness of fit test was not significant

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Each unit change for sum of left and right hippocampal volume is 10mm³. The *p*-values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B6

Multinomial Logistic Regression Analysis Examining the Relationship Between General Memory, Lateralised Difference Volume Between Left and Right Hippocampus Based on Manual Estimates and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	χ^2	df	P-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value	
General Memory	91.87	3	< .001	AD	-0.314	0.731	0.644	±	0.829	< .001
				MCI	-0.169	0.844	0.784	±	0.909	< .001
				SMC	-0.087	0.916	0.875	±	0.960	< .001
Lateralisation (raw volume)	13.65	3	.003	AD	0.020	1.021	0.973	±	1.070	.398
				MCI	-0.037	0.963	0.928	±	1.000	.047
				SMC	-0.027	0.974	0.948	±	1.001	.055

^a $\chi^2(6, N = 100) = 105.48$, Nagelkerke $R^2 = .71$, $p < .001$; Goodness of fit test was not significant

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Lateralisation (raw volume) is calculated using left – right hippocampal volume. Each unit change in lateralisation (raw volume) is 10mm³. The *p*-values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B7

Multinomial Logistic Regression Analysis Examining the Relationship Between the General Memory, Asymmetry Between Left and Right Hippocampal Volume Based on Manual Estimates and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	χ^2	df	p-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value	
General Memory	83.47	3	< .001	AD	-0.264	0.768	0.695	± 0.849	< .001	
				MCI	-0.168	0.845	0.785	± 0.909	< .001	
				SMC	-0.081	0.922	0.883	± 0.963	< .001	
Asymmetry (raw volume)	3.36	3	.339	AD	0.038	1.039	0.983	± 1.098	.174	
				MCI	0.042	1.043	0.995	± 1.094	.079	
				SMC	0.021	1.021	0.984	± 1.059	.266	

^a $\chi^2(6, N = 100) = 95.18$, Nagelkerke $R^2 = .67$, $p < .001$; Goodness of fit test was not significant

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Asymmetry (raw volume) is calculated based on the absolute difference between the left and right hippocampal volume. Each unit difference in asymmetry (raw volume) is 10mm³. The *p*-values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B8

Multinomial Logistic Regression Analysis Examining the Relationship Between General Memory, Hippocampal Lateralisation Index Based on Manual Estimates and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	X ²	df	p-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value	
General Memory	89.85	3	< .001	AD	-0.304	0.768	0.654	± 0.833	< .001	
				MCI	-0.168	0.845	0.785	± 0.911	< .001	
				SMC	-0.088	0.916	0.874	± 0.959	< .001	
Lateralisation index	12.22	3	.007	AD	0.053	1.004	0.784	± 1.417	.174	
				MCI	-0.256	0.774	0.604	± 0.991	.042	
				SMC	-0.187	0.830	0.685	± 1.004	.055	

^a $\chi^2(6, N = 100) = 104.043$, Nagelkerke $R^2 = .70$, $p < .001$; Goodness of fit was not significant.

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Lateralisation index is calculated using (left - right hippocampal volume)/(left + right hippocampal volume)*100. Each unit difference in lateralisation index is 1%. The *p*-values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B9

Multinomial Logistic Regression Analysis Examining the Relationship Between General Memory, Hippocampal Asymmetry Index Based on Manual Estimates and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	χ^2	df	p-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value	
General Memory	80.18	3	< .001	AD	-0.262	0.770	0.697	± 0.850	< .001	
				MCI	-0.165	0.848	0.788	± 0.912	< .001	
				SMC	-0.081	0.922	0.883	± 0.963	< .001	
Asymmetry index	4.19	3	.242	AD	0.293	1.340	0.929	± 1.934	.117	
				MCI	0.330	1.392	0.997	± 1.942	.052	
				SMC	0.162	1.176	0.904	± 1.529	.227	

^a $\chi^2(6, N = 100) = 96.02$, Nagelkerke $R^2 = .67$, $p < .001$; Goodness of fit was not significant.

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Hippocampal asymmetry index is the absolute value of (left - right volume)/(left + right hippocampal volume)*100. Each unit difference in asymmetry index is 1%. The *p*-values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B10

Multinomial Logistic Regression Analysis Examining the Relationship Between the General Memory, Sum of Left and Right Hippocampal Volumes Based on FreeSurfer Estimates and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	χ^2	df	p-value	Group	β	Odds Ratio	Odds ratio 95% CI		p-value	
General Memory	65.48	3	<.001	AD	-0.249	0.780	0.705	± 0.862	<.001	
				MCI	-0.161	0.851	0.793	± 0.915	<.001	
				SMC	-0.084	0.919	0.880	± 0.959	<.001	
Sum of left and right hippocampal volumes	4.48	3	.214	AD	-0.016	0.984	0.966	± 1.002	.075	
				MCI	-0.013	0.987	0.973	± 1.001	.074	
				SMC	-0.004	0.996	0.987	± 1.006	.432	

^a $\chi^2(6, N = 100) = 96.31$, Nagelkerke $R^2 = .61$, $p < .001$; Goodness of fit was not significant.

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Each unit increase for sum of left and right hippocampal volume is 10mm³. The *p*-values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B11

Multinomial Logistic Regression Analysis Examining the Relationship Between General, Lateralised Difference between the Left and Right Hippocampal Volume Based on FreeSurfer Estimations and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b					
	χ^2	df	p-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value
General Memory	91.73	3	<.001	AD	-0.249	0.736	0.651	$\pm$ 0.833	<.001
				MCI	-0.161	0.847	0.788	$\pm$ 0.910	<.001
				SMC	-0.084	0.919	0.880	$\pm$ 0.959	<.001
Lateralisation (raw volume)	9.10	3	.028	AD	-0.016	1.027	0.985	$\pm$ 1.071	.212
				MCI	-0.013	0.968	0.928	$\pm$ 1.009	.122
				SMC	-0.007	0.993	0.964	$\pm$ 1.023	.645

^a $\chi^2(6, N = 100) = 100.93$, Nagelkerke $R^2 = .69$, $p < .001$; Goodness of fit was not significant.

