

Volumetric assessment of hippocampus and cerebral white matter lesions in structural MRI

Erik Olsson

Department of Psychiatry and Neurochemistry
Sahlgrenska Academy at University of Gothenburg



UNIVERSITY OF GOTHENBURG

Gothenburg 2013

Cover illustration : by Erik Olsson & Sigun Bergstedt

Volumetric assessment of hippocampus and cerebral white matter lesions in
structural MRI

© Erik Olsson 2013

erik.olsson@neuro.gu.se

ISBN 978-91-628-8610-3

Printed in Gothenburg, Sweden 2013

Ale Tryckteam

To my father

Never give in, never give in, never; never; never; never - in nothing, great or small, large or petty - never give in except to convictions of honor and good sense ...

Winston Churchill, 1941

Volumetric assessment of hippocampus and cerebral white matter lesions in structural MRI

Erik Olsson

Department of Psychiatry and Neurochemistry
Institute of Neuroscience and Physiology
Sahlgrenska Academy at University of Gothenburg, Göteborg, Sweden

ABSTRACT

Assessment in structural MRI like hippocampal volumetry and white matter lesion (WML) assessment is receiving widespread attention and recommendation as important research and diagnostic tools. The aim in this thesis is to contribute to enhanced reliability and validity in structural MRI assessment.

The hypothesis in Paper I was that long-term survivors of head and neck cancer with lowered quality of life had radiation induced damage to the hippocampus. The main hypothesis in Paper II was that patients with mild cognitive impairment subsequently converting to Alzheimer's or vascular dementia had hippocampal atrophy. The main aim in Paper III was to explore reliability in three types of WML assessment methods. Manual hippocampal volumetry was used in Paper I and II. A visual assessment method, a manual segmentation with thresholding method, and an automatic volumetry method were used in Paper III.

Low dose radiation gave no volumetrically discernible damage to the hippocampus. Other possible effects of such radiation on the brain remain to be explored. Left hippocampal atrophy predicted conversion to dementia, which confirms its usefulness as a biomarker. Low reliability for low and medium volumes in WML assessment in clinical samples implies a need of refined methodology and reliability analysis.

Keywords: magnetic resonance imaging, hippocampal volumetry, white matter lesions, mild cognitive impairment, dementia, low dose radiation
ISBN: 978-91-628-8610-3

SAMMANFATTNING PÅ SVENSKA

Bildkvalitet, bildfiltrering och metodologi är avgörande faktorer när man vill kvantifiera sjukliga eller behandlings-specifika förändringar med strukturell hjärnabbildning med magnetkamera. Premorbid individvariation kan ligga i samma storleksordning som förväntad förändring vilket då påkallar normaliseringsmetoder. Reliabilitet är en repeterbarhetsanalys och är en viktig utvärdering av kvantitativa magnetkamera-metoder och är i fokus för denna avhandling.

Material från två studier ingår i avhandlingen. Den första artikeln analyserar sambandet mellan livskvalitet och hjärnpåverkan hos långtidsöverlevare vid strålbehandling av cancer i hals- och nackregionen, där 15 kontroller och 15 patienter skannades med en 1.5 T magnetkamera. De två följande studier är patienter ur Gothenburg MCI (mild cognitive impairment) Study, där kognitiv störning och demensutveckling följs under flera år. Deltagarna i studien skannades med en 0.5 T magnetkamera i hippocampus-substudien (26 kontroller och 42 MCI-patienter) och med en 1.5 T magnetkamera i substudien av vitsubstanslesioner (white matter lesions, WML) (28 kontroller och 124 patienter). Hippocampusvolumetri utfördes manuellt i de två första studierna och normaliserades med intrakraniell volym. WML mättes med tre metoder, Fazekas visuella skattning, en manuell segmenterings- och trösklingsmetod och en automatisk volumetrisk metod.

Interbedömar-reliabiliteten för hippocampus-volumetrin ligger i intervallet Pearson's $r = 0.84-0.94$ och intraklasskorrelationen $ICC = 0.66-0.85$ (beroende på val av ICC-mått) med likartad reliabilitet i båda studierna.

Långtidsöverlevare efter strålbehandling av cancer i hals- och nack-regionen visar inte några mätbara förändringar i hippocampusvolym i förhållande till

en matchad kontrollgrupp.

Hippocampusvolymen, särskilt på vänster sida, hos MCI patienter som senare progredierar till demens är mindre vid baslinjen än hos kontroller och stabila MCI patienter.

I båda hippocampusstudierna var reliabiliteten hos ICV-skattningarna mycket hög. ICV-normaliseringen reducerade variansen med 46% och ökade avsevärt den diskriminativa förmågan hos de uppmätta hippocampusvolymerna.

Mätning av vitsubstansförändringar i MCI-studien med tre olika typer av metoder visar acceptabel övergripande reliabilitet men låg reliabilitet för lägre volymer. Den manuella metoden visar ingen reliabilitet för låga volymer. Den bristande reliabiliteten kan härledas till bilddata från främre hjärndelar där intensitetsdistorsionen är som tydligast.

Reliabiliteten för hippocampusvolumetrin är huvudsakligen likartad mellan studierna trots att bildkvaliteten är bättre i strålbehandlings-studien. Detta kan tolkas som att segmenteringsmetoden är robust. Resultaten av de manuella WML-mätningarna indikerar att reliabiliteten vid låga volymer av WML kan påverkas av intensitetsdistorsion och att en global reliabilitetsanalys kan behöva kompletteras med analyser av reliabiliteten avseende delar av materialet.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

- I. Olsson E, Eckerström C, Berg G, Borga M, Ekholm S, Johannsson G, Ribbelin S, Starck G, Wysocka A, Löfdahl E, Malmgren H: **Hippocampal volumes in patients exposed to low-dose radiation to the basal brain. A case-control study in long-term survivors from cancer in the head and neck region.** *Radiation Oncology* 2012, **7**:202.
- II. Eckerström C, Olsson E, Borga M, Ekholm S, Ribbelin S, Rolstad S, Starck G, Edman A, Wallin A, Malmgren H: **Small baseline volume of left hippocampus is associated with subsequent conversion of MCI into dementia: the Goteborg MCI study.** *J Neurol Sci* 2008, **272**:48-59.
- III. Olsson E, Klasson N, Berge J, Eckerström C, Edman Å, Malmgren H, Wallin A: **White matter lesion assessment in patients with cognitive impairment and healthy controls: reliability comparisons between visual rating, a manual and an automatic volumetrical MRI method - the Gothenburg MCI Study.** *J Aging Research*, in press.

CONTENT

ABBREVIATIONS	IV
Foreword	vi
1 INTRODUCTION	1
1.1 Uses of MRI volumetry	1
1.2 Refinement of methodological strategies	2
1.2.1 Scanners and scanning protocols	3
1.2.2 Image quality: MRI intensity variation in normal tissue, aging and pathology	4
1.2.3 Anatomical variability, aging and pathological variability	6
1.2.4 Segmentation methods: validation and further development	12
1.2.5 Comparability	17
1.2.6 Refined power (larger studies)	19
2 PRESENTATION OF THE PAPERS	20
2.1 Background	21
2.1.1 Clinical background and aims	21
2.1.2 Methodological background	25
2.2 Materials and methods	29
2.2.1 Study characteristics of the hippocampal volumetry studies	29
2.2.2 Methods used in the hippocampal studies	30
2.2.3 Study characteristics of the WML study	38
2.2.4 Methods used in the WML study	38
2.3 Results	41
2.3.1 Methodological results of the hippocampal studies	41
2.3.2 Clinical results of the hippocampal studies	43
2.3.3 Results of the WML study	44
3 DISCUSSION	47
3.1 Discussion of the hippocampal studies	47
3.2 Discussion of the WML study	51

4	CONCLUSIONS.....	55
5	FUTURE DIRECTIONS.....	56
6	REFERENCES	59

ABBREVIATIONS

AD – Alzheimer’s disease

ADNI – The Alzheimer's Disease Neuroimaging Initiative

AUC – area under the curve

BA plot – Bland Altman plot

BDNF – brain derived neurothropic factor

BMI – body mass index

CA – cranial area

CNS – central nervous system

CSF – cerebro spinal fluid

CT – computer tomography

DTI – diffusion tensor imaging

ECT – electroconvulsive therapy

EEG – electro encephalography

fMRI – functional magnetic resonance imaging

GDS – global deterioration scale

GLM – general linear modeling

ICC – intra-class correlation

ICV – intracranial volume

MCI – mild cognitive impairment

MRI – magnetic resonance imaging

MS – multiple sclerosis

MTL – medial temporal lobe

SNP – single-nucleotide polymorphism

SPECT – single-photon emission computed tomography

TESC gene – tescalcin gene

ROC – receiver operating characteristic

SVM– support vector machines

TAO – thyroid associated opthalmopathy

TBV – total brain volume

WML – white matter lesions, white matter changes, (white matter hyperintensities, white matter hypointensities, leukoaraiosis)

Foreword

This thesis starts with an *Introduction* containing an overview of structural neuro MRI assessment and some methodological problems involved that have a bearing on the presented studies. Then follows the *Background* to the three studies included in the thesis, two about hippocampal volumetry and one about white matter lesion (WML) measurement. In the *Materials and methods* section study settings and methods are summarized for the hippocampal volumetry studies and for the WML study. In the *Results* section the studies are summarized in a hippocampus and a WML passage, and starts with an overview of reliability results where the hippocampal studies are compared. In the *Discussion* interpretations and implications of the studies are summarized. In the *Conclusion* the major findings are summarized followed by a statement of intent concerning future directions.