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Lateralisation (raw volume) is calculated using left – right hippocampal volume. Each unit change in lateralisation (raw volume) is 10mm³. The *p*-values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B12

Multinomial Logistic Regression Analysis Examining the Relationship Between General Memory, Asymmetry Between the Left and Right Hippocampal Volume Based on FreeSurfer Estimations and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b					
	χ^2	df	p-value	Group	β	Odds Ratio	Odds Ratio 95% CI		p-value
General Memory	84.51	3	<.001	AD	-0.276	0.759	0.683	$\pm$ 0.842	<.001
				MCI	-0.179	0.836	0.774	$\pm$ 0.904	<.001
				SMC	-0.085	0.918	0.879	$\pm$ 0.959	<.001
Asymmetry (raw volume)	7.51	3	.060	AD	0.038	1.039	0.998	$\pm$ 1.010	.228
				MCI	0.033	10.033	0.998	$\pm$ 1.009	.226
				SMC	-0.022	0.979	0.993	$\pm$ 1.003	.380

^a $\chi^2(6, N = 100) = 99.33$, Nagelkerke $R^2 = .68$, $p < .001$; Goodness of fit was not significant.

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Asymmetry (raw volume) is calculated based on the absolute difference between the left and right hippocampal volume. Each unit difference in asymmetry (raw volume) is 10mm³. The *p*-values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B13

Multinomial Logistic Regression Analysis Examining the Relationship Between General Memory, Hippocampal Lateralisation Index Based on FreeSurfer Estimates and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	χ^2	df	<i>p</i> -value	Group	β	Odds Ratio	Odds Ratio 95% CI		<i>p</i> -value	
General Memory	90.73	3	<.001	AD	-0.301	0.740	0.657	±	0.835	<.001
				MCI	-0.165	0.848	0.789	±	0.911	<.001
				SMC	-0.085	0.919	0.880	±	0.959	<.001
Lateralisation index	7.88	3	.049	AD	0.101	1.106	0.884	±	1.384	.378
				MCI	-0.159	0.853	0.692	±	1.052	.136
				SMC	-0.040	0.960	0.819	±	1.126	.619

^a $\chi^2(6, N = 100) = 99.70$, Nagelkerke $R^2 = .69$, $p < .001$; Goodness of fit was not significant.

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score.

Lateralisation index is calculated using (left - right hippocampal volume)/(left + right hippocampal volume)*100.

Each unit difference in lateralisation index is 1%. The p -values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Table B14

Multinomial Logistic Regression Analysis Examining the Relationship Between General Memory, Hippocampal Asymmetry Index Based on FreeSurfer Estimates and Classification in the Clinical or Healthy Control Sample (N = 100)

Predictor	Predictor Model ^a			Coefficient Estimates ^b						
	χ^2	df	<i>p</i> -value	Group	β	Odds Ratio	Odds Ratio 95% CI		<i>p</i> -value	
General Memory	80.51	3	<.001	AD	-0.272	0.762	0.687	± 0.844	<.001	
				MCI	-0.177	0.838	0.775	± 0.906	<.001	
				SMC	-0.084	0.919	0.880	± 0.959	<.001	
Asymmetry index	8.16	3	.043	AD	0.220	1.246	0.897	± 1.731	.190	
				MCI	0.231	1.260	0.948	± 1.675	.111	
				SMC	-0.075	0.927	0.715	+ 1.204	.571	

^a $\chi^2(6, N = 100) = 99.99$, Nagelkerke $R^2 = .67$, $p < .001$; Goodness of fit was not significant.

^b The reference group is the healthy controls. Each unit increase for *General Memory* is 1 standard score. Asymmetry index is calculated using the absolute value of (left - right hippocampal volume)/(left + right hippocampal volume)*100. Each unit difference in asymmetry index is 1%. The p -values presented are uncorrected for multiple comparisons. β : regression coefficient, Odds ratio: exponential of the regression coefficient.

Chapter 7

General Discussion

The previous chapter marks the last investigation of this thesis, which addressed the contribution of volumetric measures to diagnosis of mild cognitive impairment and Alzheimer's disease. This final chapter in the thesis summarises and brings together the conclusions from Chapters 1 to 6 and highlights the relationship of the findings to each individual chapter as well as to the general literature. This chapter also discusses the overall limitations to the methodology of the investigations and how it might impact the interpretation and applicability of the data in the larger context.

General Discussion

This thesis was conceptualised to examine and validate the current methods of estimating medial temporal structures and to investigate the correlates between these medial temporal structures and memory function. The medial temporal structures of interest were the hippocampus, the entorhinal cortex and the transentorhinal cortex as these structures are implicated neuropathologically early in the development of Alzheimer's disease (AD) and could therefore play a crucial role in improving time to diagnosis (Braak & Braak, 1991).

One of the core clinical criteria for diagnosing mild cognitive impairment (MCI) and dementia related to AD includes impairment in one or more cognitive domains on cognitive testing (Albert et al., 2011; McKhann et al., 2011). However, converging evidence from years of research suggests that AD development is an extended process, which involves a long, cognitively asymptomatic phase but with neuropathological changes detectable years before one would arrive at a clinical diagnosis of MCI or AD (Dubois, Hampel, et al., 2016; Sperling et al., 2011).

Providing an accurate diagnosis of MCI in its earliest stages and identifying the transition of MCI to clinical AD can be a difficult process (McKhann et al., 2011; Sperling et al., 2011). Although there is currently no definite link between a particular fluid or neuroimaging biomarker in asymptomatic individuals with consequent clinical symptomology, research supports the benefits of incorporating biomarkers for the detection and monitoring of MCI (Albert et al., 2011; Bayram et al., 2018; Dubois et al., 2010; McKhann et al., 2011; Sperling & Johnson, 2013).