1 INTRODUCTION

1.1 Uses of MRI volumetry

Quantitative assessment of brain structure volumes and pathological changes visible in magnetic resonance imaging (MRI) has two main purposes: 1) diagnosing and monitoring of diseases, and 2) to clarify treatment and intervention effects. MRI volumetry of brain structures, especially the hippocampus, has for example been used to detect pathological changes in Alzheimer's disease (AD), mild cognitive disorder (MCI), temporal lobe epilepsy, schizophrenia, depression (where enlargement of hippocampus after electroconvulsive therapy (ECT) treatment has also been seen), stress and Cushing's syndrome (growth of hippocampus in recovery) and as a side effect of radiotherapy [1,2]. These diagnostic and monitoring possibilities have so far mostly been studied on a research basis. Diagnostic use in dementia can be expected to get more widespread in a near future judging from reports about AD diagnostic criteria [3,4]. These reports state that hippocampal volumetry and white matter lesion (WML) assessment are gaining acceptance as tools in standardizing dementia diagnostics. Hippocampal volumetry is also becoming recognized as an important biomarker for detecting AD treatment effects in clinical trials [5].

1.2 Refinement of methodological strategies

Findings in structural MRI volume assessment of brain structures and pathologies do not always point in the same direction. For example, in a study of healthy aging by Sullivan et al. no significant decline could be found in the hippocampus, but rather in the temporal neocortex [6]. In a large study by Walhovd et al., consisting of five different samples, the hippocampus was in contrast found to have an accelerating atrophy in healthy aging [7]. Further, hippocampal atrophy is often found to be associated with dementia [8]. WML are often found to be associated with vascular or mixed dementia but not with dementia in general nor with AD [9,10]. However, in a recent study by Brickman et al. WML in the parietal lobe predicted time to incident AD but hippocampal volume did not [11].

In the literature of structural MRI volume assessment various cases of discrepant results show a need for refined methodological strategies. This could for example be achieved by:

1. Improvements in hardware and scanning protocols.
2. Refined imaging quality; better preprocessing in the form of distortion filtering techniques and image intensity normalization.
3. Controlling for irrelevant sources of anatomical variation.
4. Improved and preferably validated and standardized segmentation methods.

5. Enhanced comparability between studies by better specification of participant selection criteria, sufficient methodological detail in study reports and compliance to methodological standardization efforts in study design.

6. Refined power.

1.2.1 Scanners and scanning protocols

The influence of scanner vendor model, field strength and coil type differences was analyzed in a study by Kruggel et al. using the ADNI (The Alzheimer's Disease Neuroimaging Initiative) optimized MPRAGE sequence. They found that scanner hardware factors explained 30-50% of the variance of the studied variables [12]. The studied variables were signal to noise and grey/white matter contrast parameters as well as gross volumetric compartments. The scan/rescan variability for the compartments was about ten times higher when the subjects were rescanned in different scanner hardware. The recommendation hence was to include scanner hardware as a covariate when analyzing multicenter data. Similar conclusions are drawn in a study by Jovicich et al. that moreover found no significant change in variance of analyzed brain structures but a possible bias in mean volume differences after scanner upgrades [13]. In a study by Huppertz et al. inter scanner variability was found to be five times higher than intra scanner variability [14]. However, the choice of segmentation method has been found to give a larger contribution to variance than the choice of scan sequence [15]. As the three studies mentioned above were performed with different segmentation methods this may have given the largest contribution to the different (but coherent) figures concerning reproducibility errors. Hence the

best strategy for diminishing the tension between the ongoing technological development and the desire for comparability of studies may be to improve and standardize the segmentation methods.

Even if standardization of segmentation methods is crucial the reliability of volumetric assessments must be improved by achievements in other areas too. The variability resulting from MR technical causes like noise and image intensity variation can to some extent be accounted for by noise reduction, intensity normalization methods (see below).

1.2.2 Image quality: MRI intensity variation in normal tissue, aging and pathology

A confounder in structural neuro MRI is varying tissue intensities within a single scan and between subjects. In conventional MRI the intensities are not solely determined by the physical properties corresponding to the voxel but also depend on hardware related effects as magnetic field inhomogeneities and possibly cross-talk between slices [16,17]. The impact of the mentioned variation can be reduced by post scan image intensity standardization and normalization. Another option is scanning techniques for minimizing intensity distortion e.g. [18,19] but the resulting overall image quality has been ranked inferior to conventional scanning [20,21]. Another factor limiting applicability is that the product sequence in today's synthetic MRI, where intensities are closely correlated with physical properties [19], does not produce images with sufficient resolution for segmentation of e.g. temporal lobe substructures.

Only part of the intensity variation in structural brain MRI is due to hardware and imaging related errors. MRI relaxation times for both normal appearing white matter and WML vary throughout the brain [22], and between fiber tracts [23]. Although high iron content can be found in white matter due to the need of iron in myelin maintenance [24] the relaxation time variation in WML is mainly due to differences in myelinisation rather than in iron content [25]. In grey matter apparent transverse relaxation has been found to reflect differences in regional iron distribution [26].

Grey and white matter contrast in MRI has been found to have a heritable component [27]. The contrast generally decrease with age and with certain pathologies, like Alzheimer's disease (AD). The T2 relaxation times of both grey and white matter depend on age and localization but the largest cause of reduced contrast is the changes in white matter. Age affects the amount and structure of myelinisation, which reduces grey and white matter contrast. T2 relaxation times have been found to generally increase during maturation and decrease in aging, but they follow different age trajectories in gray matter structures like hippocampus, amygdala and caudatus compared to callosal, orbitofrontal, temporal and occipital white matter [28]. Different white matter pathologies may produce changes of T2 relaxation times in the same range. In Oakden et al. [29], differences in T2 relaxation components in controls and probable AD were interpreted as indications of two types of WML where one type with very low myelin water fraction was found in all probable AD but only in less than half of the controls.

The influence of age on image contrast between gray and white matter has been corrected for by general linear modeling (GLM) by Westlye et al. [30]. In their study an age related decline in contrast was observed in the frontal, temporal and parietal lobes. This study also found that adjusting for contrast

increased the sensitivity to AD for the cortical thickness estimates. Alzheimer's disease is however itself associated with a more severe myelin reduction [31], and hippocampal degeneration has been found to be associated with contrast changes in temporal and limbic grey as well as in white matter [32]. These changes in contrast can be expected to blur the measurement of disease related volume changes. The entorhinal cortex is affected in the earliest stages of AD [33] and hence its thickness is a potentially important MRI biomarker. However, in the study by Westlye et al. the effect size was significantly *reduced* in entorhinal cortex after adjustment for the influence of age on image contrast [30], which may be due to a faulty adjustment of the increased disease related changes in image contrast.

A useful intensity normalization procedure must not extinguish clinically or scientifically relevant differences. For example, age related changes in grey and white matter are often relevant to preserve in a further analysis in order to get an accurate estimate of age atrophy. An intensity normalization method that is widely used and publicly available is the N3 method [34]. N4ITK is a development of the N3 method and its source code is also publicly available [35]. Several other intensity normalization methods have been presented recently which claim to be improvements of the predecessors [36-38]. See also *Future directions*.

1.2.3 Anatomical variability, aging and pathological variability

Beside variations in image contrast, several dimensions of real variation over time and between different individuals must be considered and if possible

controlled for. The inter-individual variation in volume and shape of e.g. hippocampus is a confounder in manual segmentation and in atlas based transformations as used in some automatic procedures. Several studies have been dedicated to clarify different aspects of inter individual variation for example normal volume variation [39], ethnic brain morphology [40], infant brain growth rate [41,42], hemispherical differences related to handedness [43-45] and sulcal variability [46-48]. In the following paragraphs some sources of normal and pathological anatomical variation are described, starting with the normal spectrum.