As described in Chapter 1, there is no one ideal diagnostic tool, neither cognitive measures nor fluid and neuroimaging biomarker, for the identification of AD, especially in the MCI stages. In general, cognitive measures showed very good sensitivities and specificities in the identification of AD-related dementia. However, as reported in Chapter 1,

a review of the diagnostic test accuracies of cognitive assessments revealed a poor balance between the sensitivity and specificity values for the detection of MCI, with diagnostic cognitive measures showing a pattern of low sensitivity but high specificity for MCI. Diagnostic test accuracies of biomarkers, on the other hand, showed a pattern of high sensitivity but low specificity in the detection of people with MCI. Given this pattern of diagnostic accuracy statistics, it strongly suggests that a combination of diagnostic cognitive assessment, and fluid and neuroimaging biomarker testing would provide better MCI discrimination than a singular diagnostic approach. In fact, many studies have already shown that a multimodal approach improves the diagnostic accuracies of discriminating between AD, MCI, and healthy controls (Bauer et al., 2018; Gross et al., 2012; Trzepacz et al., 2014).

Structural Variations of the Collateral Sulcus

One of the key biomarkers that has been included as supportive evidence for a diagnosis of MCI and AD is the presence of neuronal degeneration on structural magnetic resonance imaging (MRI), as indicated by disproportionate atrophy in the medial temporal structures (Albert et al., 2011; McKhann et al., 2011). The entorhinal and transentorhinal cortices are of particular interest as these structures have been shown to be affected much earlier than the hippocampus in AD, and a review of the literature has shown close associations between the structural atrophy of these areas and AD progression (Kulason et al., 2018; Tward et al., 2017; Weiner et al., 2015).

However, as reviewed in Chapter 2, the localisation of the entorhinal and transentorhinal cortices has been complicated by the morphological variability of the collateral sulcus. The collateral sulcus, which is a major anatomical landmark for the defining the borders of the entorhinal and transentorhinal cortices, have been found to have several patterns of cortical folding, discontinuity and bifurcation along the entire length (Chau et al.,

2014; Cikla et al., 2016; Ding & Van Hoesen, 2010; Huntgeburth & Petrides, 2012; Ono et al., 1990; Ovalioglu et al., 2018; Pruessner et al., 2002; Taylor & Probst, 2008).

Chapter 2 described the variability of the entire length of the collateral sulcus, in which the collateral sulcus extends from the anterior pole of the temporal lobe to the posterior occipital cortex. In order to better understand the patterns of variability along the length of the entorhinal and transentorhinal cortices, the first study (Chapter 2) investigated the morphology of the collateral sulcus specifically along that length of interest. The results showed that 62.5% of the cases showed a single, continuous collateral sulcus along the length of the entorhinal cortex and transentorhinal cortex. In the cases of a non-continuous collateral sulcus, interruptions could occur due to (i) the presence of side branches, (ii) sulcal mergers within the different sections of the collateral sulcus, (iii) the presence of double collateral sulcus, or (iv) connections to the occipitotemporal sulcus. Furthermore, as described in the findings in Chapter 2, the patterns of variability were also different between the entire length of the collateral sulcus and the length of the collateral sulcus where the entorhinal and transentorhinal cortices were located, which emphasises the fact that the collateral sulcus is not a static anatomical landmark.

Utility of FreeSurfer in a Clinical Setting

The introduction of automated neuroimaging analysis packages such as the popular open-source software, FreeSurfer, has facilitated the quantification of structural changes using automated volumetric estimations (Fischl, 2012). The automation of structural neuroimaging analysis is an especially promising advancement as it presents an opportunity for the integration of quantitative volumetric analysis into routine clinical dementia workup (Bigler, 2015). However, the complexity of structures such as the collateral sulcus, as seen in Chapter 2 and described in previous studies, brings into question whether automated processing pipelines could adequately account for the variability and provide valid and

reliable volumetric measures for the different medial temporal structures, especially the rhinal cortices.

The validation studies on FreeSurfer (v6.0) conducted through Chapters 3 to 5, presented with some promising and some suboptimal results. Firstly, the well-established hippocampal protocol in FreeSurfer (v6.0) showed promising results, as seen in Chapter 3 and 5. Automated hippocampal volume estimations presented with excellent single-measure consistency intraclass correlation coefficient and good agreement with manual volume estimations in both clinical and control groups. Chapter 5 also showed that the results were similar regardless of the scanner field strength used to acquire the scans, or the diagnostic status of the patient, or the brain hemisphere.

On the other hand, results seen in Chapter 3 and 4, with regards to the performance of new *ex vivo* automated protocol for the volume estimation of the entorhinal and transentorhinal cortices, were much less precise. In both Chapters 3 and 4, single-measure consistency intraclass correlation coefficient between automated and manual automated entorhinal volume were generally well below .50, which indicated poor reliability (Portney & Watkins, 2013). As seen in Chapter 4, the single-measure consistency intraclass correlation coefficient achieved between automated and manual transentorhinal cortical volume estimation volumes were in the negative range, which was statistically equivalent to an intraclass correlation coefficient of zero (James et al., 1984; LeBreton & Senter, 2008). The limits of agreement between automated and manual volume estimations were also very large for both structures (Figure 4.5, Figure 4.6 and Figure 4.7), which makes the precise interpretation of these automated volumes a difficult task.

The findings from the hippocampal volume estimations studies using the newest FreeSurfer protocol lend support to the converging evidence of validity of automated hippocampal volume estimations (Clerx et al., 2015; Schmidt et al., 2018). More importantly,

the scans were obtained using routine clinical acquisition parameters that included patients who have been diagnosed with MCI or AD. The excellent results demonstrate that the hippocampal automated segmentation protocol found in FreeSurfer is robust against variable scan quality and parameters, as well as the variability in hippocampal structure that is likely to be found in the clinical population.