Hereditary and environmental variability

A general hereditary influence on cortical thickness and surface area [49] may contribute to neuroanatomical variation but differences between hemispheres do not seem to be hereditary [50]. In a study by Eyler et al. high heritability for cortical surface area of the frontal lobe and low heritability for the temporal lobe were found [51].

It has been shown in a twin study that about half of the hippocampal volume variability is explained by heredity while the frontal lobe heritable component might be as high as 95% [52]. Brain derived neurotrophic factor (BDNF) plays an important role in synaptic plasticity and neurogenesis and maintenance of neurons. BDNF is heavily expressed in hippocampus and a polymorphism in the BDNF gene has been shown to affect hippocampal volume in terms of both mean and variance [53]. Some hereditary influence could be a secondary effect from hereditary influences on total brain volume (TBV). However, a recent meta analysis study found no BDNF genetical

influence on hippocampal volume but found a single-nucleotide polymorphism (SNP) (associated with the expression of the tescalcin (TESC) gene) influence on hippocampal volume regardless of intracranial volume (ICV), brain size or disease [54].

As hippocampus is a highly plastic structure it is also affected by environmental factors e.g. learning [55].

Sex differences

Brain parenchymal volume differs between sexes with about 10 % but the difference decreases with age due to a somewhat larger age decline in males [7,56]. The hippocampal gender differences are smaller but the figures differ between studies, manually assessed volumes have shown a difference of about 3% [56] but automatic assessments have found a difference of nearly 10% [7].

Age-related atrophy and pathological changes of the brain

Age-related changes in the brain are special in that the differentiation of normal and pathological changes is difficult already on the conceptual level [57]. However, in spite of some conflicting reports (see above) it can be stated with some confidence that certain common age-related changes are not caused by separately identifiable disease processes. While Sullivan et al. [6] found no significant age decline in hippocampal volume, an accelerating age effect has been found by both automatic [7] and manual volumetry [58].

Subregional manual volumetry results by Malykhin et al. indicate that the posterior part of hippocampus might be more affected by age-related atrophy than the anterior part [59]. Taken together with indications of an anterior emotional and posterior cognitive dominance [60], this could mean relatively intact emotional processing in aging. However, a recent study found an association between anterior hippocampal atrophy and functional impairment in normal aging [61]. This study had a better field strength but Malykhin et al. may have used a more accurate segmentation protocol, which calls for further study of this issue.

Disease related shape deviations

Severe deviation from normal anatomical shape of e.g. hippocampus can occur as a delamination of cornu ammonis and gyrus dentatus in cases of marked atrophy or hydrocephalus. Although some demarcations between tissue and cerebrospinal fluid (CSF) can be more apparent, such pathological deviations from normal anatomy generally make the manual segmentation more difficult, and found volumes become much more difficult to interpret. These gross anatomical deviations can be expected to concern automatic atlas based segmentation methods even more than manual segmentation. In AD a slight rotation of the hippocampal head has been found [62] that might lead to similar problems. In the study by Adachi et al. the shape of the hippocampal body was found to be more rounded with increasing atrophy in AD [63], which is in accord with the *post mortem* MRI shape analysis in Dawe et al. [64].

Normalization

Due to systematic gender differences and other normal inter-individual variance in total and regional brain volumes regional atrophy of e.g. hippocampus can hardly be assessed directly for diagnostic purposes except as volumetric change in longitudinal MRI studies. In cross-sectional studies some correction is usually motivated, especially in small studies where e.g. gender stratification is not feasible. Ideally the directly assessed volume should be related to premorbid volume. As a substitute, a correction is often done by normalizing the volume to intracranial volume (ICV). The rationale for this procedure is the strong correlation in healthy subjects between skull size and regional brain volumes, and the technique used is based either on a linear regression of regional volume on ICV or on a simpler assumption of proportionality. In studies with high quality data and highly reliable measurements it might be motivated in the future to take non-linearities into account in the regression [65]. The aim is to improve criterion validity [66], and the confounding effects of gender (and to some extent the normal age related volume variation) can be controlled by the normalization [56,67].

Assessment of other age-related or pathological changes such as white matter lesions (WML) is not necessarily affected by normal individual variation in the same way and normalization by estimated premorbid volume is here not clearly motivated. However, there are findings where patients with high brain reserve as estimated with ICV are more resistant to the cognitive effects of WML pathology [68,69].

Total brain volume (TBV) has also been used for normalization of regional measurements and comparisons between this and ICV normalization have given conflicting results. In Jack et al. [70], ICV normalization resulted in the

most consistent reduction in variance but in Free et al. [71] the strongest correlation was found between TBV and hippocampal volume. Bigler et al. [72] found TBV normalized hippocampal volumes to give the best separation of controls and patients, while ICV normalized hippocampal volumes did not improve the classification over absolute volumes. Several other normalization methods have been proposed but only TBV, cranial area (CA) and ICV correlate with hippocampal volumes.

The best way for normalization is mainly an open question but the pathology studied and cross sectional or longitudinal study design influence the choice. TBV normalization is problematic in pathologies where cerebral atrophy occurs. Also, the aging brain normally loses volume, while the intracranial vault does not, although a possible secular trend caused by changes in socioeconomic and nutritional conditions must be considered in cross-sectional studies [73]. In comparison with the whole brain the hippocampus remains fairly intact in normal aging [6,74]. This speaks in favor of ICV normalization in studies of the hippocampus. A combination of TBV and ICV may be the best generalized method for normalization [75]. This is however more time consuming but a fully automated method would remove this obstacle.

The result of an ICV normalization also depends on the method used to estimate ICV. A total ICV segmentation of all slices in a scan series is optimal but generally too time-consuming so many simpler estimates have been developed. Obenaus et al. [76] compares four simple head-size related measures and the most reliable one for normalization of pediatric hippocampal volumes was a measure of intracranial diameter developed in [77]. Ferguson et al. [78] uses intracranial area for one selected slice. In contrast Eritaia et al. [79] analyzes the accuracy and efficiency of ICV

segmentation tracing all variants of x equidistant 0.938 mm slices and finds that the correlation with a full segmentation soon rises to acceptable levels. The authors recommend every 10th slice as a rule of thumb and this recommendation has been followed in several studies. See also *Further directions*.

1.2.4 Segmentation methods: validation and further development

Validation of segmentation procedures

Theoretically, reliability is an estimate of the agreement to the true score and intra-class correlation (ICC) estimates the ratio of true variance to (true variance plus error variance). The intra-class correlation between two ratings is often used as a reliability estimate in MR volumetry [80]. Different versions of ICC and their relevance will be discussed below in connection with Paper I. The true variance is the individual anatomical variation and error variance is the errors accumulated in the imaging and segmentation procedures. However, reliability studies can in practice only provide indirect evidence for the validity of a segmentation method. Although an unreliable method cannot be wholly accurate, perfect reliability does not imply that the method is free from systematic imaging or segmentation errors.

To know the accuracy of a MR-segmentation of a brain structure the ideal would be to know the real demarcation. Available techniques for validation comprise post mortem MRI, other post mortem investigations, resected tissue in temporal lobectomy, MRI phantoms, comparisons with manual “golden

standards” and assessment of the predictive power of the method with respect to some biological variable, e.g. a clinical category or a biomarker (also referred to as criterion validity).

In a study by Lee et al. hippocampi resected in temporal lobectomy was used for MRI validation [81]. Hippocampal MRI assessed volume was shown to correlate to neuronal count. In several studies it has also been found that MR volumetry of formalin fixed brains is a valid method through comparison with histological neuronal count [64,82]; and scan parameter guidelines for *post mortem* MRI of formalin fixed brains have been presented [83]. However, there are changes in shape, volume and MR signal intensities in post mortem tissue, which implies difficulties if one want to use it as a tool for validation of *in vivo* MR volumetry. In a study by Van Duijn et al. hypointensity artifacts related to formalin fixation were found to be indistinguishable from brain pathology [84].

Phantoms can be used for optimization of acquisition sequences [85], to identify scanner errors in multi site studies [86], for control of scanner drift [21] and for rater validation. Optimal voxel size has been studied in [87] where a simple phantom was used to compare voxel dimensions in 3D-MRI. It was concluded that voxels with both an isotropic shape and small volume give the best volumetric results. Similar results have been shown in [88]. Phantoms can further be used for intra- and inter-site comparisons [89]. For rater validation a phantom that is an accurate replica of a neuroanatomical region would be preferred. The difficulty lies in building a phantom with determinable volumetric compartments that has signal intensities comparable to both (e.g.) hippocampus and its environment. It is also possible to use *digital* phantoms for optimization and validation of post scan procedures [90].

Evaluation of predictive performance

Predictive performance can play an important role in the validation of segmentation methods. Several examples will be given below in the discussion of the hippocampal studies. In this introduction, only a certain methodological problem that has to do with the measurement of predictive power will be mentioned. It is important because advanced statistical methods for optimizing classification and prediction from multidimensional input data, e.g. MRI volumes together with biochemical and psychometric parameters, are becoming more and more common.