Although the encouraging results for FreeSurfer hippocampal automated segmentation suggest a strong potential for translational into clinical practice, the results from the entorhinal and transentorhinal automated volume estimation presented a very different picture. Automated measurement of these temporal structures has been known to be particularly challenging as it is a common area for imaging artefacts (Desikan et al., 2010; Desikan et al., 2006; McCarthy et al., 2015). Unlike the hippocampus that is larger and well-defined by the alveus, the proximity of the smaller entorhinal and transentorhinal cortices to the tentorium cerebelli and the adjacent cortex makes identification of the cortical borders more challenging (Augustinack et al., 2013; Fischl et al., 2009; Franko et al., 2014; Yushkevich, Pluta, et al., 2015). In addition, the variability of the collateral sulcus, as described in the findings from Chapter 2 as well as in other studies reviewed in Chapter 2, adds another layer of complexity to the development of a fully automated segmentation. The results show that currently, the complexity of the collateral sulcus and the impact on localising the neighbouring structures, could not be readily resolved in the current FreeSurfer software (Augustinack et al., 2013; Wolk et al., 2017; Xie et al., 2017; Yushkevich, Pluta, et al., 2015).

Contribution of Volumetric Measures to Diagnosing Alzheimer's Disease

Besides the availability of MRI facilities as a routine clinical tool, one of the compelling reasons for including structural MRI as a neuroimaging biomarker is the close relationship between the characteristic medial temporal atrophy with cognitive dysfunction

throughout the clinical phases of AD (Sperling et al., 2011). The close relationship between medial temporal atrophy and disease severity has been documented in numerous studies, and many studies have also suggested that the presence of medial temporal atrophy early in the disease could serve as good predictors of future progression (Burggren et al., 2011; Hyman et al., 1984; Kulason et al., 2018; Maass et al., 2018; Pennanen et al., 2004; Reitz et al., 2009; Stoub et al., 2010; Tapiola et al., 2008; Tward et al., 2017; Van Hoesen et al., 1991; Velayudhan et al., 2013; Venneri et al., 2011; Xie et al., 2018).

The group comparison results from Chapter 6 supported previous findings, in which the left hippocampus was found to be consistently smaller in both the MCI and AD samples, compared to healthy controls and clinical patients who were diagnosed with subjective memory concerns. The right hippocampus and right entorhinal cortex were also found to be significantly smaller in the AD group than in the healthy controls or people with subjective memory concerns. Patients with subjective memory concerns are an interesting clinical population as some studies have noted atrophic changes in the medial temporal structures in these patients, hence suggesting that these patients represent an at-risk group for future cognitive decline (Yue et al., 2018; Zhao et al., 2019). However, the findings from Chapter 6 demonstrated that the medial temporal volumes between the healthy controls and people with subjective memory concerns were not significantly different from each other.

Typically, patients would present themselves at a memory clinic due to memory inefficiencies observed by themselves or others. Looking back at the review in Chapter 1, the ability of memory measures to detect people with true MCI is good but less than ideal, with sensitivity values that were generally more than .70 but less than .80. However, structural measures could provide another avenue for evaluation to the causes of these self-reported memory inefficiencies. The results from Chapter 6 suggests that in the cases where there is no objective memory impairment on testing and there are no detectable neuropathological

changes on the MRI, it is less likely to be a missed MCI diagnosis. The functional memory impairment is likely driven by other non-organic factors that has be further considered to form a clearer clinical picture. Such a conclusion would also be consistent with the opinions of the authors involved in the Alzheimer's Biomarkers Standardization Initiative, who expressed that biological biomarkers would serve as good tools to rule out an AD-related diagnosis (Molinuevo et al., 2014).

Chapter 6 was also designed to investigate the discriminatory ability of medial temporal volumes between different diagnostic groups and healthy controls. Surprisingly, although the left and right hippocampal volumes provided unique contribution to the ability to discriminate between the different samples, neither the left nor right hippocampal volumes significantly improved discrimination between the various samples. In general, besides the right entorhinal volume, the left entorhinal cortex and transentorhinal cortex provided limited diagnostic utility in this study. The improvements to classifications contributed by the addition of medial temporal volume estimations were also particularly modest in the MCI samples.

An interesting observation that has arisen from this final study in Chapter 6 is the disease-dependent relationship between memory and medial temporal structural volumes. In other words, smaller volumes may not necessarily indicate poorer memory performance in healthy individuals. Such findings have also been reported in other studies. Apostolova et al. (2010) found associations between the hippocampal volume and memory in the samples with MCI and AD but not in the healthy controls. In the same vein, Dickerson et al. (2009) also only found correlations between entorhinal and perirhinal cortices with memory performance in the AD group but not the healthy controls. The presence of a disease-dependent relationship suggests that in the cases in which there are only minor disruptions to the medial temporal structures, such as during the asymptomatic stages of AD or very early stage MCI,

individuals could effectively recruit other memory processing networks (Grady, 2012). Furthermore, it has also been proposed that cell death do not necessarily occur in the face of pathology but continues to exist in a functioning but defective state (Stokin & Goldstein, 2006; Wirths & Bayer, 2010). As the number of abnormal neurons increase with the infiltration of tau pathology, cognitive impairment becomes ultimately apparent not directly due to neuron loss, but when there is an overwhelming number of dysfunctional neurons (Braak & Del Tredici, 2015).

The finding in Chapter 6 highlights the complexity of identifying people with MCI. Medial temporal volumes had limited ability in discriminating the different diagnostic groups but at the same time, an important point of consideration is that, as mentioned above, the relationship between memory performance and volumes is not necessarily direct. Subtler changes, such as those seen in people with MCI, might not be easily detectable. As discussed in Chapter 1, the current tools available are still lacking in the ability to accurately diagnose this highly heterogeneous MCI group, and may require a combination of neuropsychological measures, and various biological biomarkers in tandem with other clinical investigations to improve detection (Albert et al., 2011; Bauer et al., 2018; Dubois, Hampel, et al., 2016; Gross et al., 2012; Klekociuk et al., 2016; Klekociuk et al., 2014; Summers & Saunders, 2012). Considering the early neuropathological involvement of these medial temporal structures early in the AD process and the accessibility of MRI in various clinical settings, it would remain of great interest to consider the combination of structural MRI with other biomarkers that could significantly improve detection of MCI in a clinical setting.

Methodological Considerations

All the studies in this thesis, Chapter 2 to Chapter 6, used a cross-sectional design, which could have limited the ability to draw any conclusion about relevant longitudinal changes that might occur. However, there is very little reason to believe there would be

longitudinal changes to sulcal patterns. Although sulcal depth and width have been found to change with age-related cortical atrophy, the sulcal folding patterns are found to remain stable over time (Cachia et al., 2016; Jin, Zhang, Shaw, Sachdev, & Cherbuin, 2018). Therefore, the findings in Chapter 2 would most likely represent the most common patterns of cortical folding of the collateral sulcus that one might expect to find across one's lifetime.

One of the main limitations of cross-sectional investigations is that it is difficult to ascertain whether a patient diagnosed with MCI had true progressive AD-related MCI, which is a limitation that might be particularly relevant to the study detailed in Chapter 6. As mentioned in the limitations of Chapter 1 and Chapter 6, MCI is not necessarily a stable longitudinal diagnosis, and the diagnosis of MCI, as a construct, is in itself highly heterogeneous (Albert et al., 2011; Klekociuk et al., 2016; Klekociuk et al., 2014; Summers & Saunders, 2012). The identification of people with MCI had been primarily centred on memory dysfunction, but a subset of people with MCI present with impairment in non-memory domains, and both amnesic and non-amnesic MCI could have impairments in a single or multiple cognitive domains (Albert et al., 2011; Petersen et al., 2009). However, divisions into the different MCI subtypes have yielded limited predictive utility (Klekociuk et al., 2014; Summers & Saunders, 2012), and determination of cognitive impairments have been undermined by statistical issues related to variability in test reliability (Klekociuk et al., 2016).

In the most recent MCI core clinical criteria detailed in Albert et al. (2011), there was no reference to the different amnesic and non-amnesic, or single and multi-domain cognitive impairment subtypes. The authors did, however, emphasise the importance of ruling out other diseases such as Lewy bodies or vascular disease to increase confidence that the underlying aetiology for cognitive impairment in MCI is most consistent with AD. The inclusion of additional biomarkers such as CSF amyloid-beta and tau measures would have

increased the level of confidence in the diagnosis that the cognitive impairment is of an AD pathology. However, from a clinical perspective there are still limitations in the current state of knowledge about the role of biomarkers in AD and Albert et al. (2011) emphasised that the overall clinical and cognitive presentation should be the main basis of a clinical diagnosis. Notably, all the clinical patient samples included in Chapters 3 to 6 underwent a thorough diagnostic workup that included evaluations from a neurologist, comprehensive neuropsychological examination, neuroimaging and other necessary clinical examinations as required for clarification. The use of composite scores obtained from comprehensive cognitive test batteries addresses the psychometric limitations of short cognitive measures (Klekociuk et al., 2016), while extensive clinical examinations provides the evidence for non-AD aetiology if present (Albert et al., 2011). Altogether, the aggregate information obtained from this evaluation process allows for the narrowing down to the most probable diagnosis of AD-related MCI.

Another limitation in these series of investigations in Chapters 3 to 6 was that there was no intentional blinding protocol implemented during cognitive assessment or MRI acquisition. However, no clinician who contributed to the cognitive assessment or the neuroimaging results knew the final clinical diagnosis that was used as the reference standard in these studies.

A noteworthy development in the neuroimaging analysis landscape is the new longitudinal processing pipeline that has been released by FreeSurfer (Reuter, Schmansky, Rosas, & Fischl, 2012). Undoubtedly, evaluation of the rates of atrophy would be a valuable line of research that this thesis was not able to pursue. However, the value of this series of cross-sectional validation studies, covered from Chapter 3 to 5, is the ascertainment of the clinical utility of the tools that could be used to evaluate a patient at the first visit. The ability to incorporate an automated tool such as FreeSurfer at first evaluation would facilitate the

provision of valuable quantitative information in terms of determining clinical diagnosis and future follow-up.

None of the samples included for analysis in the studies exceeded a size of 100, which may have led to underpowered studies. Self-selection bias for both the clinical and control samples might have also introduced cohort effects that could not be easily discerned from the results given the nature of this project was largely focussed on using routine clinical data. However, the research design used in the thesis involved methods quality representative of many published studies. In Chapter 2, only high-quality MRI scans of healthy older adults without any AD-related pathology were included. From Chapters 3 to 6, the patients included were clinically very well-characterised using methods described in the relevant chapters and summarised in the preceding paragraph. Furthermore, the memory performance of individuals included in these studies were evaluated using the full Weschler Memory Scale, in which the various memory index scores have been shown to have the best available reliability coefficients (Kreutzer et al., 2018; Weschler, 2009). Certainly, the rigour of the study would have been enhanced with multiple manual tracers, but the high inter- and intra-rater reliabilities reported in Chapters 3 to 5 provided confidence in the veracity of these manual measures that had strictly followed manual segmentation protocol. Moreover, the newest version of FreeSurfer was used throughout the studies, with manual delineation being used as the gold standard “benchmark” for comparison with the performance of the automated pipelines.