Predictive performance is often evaluated by sensitivity and specificity, positive and negative predictive value and receiver operating characteristic (ROC) analysis. E.g. in Geremia et al. [91] a random forest classification showed a significant improvement compared to an earlier automatic method using multi-modal MRI in multiple sclerosis (MS). In Chincarini et al. [92] volumes of medial temporal lobe (MTL) structures and their intensity and textural features were classified in an analysis using a random forest classifier followed by a support vector machine (SVM) classifier resulting in a very high area under the curve $AUC = 0.97$. When evaluating several covariates logistic regression and other data mining analyses can be used. In a data mining comparison study of psychometric data in an MCI and AD sample, SVM showed the highest AUC (0.90) but low specificity. Somewhat paradoxically, for random forests and linear discriminant analysis the overall accuracy were considered higher with $AUC = 0.73$ and 0.72 respectively and acceptable sensitivity and specificity [93]. Hence, predictive performance evaluation may in turn need evaluation and guidelines.

Accuracy conditions for automatic and manual segmentation

For enhanced accuracy in structural MRI all stages from the scanning to segmentation are important but in this dissertation the segmentation aspects is in focus.

Automatic and manual segmentation are the both ends along a continuum of technical tools used for MR volumetry and thickness measures. The automatic method most frequently used today is FreeSurfer [94,95], which assesses about forty subvolumes of the whole brain. FreeSurfer uses a probabilistic atlas generated from manually segmented MR scans to execute the segmentation. FreeSurfer has been frequently used for neuroanatomical subregional volumetry and has been shown to be comparable in accuracy to manual labeling for many tasks [94,95], and to perform well compared to other automated segmentation tools [96]. FreeSurfer is available for download online (<http://surfr.nmr.mgh.harvard.edu/>).

Manual segmentation is often regarded as too time-consuming but some of the available “automatic” methods including FreeSurfer may also be quite time-consuming due to a possible need of extensive manual editing for each patient.

Automatic methods depend on manual ones in two ways. As a rule, they rely on manually segmented atlases. Also, the validation of an automatic method is often done by comparing with manual segmentation as a golden standard [97-99]. In order to get better automatic volumetry, and to further improve the softwares, the manual methods also need to be improved. The hippocampi have complex demarcations to the surrounding tissue in MR images and an accurate segmentation procedure puts pressure on reliability

for human raters. It is possible to enhance reliability at the expense of validity but all such strategies will decrease the quality of the evaluation of methods regardless of whether it is on automatic or manual.

Improvements in manual segmentation

Although one of the main goals of structural MRI research is to develop methods for clinical use, this does not mean that a clinical approach on all levels is the best way forward. High accuracy is a more important goal than speed when developing and evaluating methods sensitive for a certain pathology or change. Rather than directing the research towards ordinary clinical conditions from the beginning it is probably a better strategy to try to acquire all relevant knowledge about the task and then implement this knowledge in an accurate solution. For example, to develop a segmentation protocol for hippocampal volumetry or regional or total WML volumetry it is important to evaluate anatomical and cytoarchitectural demarcations in modalities with more information. This could be done by neuropathological histology studies or by e.g. *ex vivo* long time high resolution scanning [100] and *in vivo* ultra high resolution by repeated acquisitions [101]. The results should then be taken into consideration in the regular MRI procedure.

A segmentation protocol for standardized demarcation of e.g. hippocampus can be established by evaluating the reliability figures for candidate methods with different demarcation criteria [102]. However, excluding or including candidate subregions e.g. white matter structures like alveus or fimbria in the hippocampal region of interest is also a question of relevance for the studied

pathology. This means that etiological aspects as well as predictive performance evaluation must be considered.

1.2.5 Comparability

Methodological standardization

It is often difficult or unfeasible to compare segmented volumes and statistical results between research centers due to technical and methodological differences. To enhance validity, an expanded cooperation between centers regarding standardization of scan parameters, anatomical demarcation criteria, segmentation technique and multirater segmenting has been requested [103,104]. Within ADNI, The Alzheimer's Disease Neuroimaging Initiative [105], optimized MRI protocols have been developed to enhance comparability between centers [21] and a survey of manual hippocampal volumetry protocols has been published in order to develop a standardized protocol [102]. ADNI standardized sets of MRI data can be downloaded and results and methods used for analysis of these data can be reported to ADNI for comparison [106]. Simulated MR images of brains are available at BrainWeb from McGill University and real data from several sources are found at the Internet Brain Segmentation Repository. They can all be processed and segmented for training, optimization or validation studies [107]. The availability of extensive standardized normative data may also to some extent reduce the need for large control groups in MR studies, see e.g. [108]. A possible drawback of standardization is however the continuous need of improvement in most aspects of neuroimaging research,

which makes it hard to settle definitions and other standardizations permanently.

Another important task for enhanced comparability is reporting studies in sufficient detail and standardization guidelines has been proposed both for reliability [109] and diagnostic utility studies [110].

Patient selection, exclusion and classification

In epidemiologic studies it has been found that women, persons with higher socioeconomic status, persons with higher education, married persons and employed persons are more likely to participate in a study but there is little evidence for a sampling bias due to the mentioned participant properties [111]. However, better health due to e.g. higher socioeconomic status implies a possible sampling bias, which calls for further studies on this issue including potential cultural differences. Beside these general and well-known issues there are some possible sources of bias that are more specific to structural MRI. A recent fMRI study of multiple sclerosis (MS) found an association between severity of disease and movement artifacts in the MR images. This association may be due to exhausted cerebral resources, fatigue and impaired motor control secondary to cognitive impairment [112]. This hypothesis is relevant for the impact of the exclusion of structural MRI scans with movement artifacts in studies of other patients with cognitive impairment. Movement artifacts add substantial uncertainty in the segmentations and this is the rationale for exclusion of scans with such artifacts. This kind of exclusion may however neglect a specific group of participants and differences in exclusion criteria may limit comparability and contribute to discrepant results between studies.

Further, there is also a possible ascertainment bias where verbally mediated cognitive tests may be more sensitive to selecting left hippocampal pathology [113]. Differences in reported prevalence of e.g. MCI would probably be decreased by standardization of operational definitions of MCI types [114] which would benefit the efforts to achieve comparable and perhaps more concordant overall results.

1.2.6 Refined power (larger studies)

Increased sample sizes will increase the probability of getting significant results but at a larger cost and slower study throughput. Power calculations are used to determine the sample size required for finding a clinically meaningful effect with sufficient probability. Hence, power calculations are important in study design when sample size can be fully controlled. In a study by Morra et al., a sample size of $n = 40$ was required to differentiate hippocampal volumes between AD patients and controls [115]. Ard et al [116] found that for longitudinal treatment trial studies in AD, the required sample size to detect a 25% reduction of the speed of progression was about $n = 100$. For an agent supposed to reduce the speed of the disease-specific atrophy with the same amount beyond projected age-related atrophy, the required sample size would be about four times higher.

When planning a study it is important to account for measurement reliability in the power calculations; e.g. in a case with an ICC of 0.5 the required sample size will be doubled compared to the case with perfect reliability and with an ICC of 0.9 it will be up to 20 % higher [117]. Power is however not the only factor determining sample size. See Discussion.

2 PRESENTATION OF THE PAPERS

This thesis is based on three papers. In two of the papers hippocampal volumetry is used and in the third different assessment methods for white matter lesions (WML) are compared. Olsson et al. [118] is a case-control study of hippocampal volumes in irradiation treated long-term survivors of head and neck cancer. Eckerström et al. [119] is a longitudinal study of hippocampal volumes in mild cognitive impairment (MCI) that was published in 2008. Olsson et al. [120] is a comparison of the reliability of three methods for WML estimation. There are considerable areas of overlap between the hippocampal and WML studies, which motivates presenting them together. The second hippocampal study and the WML paper both concern patients with mild cognitive disorder (MCI) and Alzheimer's Disease (AD), and our group intends to use hippocampal and WML measurements together (also with other indicators) in composite diagnostic measures. Further, several methodological issues pertain to all three studies: especially segmentation difficulties because of intensity variations and other deficiencies of the MR images, and questions about how to measure reliability and how to interpret reliability figures.

2.1 Background

2.1.1 Clinical background and aims

Hippocampal studies

Paper I (The low dose radiation study)

Side effects of high dose radiation therapy directed to the CNS are a well-known concern [121,122]. A plausible hypothesis is that some of these side effects are due to hippocampal damage. The hippocampus is regarded as a neurogenic region of the brain, with the presence of both precursor cells and a microenvironment suitable for production of new neurons [123]. The neurogenic cells are also known to be radiosensitive [124,125]. Children with a slowed cognitive development after treatment of medulloblastoma had a delayed development of their hippocampi [126,127]. Further, a post-mortem study on patients treated with chemotherapy and cranial irradiation, some with reported memory deficits showed profoundly reduced hippocampal neurogenesis [128]. These findings support the hypothesis that neurocognitive impairment after CNS-directed radiation therapy in childhood to some degree is due to a hampered hippocampal neurogenesis [128,129]. If this is so, it has high priority to shield the hippocampus even more than what is standardly done when radiation therapy to the CNS is given.

Less is known about the effects on the brain of low radiation doses, which may result from treatment of cancer outside the CNS, although there is some clinical and laboratory evidence of such effects [130,131]. How the damage to hippocampal neurogenic cells at adult age could lead to cognitive and other impairments as well is not obvious. From a clinical point of view it is

urgent to determine if there are such effects and if so, what the mechanisms behind them are. Better shielding of hippocampus could be one solution to reduce symptoms also at lower doses. A finding of a low-dose effect could further lead to extended restrictions for X-ray investigations in the head and neck region.

Radiotherapy to patients with cancer in the head-and-neck region will result in a low dose to the basal parts of the brain. In a recent study [132], fifteen long-term survivors of such treatment were identified and compared with 15 controls matched for age, sex and BMI. Several quality of life dimensions were significantly compromised in patients compared to controls, which might be related to a negative effect on the CNS of the radiation therapy. The aim of Paper I was to investigate whether the lowered quality of life of the patients was associated with reduced hippocampal volumes. The material was also used for the development and validation of a new intracranial volume (ICV) estimation method (later used in the MCI study). The results of this validation are reported in this thesis but not in the paper.

Paper II (The MCI study)

The term mild cognitive impairment (MCI) describes a state where the cognitive functions are more impaired than would be expected from aging alone but not enough to be described as dementia. In dementia the cognitive functions are impaired to such a degree that it affects the daily living. Quite similar syndromes occur from various causes in younger age groups but the term MCI is almost exclusively used for impairments at older age (or as pre-dementia at younger age). The etiology of MCI is multi-factorial and the

prognosis differs within the group [133,134]. Some MCI patients eventually convert to dementia, but many remain stable and some even improve.

Before our study was carried out, cross-sectional volumetric MRI studies had found some evidence that the hippocampus is significantly smaller in MCI compared to controls and strong evidence that it is smaller in Alzheimer's disease (AD) groups [2,135]. Longitudinal studies had also been performed to investigate if volumetry of various structures in the brain could predict which MCI subject would convert to AD. Hippocampal and entorhinal volumes had been shown to predict this conversion [2,136-140]. The majority of neuroimaging studies published in the field of MCI had focused on the development of AD, the most common form of dementia. Fewer papers studied vascular dementia, the second most common form of dementia [141].

Our study, which is part of the larger Gothenburg MCI study (see below), was the first in a series of cross-sectional and longitudinal investigations with structural MRI in subjects with MCI who either convert or do not convert to dementia at follow up. It was based on a subset of patients and controls that were scanned with a 0.5 T MRI scanner. Paper II tests the hypothesis that baseline hippocampal volumes in MCI patients can predict conversion (at the first follow-up) to dementia. We also address the following issues: asymmetries (left compared to right hippocampal volume), clinical subgroup differences (conversion to AD or non-AD), longitudinal volume changes and the usefulness of ICV normalization.

The white matter lesion study

Paper III

The white matter lesions (WML) study focuses on MCI and dementia. Some amount of WML is associated with normal aging but high WML load entails an increased risk for stroke, cognitive decline, dementia and death [142,143]. It is customary to use CT or MRI to detect WML, and WML are included among the diagnostic criteria for subcortical vascular dementia e.g. in accordance with [144]. In atrophy research, WML is also an important variable possibly confounding primarily gray matter volume in automatic segmentation methods [145]. WML are due to demyelination, axonal loss, gliosis and edema, and mainly affect information processing speed and executive function in cognition [146]. Studies by our group concerning the value of WML assessment in the diagnosis and prediction of dementia have been carried out and are ongoing; a paper co-authored by the author of this thesis shows that white matter lesion load correlates with low hippocampal volume [147].

Several issues concerning the diagnostic significance of WML are still unresolved. There are reports of regional WML or fiber tract integrity association with ischemic or non-ischemic etiology [148], specific domains of cognitive decline [149,150] and pathology [151,152] but other studies find no point in separating WML subregions [153]. Wakefield [154] finds the predictive performance of total WML nearly as good as any regional WML assessment for the functional decline variables mobility, urinary incontinence severity, executive function and processing speed.

Methodological differences (see next section) may explain a large part of these controversies and there is an urgent need for harmonized standards in

WML assessment [155,156]. As long as no consensus is established regarding WML properties or regions important for cognitive impairment of different etiologies, global WML assessment is still a major issue and visual rating scales, manual or automatic volumetry are the methodological types available. The aim of the present study was to assess the inter-rater and inter-method reliability of three commonly used methods for such global assessment; their diagnostic power will be compared in a later paper (see *Future directions*).

2.1.2 Methodological background

Hippocampal volumetry

The definition of hippocampus used here (cornu ammonis and gyrus dentatus) and hippocampal formation (cornu ammonis, gyrus dentatus, subiculum and often entorhinal cortex) follows Duvernoy's sectional anatomy of the hippocampus [157]. Volumetry of the whole hippocampus is often regarded as the best method for quantifying MR detectable pathologies in the hippocampus, at least at field strengths less than 3 T. Which segmentation methods to employ differ between centers, the pathology to be diagnosed, the field strength of the scanner and the preferred trade-off between accuracy and time consumption. The technical conditions regarding image resolution and contrast mainly depend on field strength: in 0.5–1.5 T MR images the demarcations are often adjusted to something in between the hippocampus and hippocampal formation definitions. This is because the 0.5–1.5 T MR images normally lack demarcation information about the borders of the subiculum and its subregions. In MR images from 3 T

scanners and higher it is possible to separate subregions of cornu ammonis (CA1–CA4) and gyrus dentatus in the hippocampus [158].

Earlier, the author of this thesis worked within a group that included researchers in biomedical engineering and that had as aim to develop a fully automatic method for hippocampal volumetry. After having reached some promising preliminary results [159-163] the group had to leave this research thread due to lack of funding. The work with an automatic method had then already involved the development of a new manual segmentation method that had been applied to several datasets. The measurements for low dose radiation study presented here as well as those for the m44 MRI study (see *Future directions*) were performed in 2004–2005 with the purpose to develop the automatic software. After 2006, the main focus for the author’s research instead became to further develop the manual method and to apply it to other clinical data in order to illuminate clinical problems.

At an early stage our segmentation protocol used external demarcation criteria: the anterior commissure for the anterior limit, and the inferior and superior colliculi for the posterior limit of the hippocampus. This procedure makes the segmentation easier and as a rule results in better reliability, but it is not a valid measure of hippocampal volume. It is inapplicable to obliqueness like if the brain lies slightly diagonally in relation to the intracranial vault, or if the patient’s head is oblique in relation to the scanner orientation. The external anterior and posterior landmark method *a priori* misses a large part of the possibilities to detect hemispherical asymmetry. The left hippocampus generally occupies more coronal slices than the right hippocampus in females yet the right hippocampus is bigger [56]. Also, interindividual differences in the location of the hippocampi in relation to the mentioned external landmarks are clearly possible. This early method may

not have any application today but our comparison of it with the full segmentation may still be of some interest. These results are reported in the thesis but not in Paper I.

In an earlier phase of the segmentation method the thin white matter alveus was excluded. Because of partial volumes effects and low resolution it is difficult to determine the thickness of this layer on 0.5–1.5 T MR images, and in Paper I and II the alveus is included in the hippocampal segmentation (see Figure 2, Paper II).

The segmentation software was first called Hipposegm but the name of its recent versions is MIST (Medical Imaging Segmentation Tool). It was developed for manual, volumetric segmentation of brain structures – not only hippocampus and other MTL structures but also for example ICV and WML – from MRI data. Hipposegm, as used in Paper I and II, offered segmentation in coronal, sagittal and transversal views, adjustable interpolation methods, noise reduction and 3D visualizations. MIST adds segmentation of different views at multiple windows, reformatting of the image data, a capability to resize the visualization windows and to get the visualization size in accord with the actual size of the voxels, intensity adjustments, threshold segmentation and an optional random left/right display of the MR images [164]. MIST is developed in MATLAB and at the moment not for public use.

White matter lesion assessment

WML visible in MRI reflect demyelination, axonal loss, gliosis or edema however structural MRI comprises no lucid separation between these pathological substrates.