Overall Conclusion

In this thesis, volumetric validation studies of FreeSurfer (v6.0) automated segmentation pipelines, based on the structural MRI scans of healthy controls and people with AD, MCI, or subjective memory concerns, provided promising results for the hippocampal measures. However, the results were suboptimal for the entorhinal cortex and

the transentorhinal cortex. These results suggest that substantial refinement of the current algorithms in FreeSurfer (v6.0) is required before fully automated segmentations of these structures could be implemented and used in a clinical context. An important point of consideration for refinement would include the need to address the variable patterns of collateral sulcal folding into the segmentation algorithms. Although the relationship between medial temporal structural volumes, accuracy of diagnosis and memory performance is complicated, the evidence of early neuropathological involvement of these medial temporal structures support the continual roles that these structures will play into the development of a multimodal approach to diagnosing AD in the future.

References

- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to 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), 270-279. doi:10.1016/j.jalz.2011.03.008
- Apostolova, L. G., Morra, J. H., Green, A. E., Hwang, K. S., Avedissian, C., Woo, E., . . . Thompson, P. M. (2010). Automated 3D mapping of baseline and 12-month associations between three verbal memory measures and hippocampal atrophy in 490 ADNI subjects. *Neuroimage*, 51(1), 488-499. doi:10.1016/j.neuroimage.2009.12.125
- Augustinack, J. C., Huber, K. E., Stevens, A. A., Roy, M., Frosch, M. P., van der Kouwe, A. J., . . . Alzheimer's Disease Neuroimaging, I. (2013). Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. *Neuroimage*, 64, 32-42. doi:10.1016/j.neuroimage.2012.08.071
- Bauer, C. M., Cabral, H. J., & Killiany, R. J. (2018). Multimodal discrimination between normal aging, mild cognitive impairment and Alzheimer's disease and prediction of cognitive decline. *Diagnostics*, 8(1), 20. doi:10.3390/diagnostics8010014
- Bayram, E., Caldwell, J. Z. K., & Banks, S. J. (2018). Current understanding of magnetic resonance imaging biomarkers and memory in Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 4, 395-413. doi:10.1016/j.trci.2018.04.007
- Bigler, E. D. (2015). Structural image analysis of the brain in neuropsychology using magnetic resonance imaging (MRI) techniques. *Neuropsychology Review*, 25(3), 224-249. doi:10.1007/s11065-015-9290-0

- Braak, H., & Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4), 239-259. doi:10.1007/bf00308809
- Braak, H., & Del Tredici, K. (2015). The preclinical phase of the pathological process underlying sporadic Alzheimer's disease. *Brain*, 138(Pt 10), 2814-2833. doi:10.1093/brain/awv236
- Burggren, A. C., Renner, B., Jones, M., Donix, M., Suthana, N. A., Martin-Harris, L., . . . Bookheimer, S. Y. (2011). Thickness in entorhinal and subicular cortex predicts episodic memory decline in mild cognitive impairment. *International Journal of Alzheimer's Disease*, 2011, 956053. doi:10.4061/2011/956053
- Cachia, A., Borst, G., Tissier, C., Fisher, C., Plaze, M., Gay, O., . . . Raznahan, A. (2016). Longitudinal stability of the folding pattern of the anterior cingulate cortex during development. *Developmental cognitive neuroscience*, 19, 122-127. doi:10.1016/j.dcn.2016.02.011
- Chau, A. M., Stewart, F., & Gragnaniello, C. (2014). Sulcal and gyral anatomy of the basal occipital-temporal lobe. *Surgical & Radiologic Anatomy*, 36(10), 959-965. doi:10.1007/s00276-014-1294-6
- Cikla, U., Menekse, G., Quraishi, A., Neves, G., Keles, A., Liu, C., . . . Baskaya, M. K. (2016). The sulci of the inferior surface of the temporal lobe: An anatomical study. *Clinical Anatomy*, 29(7), 932-942. doi:10.1002/ca.22767
- Clerx, L., Gronenschild, E. H. B. M., Echavarri, C., FransVerhey, Aalten, P., & Jacobs, H. I. L. (2015). Can FreeSurfer compete with manual volumetric measurements in Alzheimer's disease? *Current Alzheimer Research*, 12(4), 358-367. doi:10.2174/1567205012666150324174813
- Desikan, R. S., Cabral, H. J., Settecase, F., Hess, C. P., Dillon, W. P., Glastonbury, C. M., . . . Fischl, B. (2010). Automated MRI measures predict progression to Alzheimer's

- disease. *Neurobiology of Aging*, 31(8), 1364-1374.
doi:10.1016/j.neurobiolaging.2010.04.023
- Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Killiany, R. J. (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. doi:10.1016/j.neuroimage.2006.01.021
- Dickerson, B. C., Feczko, E., Augustinack, J. C., Pacheco, J., Morris, J. C., Fischl, B., & Buckner, R. L. (2009). Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of Aging*, 30(3), 432-440. doi:10.1016/j.neurobiolaging.2007.07.022
- Ding, S. L., & Van Hoesen, G. W. (2010). Borders, extent, and topography of human perirhinal cortex as revealed using multiple modern neuroanatomical and pathological markers. *Human Brain Mapping*, 31(9), 1359-1379. doi:10.1002/hbm.20940
- Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., Dekosky, S. T., Barberger-Gateau, P., . . . Scheltens, P. (2010). Revising the definition of Alzheimer's disease: A new lexicon. *The Lancet Neurology*, 9(11), 1118-1127. doi:10.1016/s1474-4422(10)70223-4
- Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., . . . Washington Dc, U. S. A. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 12(3), 292-323. doi:10.1016/j.jalz.2016.02.002
- Fischl, B. (2012). FreeSurfer. *Neuroimage*, 62(2), 774-781. doi:10.1016/j.neuroimage.2012.01.021

- Fischl, B., Stevens, A. A., Rajendran, N., Yeo, B. T., Greve, D. N., Van Leemput, K., . . . Augustinack, J. C. (2009). Predicting the location of entorhinal cortex from MRI. *Neuroimage*, 47(1), 8-17. doi:10.1016/j.neuroimage.2009.04.033
- Franko, E., Insausti, A. M., Artacho-Perula, E., Insausti, R., & Chavoix, C. (2014). Identification of the human medial temporal lobe regions on magnetic resonance images. *Human Brain Mapping*, 35(1), 248-256. doi:10.1002/hbm.22170
- Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Reviews Neuroscience*, 13(7), 491-505. doi:10.1038/nrn3256
- Gross, A. L., Manly, J. J., Pa, J., Johnson, J. K., Park, L. Q., Mitchell, M. B., . . . McLaren, D. G. (2012). Cortical signatures of cognition and their relationship to Alzheimer's disease. *Brain Imaging and Behavior*, 6(4), 584-598. doi:10.1007/s11682-012-9180-5
- Huntgeburth, S. C., & Petrides, M. (2012). Morphological patterns of the collateral sulcus in the human brain. *European Journal of Neuroscience*, 35(8), 1295-1311. doi:10.1111/j.1460-9568.2012.08031.x
- Hyman, B. T., Van Hoesen, G. W., Damasio, A. R., & Barnes, C. L. (1984). Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. *Science*, 225(4667), 1168-1170. doi:10.1126/science.6474172
- James, R. L., Demaree, G. R., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69(1), 85-98. doi:10.1037/0021-9010.69.1.85
- Jin, K., Zhang, T., Shaw, M., Sachdev, P., & Cherbuin, N. (2018). Relationship between sulcal characteristics and brain aging. *Frontiers in Aging Neuroscience*, 10, 339-339. doi:10.3389/fnagi.2018.00339

- Klekociuk, S. Z., Saunders, N. L., & Summers, M. J. (2016). Diagnosing mild cognitive impairment as a precursor to dementia: Fact or fallacy? *Australian Psychologist*, 51(5), 366-373. doi:10.1111/ap.12178
- Klekociuk, S. Z., Summers, J. J., Vickers, J. C., & Summers, M. J. (2014). Reducing false positive diagnoses in mild cognitive impairment: The importance of comprehensive neuropsychological assessment. *European Journal of Neurology*, 21(10), 1330-+. doi:10.1111/ene.12488
- Kreutzer, J. S., DeLuca, J., & Caplan, B. (2018). *Encyclopedia of Clinical Neuropsychology* (Second edition. ed.): Springer International Publishing.
- Kulason, S., Tward, D. J., Brown, T., Sicat, C. S., Liu, C. F., Ratnanather, J. T., . . . Alzheimer's Disease Neuroimaging, I. (2018). Cortical thickness atrophy in the transentorhinal cortex in mild cognitive impairment. *NeuroImage: Clinical*, 21, 101617. doi:10.1016/j.nicl.2018.101617
- LeBreton, J., & Senter, L. J. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815-852. doi:10.1177/1094428106296642
- Maass, A., Lockhart, S. N., Harrison, T. M., Bell, R. K., Mellinger, T., Swinnerton, K., . . . Jagust, W. J. (2018). Entorhinal tau pathology, episodic memory decline, and neurodegeneration in aging. *Journal of Neuroscience*, 38(3), 530-543. doi:10.1523/JNEUROSCI.2028-17.2017
- McCarthy, C., 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. doi:10.3389/fnins.2015.00379/full
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to 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), 263-269. doi:10.1016/j.jalz.2011.03.005
- Molinuevo, J. L., Blennow, K., Dubois, B., Engelborghs, S., Lewczuk, P., Perret-Liaudet, A., . . . Parnetti, L. (2014). The clinical use of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis: A consensus paper from the Alzheimer's Biomarkers Standardization Initiative. *Alzheimer's & Dementia*, 10(6), 808-817. doi:10.1016/j.jalz.2014.03.003
- Ono, M. M. D., Kubik, S., & Abernathey, C. D. (1990). *Atlas of the cerebral sulci*: Stuttgart : G. Thieme Verlag ; New York : Thieme Medical Publishers, 1990.
- Ovalioglu, A. O., Ovalioglu, T. C., Canaz, G., & Emel, E. (2018). Morphologic variations of the collateral sulcus on the mediobasal region of the temporal lobe: An anatomical study. *World Neurosurgery*, 118, e212-e216. doi:10.1016/j.wneu.2018.06.156
- Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., Hanninen, T., Laakso, M. P., . . . Soininen, H. (2004). Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiology of Aging*, 25(3), 303-310. doi:10.1016/s0197-4580(03)00084-8
- Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., . . . Jack, C. R., Jr. (2009). Mild cognitive impairment: ten years later. *Archives of Neurology*, 66(12), 1447-1455. doi:10.1001/archneurol.2009.266
- Portney, L. G., & Watkins, M. P. (2013). *Foundations of clinical research: Applications to practice* (Vol. Third edition). Harlow: Pearson.
- Pruessner, J. C., Kohler, S., Crane, J., Pruessner, M., Lord, C., Byrne, A., . . . Evans, A. C. (2002). Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal

- cortex from high-resolution MR images: considering the variability of the collateral sulcus. *Cerebral Cortex*, 12(12), 1342-1353. doi:10.1093/cercor/12.12.1342
- Reitz, C., Brickman, A. M., Brown, T. R., Manly, J., DeCarli, C., Small, S. A., & Mayeux, R. (2009). Linking hippocampal structure and function to memory performance in an aging population. *Archives of Neurology*, 66(11), 1385-1392. doi:10.1001/archneurol.2009.214
- Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, 61(4), 1402-1418. doi:10.1016/j.neuroimage.2012.02.084
- Schmidt, M. F., Storrs, J. M., Freeman, K. B., Jack, C. R., Jr., Turner, S. T., Griswold, M. E., & Mosley, T. H., Jr. (2018). A comparison of manual tracing and FreeSurfer for estimating hippocampal volume over the adult lifespan. *Human Brain Mapping*, 39(6), 2500-2513. doi:10.1002/hbm.24017
- Sperling, R., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., . . . Phelps, C. H. (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. doi:10.1016/j.jalz.2011.03.003
- Sperling, R., & Johnson, K. (2013). Biomarkers of Alzheimer disease: Current and future applications to diagnostic criteria. *Continuum*, 19(2 Dementia), 325-338. doi:10.1212/01.CON.0000429181.60095.99
- Stokin, G. B., & Goldstein, L. S. B. (2006). Axonal Transport and Alzheimer's Disease. *Annual Review of Biochemistry*, 75(1), 607-627. doi:10.1146/annurev.biochem.75.103004.142637