There are several visual rating methods for WML research purposes [165,166]. Among these the Fazekas visual rating [167,168] is frequently used in research and it has been shown to have good reliability compared to two other rating scales [166] but the results on its correlation to volumetric assessment diverge. For example, the Fazekas visual rating has shown the highest [165] and lowest [169] correlation to volumetric assessment compared to other visual rating scales.

Manual volumetric assessments have shown higher reliability than visual rating scales [165,169] and would be valuable in clinical settings if made less labour-intensive. In Gouw et al. [170] visual rating of WML and WML volumetry had similar correlations to neuropsychological performance, but in Garrett et al. [171] a correlation to neuropsychological performance was only found using WML volumetry and not in visual WML rating. Several segmentation and thresholding techniques have been used to manually assess WML volume in the literature but one of the few techniques with a methodological description is reported in Gurol et al. [172] and implemented in the software MRICron.

Several automatic volumetry methods have been developed for WML classification [173-175]. FreeSurfer is one of these methods and contains automatic assessment of neuroanatomical subregions as well as WML hypointensity volumes. However, few publications have reported FreeSurfer WML volumes at all and only one study reports inter-method reliability figures [176]. The latter study also reported that regional WML predicted executive dysfunction. One study found that total WML was significant for AD versus controls [177] and another found that total WML predicted functional decline almost as well as the best predicting regional WML regions [154].

2.2 Materials and methods

2.2.1 Study characteristics of the hippocampal volumetry studies

The low dose radiation study (Paper I)

In 2002, one hundred and one individuals treated for malignancies in the head and neck region were identified from the local database of the Department of Oncology at Sahlgrenska University Hospital in Gothenburg. They had received radiotherapy to the neck and base of the skull during 1992 to 1998 due to cancer in the epipharynx or oropharynx. Out of these 101, fifteen patients (11 men and 4 women, mean age 56 years, range 31-65) with no sign of recurrence participated in a final endocrinological and radiological study. The process of selection that resulted in this reduction of study size is described in detail in [178] that also reports the endocrinological data. Fifteen healthy controls matched for age, sex and BMI were recruited. Relatives or close friends were prioritized as controls in order to adjust for social status. Patients and controls all underwent an MRI examination of the brain on a Philips Gyroscan Intera 1.5 T scanner.

The MCI study (Paper II)

The Gothenburg MCI study [134] is a clinically based longitudinal project that aims at identifying neurodegenerative, vascular and stress-related disorders prior to the development of dementia. At baseline patients and controls undergo investigations including neurological, psychiatric, cognitive

screening, neuropsychological testing, MRI, SPECT, EEG, sampling of blood and CSF. The subjects are classified according to the Global Deterioration Scale (GDS) [179] where GDS 2–3 is MCI and GDS 4 mild dementia. At the biannual follow-ups, most of these investigations are repeated. MRI is done at the first (two-year) and the third (six-year) follow-up. Somatic disorders of possible neurological and neuropsychiatric relevance are noted at each visit. Controls are recruited from other medical studies and organizations for senior citizens.

Different MRI scanners have been used since the start of the study in 1999. A 0.5 T magnet (Philips NT5) was used for the part of the study presented in Paper II. The subjects (n = 68) include 21 patients with MCI both at baseline and follow-up (stable MCI or MCI-s), 21 patients with MCI who converted to mild dementia at follow-up (converting MCI or MCI-c), and 26 controls. Of the MCI-c patients 13 converted to AD and 8 to non-AD dementia. The majority (n = 48) of the subjects were scanned twice with the 0.5 T scanner: MCI-c* (n = 14), MCI-s* (n = 15) and Controls* (n = 19).

2.2.2 Methods used in the hippocampal studies

MR-imaging

Hippocampal segmentation was in both studies done on coronal slices scanned perpendicularly to the hippocampal principal axis. The following scan parameters were used.

Table 1. Scan parameters

Study Scanner	Radiation Philips Gyroscan Intera 1.5T				Mild cognitive impairment Philips NT 5 0.5T		
	Coronal	Coronal	Axial	Sagittal	Coronal	Axial	Sagittal
Orientation	Coronal	Coronal	Axial	Sagittal	Coronal	Axial	Sagittal
Slice thickness (mm)	2.4	3	2	5	1.5	1.5	4
Spacing b. slices (mm)	1.2	3	2	6	1.5	1.5	4.4
Acquisition sequence	3D T1 FFE	2D TR IR	2D T1 IR	T2 W/TSE	3D T1 FFE	3D T1 FFE	T2 TSE
Repetition time (ms)	25	2150	2500	5830	30	30	4682
Echo time (ms)	4.6	10	10	110	10	10	120
Flip angle (°)	30	90	90	90	40	40	90
Field of view (mm)	230	200	200	250	220	220	230
Pixel spacing (mm)	0.45	0.39	0.39	0.98	0.86	0.86	0.45
Matrix size	512x512	512x512	512x512	256x256	256x256	256x256	512x512

Anatomical principles

Following Jack et al. and Cook et al. [180,181], our segmentation protocol uses the hippocampal/uncal fissure as medial border, which means that only the part of subiculum contiguous with the hippocampus is included. In other segmentation protocols it is common to include the whole of subiculum in the hippocampal region and the volume estimate is then about 15% higher than estimates based on the Jack and Cook protocols. The white matter structure fimbria is the main output pathway from the hippocampus and it is excluded from the hippocampal region of interest. The thin white layer alveus is included because low resolution and partial volume effects make it difficult to demarcate from other parts of hippocampus on 0.5 T and 1.5 T scanners [182]. The alveus is a major pathway to CA1 from entorhinal cortex. Axons from CA1 through the alveus are also the main output from the hippocampus both to the pathway fimbria and to the subiculum. As mentioned above the whole hippocampus was segmented and external demarcation criteria for it were not used.

Software & hardware

An earlier version of MIST (see above) called Hipposegm was used. The coronal slices were pen segmented on an interactive Wacom PL-400 or Cintiq LCD tablet screen.

Internal landmark setting

The segmentation consists of two parts: 1) Internal landmark setting done by point-wise cross setting in the sagittal view of the reformatted coronal image. 2) Segmentation of the hippocampus in the coronal series done by continuous pen drawing and point-wise sampling.

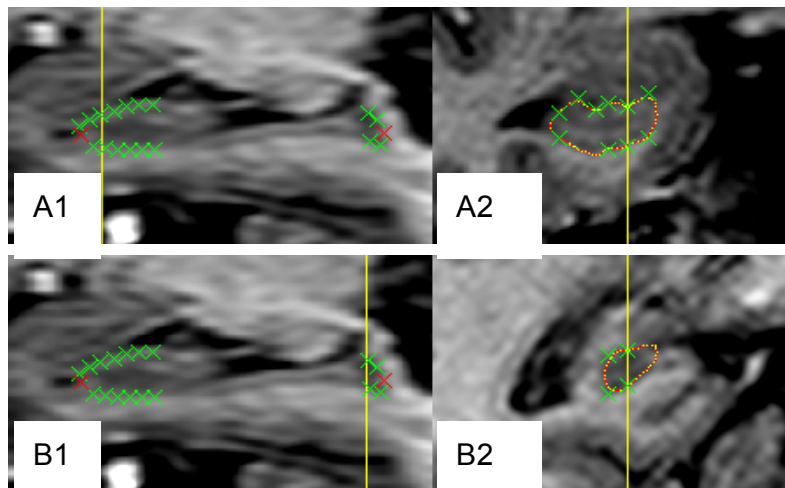


Figure 2. A1. Landmarks set in the reformatted sagittal slice of the hippocampal region. The yellow line in A1 shows the position of the coronal slice in A2. The red crosses are used for anterior and posterior limits and the green crosses are used for other limits.

A2. Green landmarks transformed into the coronal view to guide the segmentation (red line) in the anterior hippocampal head.

B1. Landmarks set in the reformatted sagittal slice of the hippocampal region. The yellow line shows the position of the coronal slice in B2.

B2. Landmarks (crosses) transformed into the coronal view to guide the segmentation (red line) in the most difficult parts of the hippocampal tail.

Landmark crosses are placed in the sagittal view in order to guide the rater where the demarcation in the original coronal image is difficult to interpret (Figure 2, from Paper I.). The most difficult parts are the anterior part where the hippocampal head reveals no border to amygdala and the posterior parts in the hippocampal tail. A careful segmentation of the hippocampal tail is important because that a substantial volume of hippocampus is in the tail and its morphology differs by sex [56]. The landmark crosses are color coded by the rater to indicate differences in certainty about their correct position.

Segmentation

The segmentation of the hippocampus starts in the body, which is the easier part, and ends in the more difficult head and tail. Starting in the first slice of the body posterior to the hippocampal head the segmentations proceed to the last slice of the tail. Then the segmentation continues, with a glance of the starting slice in the body, from the first slice of the hippocampal head to the most anterior slice.