- Stoub, T. R., Rogalski, E. J., Leurgans, S., Bennett, D. A., & deToledo-Morrell, L. (2010). Rate of entorhinal and hippocampal atrophy in incipient and mild AD: Relation to memory function. *Neurobiology of Aging*, 31(7), 1089-1098. doi:10.1016/j.neurobiolaging.2008.08.003
- Summers, M. J., & Saunders, N. L. (2012). Neuropsychological measures predict decline to Alzheimer's dementia from mild cognitive impairment. *Neuropsychology*, 26(4), 498-508. doi:10.1037/a0028576
- Tapiola, T., Pennanen, C., Tapiola, M., Tervo, S., Kivipelto, M., Hänninen, T., . . . Soininen, H. (2008). MRI of hippocampus and entorhinal cortex in mild cognitive impairment: A follow-up study. *Neurobiology of Aging*, 29(1), 31-38. doi:10.1016/j.neurobiolaging.2006.09.007
- Taylor, K. I., & Probst, A. (2008). Anatomic localization of the transentorhinal region of the perirhinal cortex. *Neurobiology of Aging*, 29(10), 1591-1596. doi:10.1016/j.neurobiolaging.2007.03.024
- Trzepacz, P. T., Yu, P., Sun, J., Schuh, K., Case, M., Witte, M. M., . . . Hake, A. (2014). Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to Alzheimer's dementia. *Neurobiology of Aging*, 35(1), 143-151. doi:10.1016/j.neurobiolaging.2013.06.018
- Tward, D. J., Sicat, C. S., Brown, T., Bakker, A., Gallagher, M., Albert, M., & Miller, M. (2017). Entorhinal and transentorhinal atrophy in mild cognitive impairment using longitudinal diffeomorphometry. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 9, 41-50. doi:10.1016/j.dadm.2017.07.005
- Van Hoesen, G. W., Hyman, B. T., & Damasio, A. R. (1991). Entorhinal cortex pathology in Alzheimer's disease. *Hippocampus*, 1(1), 1-8. doi:10.1002/hipo.450010102

- Velayudhan, L., Proitsi, P., Westman, E., Muehlboeck, J. S., Mecocci, P., Vellas, B., . . . AddNeuroMed, C. (2013). Entorhinal cortex thickness predicts cognitive decline in Alzheimer's disease. *Journal of Alzheimer's Disease*, 33(3), 755-766. doi:10.3233/jad-2012-121408
- Venneri, A., Gorgoglione, G., Toraci, C., Nocetti, L., Panzetti, P., & Nichelli, P. (2011). Combining neuropsychological and structural neuroimaging indicators of conversion to Alzheimer's disease in amnesic mild cognitive impairment. *Current Alzheimer Research*, 8(7), 789-797. doi:10.2174/156720511797633160
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., . . . Alzheimer's Disease Neuroimaging, I. (2015). 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's & Dementia*, 11(6), e1-120. doi:10.1016/j.jalz.2014.11.001
- Weschler, D. (2009). *WMS-IV technical and interpretive manual*. San Antonio, Tex: Pearson.
- Wirths, O., & Bayer, T. A. (2010). Neuron loss in transgenic mouse models of Alzheimer's disease. *International Journal of Alzheimer's Disease*, 2010. doi:10.4061/2010/723782
- Wolk, D. A., Das, S. R., Mueller, S. G., Weiner, M. W., Yushkevich, P. A., & Alzheimer's Disease Neuroimaging, I. (2017). Medial temporal lobe subregional morphometry using high resolution MRI in Alzheimer's disease. *Neurobiology of Aging*, 49, 204-213. doi:10.1016/j.neurobiolaging.2016.09.011
- Xie, L., Das, S. R., Wisse, L. E. M., Ittyerah, R., Yushkevich, P. A., Wolk, D. A., & Alzheimer's Dis Neuroimaging, I. (2018). Early tau burden correlates with higher rate of atrophy in transentorhinal cortex. *Journal of Alzheimer's Disease*, 62(1), 85-92. doi:10.3233/jad-170945

- Xie, L., Pluta, J. B., Das, S. R., Wisse, L. E. M., Wang, H., Mancuso, L., . . . Yushkevich, P. A. (2017). Multi-template analysis of human perirhinal cortex in brain MRI: Explicitly accounting for anatomical variability. *Neuroimage*, 144(Pt A), 183-202. doi:10.1016/j.neuroimage.2016.09.070
- Yue, L., Wang, T., Wang, J., Li, G., Wang, J., Li, X., . . . Xiao, S. (2018). Asymmetry of hippocampus and amygdala defect in subjective cognitive decline among the community dwelling Chinese. *Frontiers in Psychiatry*, 9, 226. doi:10.3389/fpsyt.2018.00226
- Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S. L., Gertje, E. C., . . . Wolk, D. A. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, 36(1), 258-287. doi:10.1002/hbm.22627
- Zhao, W., Wang, X., Yin, C., He, M., Li, S., & Han, Y. (2019). Trajectories of the hippocampal subfields atrophy in the Alzheimer's disease: A structural imaging study. *Frontiers in Neuroinformatics*, 13, 13. doi:10.3389/fninf.2019.00013