The segmentation strategy for each particular slice is decided by balancing four aspects: 1) The hippocampal borders visible in the image, 2) The hippocampal borders in the images of contiguous slices, 3) The hippocampal borders shown in the most appropriate section in Duvernoy's atlas [157], and 4) The presence of landmark crosses in the slice.

Frequent consultations of Duvernoy's sectional anatomy of the hippocampus [157] and the segmenting guidelines elaborated by Pantel et al. [183,184], Cook et al., Jack et al. and Pruessner et al. [104,180,181] do facilitate a stable interpretation approach. In the recent survey [102] of segmentation protocols the protocol in Convit et al. [185] is the most similar to the hippocampus

segmentation protocol developed in our group but in our protocol a full segmentation of the hippocampal tail [56] is included.

3D visualization

After segmenting the whole hippocampus it is 3D visualized to check for deviations from curvature expectations. E.g. hippocampus should have a smooth curvature on the lateral side at the border between the hippocampus and the lateral ventricle. On the inferior side, the border between the hippocampus and the white matter of the parahippocampal gyrus should also be smooth. Discontinuities from the expectations regarding curvature could be a result of patient movements that results in every second slice displacement, and the rater must be open to this possibility. Unexpected discontinuities in curvature that most probably are the results of such obvious errors can then be adjusted for in the segmentations.

Intracranial volume normalization by the three orthogonal slices method

As mentioned in the section above on Normalization, several different estimates of ICV are in use ranging from simple distances *via* single slices to a subset of all slices in one direction. Using the images from the low dose radiation study for validation, a new method was developed based on the idea that three orthogonal images would capture most of the 3D shape of the intracranial cavity. The new method was not used in the radiation study (Paper I) since we there had access to a full ICV segmentation from the validation, but it was used in the MCI study (Paper II). Since only a summary is given in Paper II, the method is presented in detail here.

Using the Hipposegm software, the contours of the intracranial cavity are drawn in four slices (from three orthogonal scans) chosen in relation to landmarks where the area of the cavity cross-section can be expected to be maximal (Figure 3). Two slices (one on each side of the midline) are needed in the sagittal view in order to avoid the hemispheric sulcus and the falx cerebri; their mean area is used as the sagittal intracranial area.

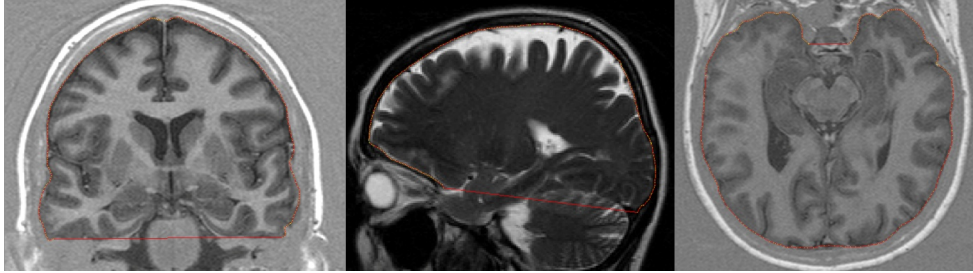


Figure 3. Coronal slice, one of the two sagittal slices and transversal slice used in the Orthogonal ICV method.

The resulting three coronal, sagittal (mean of two) and transversal areas were denoted ICA_{cor} , ICA_{sag} and ICA_{trans} and were applied in the formula:

$$(1) \quad ICV_{ell} = \frac{4}{3\sqrt{\pi}} (ICA_{cor} * ICA_{sag} * ICA_{trans})$$

The rationale of this is that the formula for the volume V_{ell} of an ellipsoid with radii a, b and c respectively is:

$$(2) \quad V_{ell} = 4/3 * \pi * a * b * c$$

If three ellipses have radii (a, b), (b, c) and (a, c) respectively, the product P of their areas becomes

$$(3) \quad P = \pi^3 * a^2 * b^2 * c^2$$

and the square root of P is identical with the expression in (2) except for a scaling factor $\frac{4}{3\sqrt{\pi}}$. Hence, using the simplifying assumption that ICA_{cor}, ICA_{sag} and ICA_{trans} are ellipses that pairwise share two diameters, (1) means approximating the intracranial cavity with an ellipsoid having its main axes parallel to the normals of the segmented slices.

The relation between ICV_{ell} and “true” ICV was investigated in connection with the low dose radiation study. ICV_{ell} as assessed by two raters in the 30 subjects was compared with the total intracranial volume (ICV_{tot}) as calculated from full segmentations by the same raters, using sagittal slices with 6 mm spacing. As reported in the Results section below, ICC of both measures as well as their intercorrelation was very high. The linear regression equation relating ICV_{ell} and ICV_{tot} was therefore used in the MCI study to get a final estimate of true ICV from ICV_{ell}.

In both studies, we followed the normalization procedure described by Jack et al. [70]. This means calculating the regression of the left and right hippocampal volume on the ICV estimate that one uses (for simplicity abbreviated ICV in formula 4). The underlying assumption is that these regressions catch most of the variance in hippocampal volume due to differences in intracranial volume. In the MCI study we used the baseline

scan of the control subjects (n = 26) for the regressions. In the radiation study, the whole material was used. To get a normalized volume V_{norm} from an absolute volume V_{abs} , the following formula was applied:

$$(4) \quad V_{\text{norm}} = V_{\text{abs}} - k * (\text{ICV} - \text{Mean}(\text{ICV}))$$

where k is the detected slope of the regression line (the regression coefficient) and $\text{Mean}(\text{ICV})$ refers to the mean estimated ICV in the whole material.

Note that because a linear function of a linear function is another linear function, one may use ICV_{ell} as ICV in the normalizing formula (4).¹

Today it is common to use a direct proportionality formula for normalizing. However, the regression method accords better with the true relation between hippocampal and intracranial volumes. It does not produce spurious correlations between normalized volumes and ICV and may be more robust to measurement errors than the proportional method [75].

Our method for ICV estimation will be compared with other alternatives in an ongoing study (see *Further directions*).

¹ There are two inconsequential errors in Paper II concerning ICV. First, the formula for ICV_{ell} is misprinted on p. 52 (but the correct formula was used). Second, ICA calculations were done using the parameter *pixel spacing* instead of *spacing between slices*, which means that a scaling error was introduced in formula (4) (p. 37).

However, for the reason just given the latter error does not affect the normalization in any way.

2.2.3 Study characteristics of the WML study

From the Gothenburg mild cognitive impairment (MCI) study (see above), 124 subjects between 40 and 86 years of age (Mean 65.6, SD 7.7 years) with subjective or objective cognitive impairment were recruited together with 28 controls. Like in the hippocampus MCI substudy, subjects with a GDS 2 or GDS 3 score received an MCI classification. MCI subjects remaining in the GDS 2 to GDS 3 range at the first follow-up were classified as “MCI stable” while MCI subjects receiving a GDS 4 (mild dementia) or higher classification at follow-up were classified as “MCI converting”. Of the 128 patients, 69 were MCI stable, 9 were MCI converting and 46 had dementia at baseline.

2.2.4 Methods used in the WML study

The three different types of WML assessment methods chosen in the present WML study consist of 1) visual rating, 2) manual segmentation and thresholding, and 3) automatic volumetry. Visual rating can be apprehended as an intermediate form between total WML and regional WML assessment in the sense that only the largest focal occasion of WML is rated. In contrast to the visual rating the volumetric methods compared can be used for total WML assessment as well as for voxel-based general linear model analysis of association with e.g. episodic memory and executive function as in Smith et al. [176]. Hence the visual rating can be considered narrower than the more versatile volumetric methods.

The visual rating was performed with the Fazekas visual rating scale [167] with modifications as described below and in Paper III. The scale was chosen because of its simplicity and fastness. Manual segmentation and thresholding

was performed with the MRICron software [172] and was done because it is commonly regarded as the golden standard. The method was adapted from Gurol et al. [172] and is one of few manual WML assessment methods reported in sufficient detail. A modification was deemed necessary to deal with the intensity inhomogeneities in the images (see below). The automatic method used was FreeSurfer [94,95] and it was chosen because of its common use for brain MRI parcellation together with the further need to validate its classification of WML.

WML was measured on the first 1.5 T MR imaging acquisitions from a Siemens Symphony scanner in the Gothenburg MCI study between the years 2005 and 2007. Axial T2 weighted images were used for the Fazekas rating, coronal FLAIR images were used in the MRICron software and coronal T1 weighted images were analyzed with the FreeSurfer package. All raters and operators were blinded for the identity and cognitive status of the subjects.

Table 2. Scan parameters

Study Scanner	White matter lesion		
	Siemens symphony 1.5T		
Orientation	Coronal	Axial	Coronal
Slice thickness (mm)	5	5	1
Spacing between slices (mm)	5.5	6	1
Acquisition sequence	2D FLAIR	T2 2D TSE	T1 3D MPR
Repetition time (ms)	8780	5410	1610
Echo time (ms)	112	104	2.38
Flip angle (°)	150	150	15
Field of view (mm)	230x187	230x187	250x203
Pixel spacing (mm)	0.45x0.45	0.45x0.45	0.49x0.49
Matrix size	512x416	512x416	512x416

Visual rating

WML severity was rated using a modified Fazekas scale [167,168]. In accordance with the recommendations by Fazekas et al. periventricular WML and deep WML were not rated separately [186]. Further, in contrast to the original four-graded scale only three grades were used as in Inzitari et al. [168]. In the present study grade 1 therefore also included possible absence of WML. For each subject the slice with the largest WML occurrence visible was used to determine the WML load. The assessments were performed independently by two raters (JB and JE) who compared the images with the template images of different grades of WML given by [168]. The template images were in axial orientation. Metrical measurements were not used and only cerebral WML were included in the ratings.

MRICron

Cerebral WML segmentation and intensity thresholding were performed on coronal 5 mm FLAIR images using the MRICron software and the method described by Gurol et al. [172]. The segmentation method consists of an initial rough manual demarcation of WML followed by thresholding. In the present material, artifactual image intensity differences between slices and between series (using the same FLAIR sequence) were common and no automatic intensity normalization between patients was done. Instead the MRICron grayscale mapping was used on a window setting containing all brain tissue in a certain slice (containing the quadrigeminal plate). Due to intensity inhomogeneities, the method adapted from Gurol et al. [172] required a modification towards a more WML specific manual segmentation. One rater (EO) demarcated 152 subjects, while a second rater (JB)

independently demarcated 27 randomly selected subjects for the determination of inter-rater reliability.

FreeSurfer

Hypointensity volumetrics determined as WML were estimated by an operator (NK) running the FreeSurfer analysis (stable release version 4.0.5). Manual edits were done by the operator (NK) to reduce inaccuracies in white and grey matter classification. The segmentation process has been described in detail elsewhere [94,95].

2.3 Results

2.3.1 Methodological results of the hippocampal studies

Validation of the method

ICV segmentation and estimation: reliability

The full ICV segmentations in the low dose radiation study had a raw correlation (Pearson's r) of 0.987 between two raters (CE and EO). The interrater reliability in terms of ICC (two-way mixed model, average measure

reliability, absolute agreement) was 0.992. The mean value of the two raters' full ICV estimates was used as the ICV measure in the study.

The correlation (Pearson's r) between two raters' (AW and EO) ICV estimate from the four-slice method, ICV_{ell} , in the same material was 0,998. This estimate was not further used in the study.

In the MCI study, the intrarater reliability for the ICV estimate ICV_{ell} (single measure absolute agreement ICC, based on data from 26 subjects) was 0.986.

Hippocampal volumes: reliability and variance

In the low dose radiation study, the correlation (Pearson's r) between EO's and CE's final measurements of total hippocampal volumes was 0.854 and the intraclass correlation (ICC; two-way mixed model, single measure reliability, absolute agreement) was 0.764. Consistency ICC (two-way mixed model, single measure absolute agreement reliability) was 0.852.

In the MCI study, the intrarater reliability for the hippocampal segmentation (based on 30 repeated segmentations by CE) was as follows: Pearson's $r = 0.937$, single measure absolute agreement ICC = 0.712. The interrater reliability (CE's second segmentation vs EO's segmentation of the same 30 series) was: Pearson' $r = 0.935$, single measure absolute agreement ICC = 0.663.

In the low dose radiation study, ICV normalization reduced the overall variance in total hippocampal volume with 46% (Paper I, Table 4).

External or internal landmarks: a comparison

Measurements using external landmarks were done early in the low dose radiation study but were not reported in Paper I because of their low validity. The correlation (Pearson's r) between EO's full hippocampus segmentations and the segmentations using external landmarks was only 0.55 for right or left hippocampus taken separately. Using the external landmarks, the full volumes were cropped in order to allow a better comparison with the earlier volumes found using external landmarks. The correlations between the cropped volume and the old volumes were 0.87 for right hippocampus and 0.91 for the left. The correlation of the remaining anterior or posterior volumes of the whole hippocampus with each other and with the old volumes (using external landmarks) were however all weakly *negative*. The number of hippocampal slices anterior to the anterior landmark and the number of slices posterior to the posterior landmark were also negatively correlated to each other. For a possible explanation of this see Discussion.

2.3.2 Clinical results of the hippocampal studies

The low dose radiation study

We did not find any significant difference with respect to left, right or total hippocampal volume between patients and controls. The mean differences were all close to zero and the two-tailed 95% confidence interval for normalized total hippocampal volume does not include more than an 8% lower mean volume for the patients. Normalizing eradicated the differences in absolute volumes between men and women. An expected age trend was found. For more results and numerical data, see Paper I.

The MCI study

Hippocampal volumetry could predict conversion to dementia in both the AD and the non-AD subgroup of converters. Contrary to some earlier studies, it was found that left hippocampal volume had the best predictive power. Cut off points for individual discrimination were shown to be potentially useful. The converting MCI group had a significantly higher rate of hippocampal volume loss than the stable MCI group. For more results and numerical data, see Paper II.

An interesting side finding emerged from the puzzling observation that the control group had a *higher* mean rate of hippocampal volume loss between the baseline and follow-up examinations than any of the patient groups. An inspection of the medical journals showed that the high mean volume loss was almost entirely explained by the data from a subgroup of six control subjects that experienced considerable medical and/or psychiatric problems, such a breast cancer or stroke, after their inclusion. Such problems were clearly overrepresented in the control group. The finding still remains to be followed up with a comprehensive analysis of the differences between the groups.

2.3.3 Results of the WML study

Regarding inter-rater reliability, the Spearman rho coefficient was 0.89 for the modified Fazekas visual rating scale and 0.60 for the manual MRIcron WML volumetry. The rho value was 0.65 for the inter-method correlation between FreeSurfer automatic WML volumetry and manual WML volumetry (Paper III, Table 3). No significant systematic differences between raters

were found for Fazekas visual rating or for MRICron manual volumetry in the Wilcoxon matched pairs test.

In order to evaluate reliabilities in the dense and the sparse parts of the data distribution (Paper III, Figure 2), respectively, the WML volumes were separated into tertiles (Table 3, below). For the aggregated lower two tertiles the Fazekas inter-rater reliability (Spearman's rho) measured 0.65 and for the upper WML tertile 0.92; this difference was significant. The manual MRICron inter-rater reliability was non-significant for the lower two tertiles, which make up about 10 percent of the whole volume range, but significant for the upper WML tertile with an inter-rater reliability of 0.94. When separating the lower aggregated WML tertiles into an anterior and a posterior part with respect to their location in the brain a significant reliability could be seen for the posterior part with a rho value of 0.56. There was still no significant reliability for the anterior part. For the inter-method reliability between the manual MRICron volumetry and the automatic FreeSurfer volumetry the rho value was also lower, 0.38 in the lower aggregated tertiles than in the upper tertile, 0.74.

Table 3. WML reliability

	Kendall's tau-b	ICC all tertiles	ICC below T3	ICC in T3	Rho all tertiles	Rho below T3	Rho in T3
Inter-rater reliability							
Fazekas	0.88**	0.87**	0.65**	0.86**	0.89**	0.65**	0.92**
Manual WML Total	0.45*	0.97**	0.19	0.95**	0.60**	-0.03	0.94**
Manual WML Anterior					0.63**	0.23	0.94**
Manual WML Posterior					0.82**	0.56*	1.00**
Inter-method reliability							
Manual WML vs Automatic WML	0.48**	0.51**	0.14**	0.44**	0.65**	0.38**	0.74**

* $p < 0.05$, ** $p < 0.01$

Bland-Altman (BA) plots were also used to illustrate the absolute volume inter-rater and inter-method differences. For the results see Paper III, especially Figure 1.

The volumes found by FreeSurfer automatic volumetry differed significantly from those of the manual MRICron volumetry. Almost all FreeSurfer volumes were lower than the corresponding manual MRICron volumes (Paper III, Figure 1). A visual inspection of a number of scans revealed some possible explanations for this, but it could also be seen that FreeSurfer was more inclusive in some respects. Since the results of the visual inspection are tentative they are not reported in this section but only in the *Discussion*.

3 DISCUSSION

3.1 Discussion of the hippocampal studies

Reliability and validity

The interrater reliability is similar in the two hippocampal studies in spite of differences in image quality, disease and rater experience. This can be interpreted as an indication of robustness in the segmentation method.

The data from the material in the low dose radiation study clearly show that a hippocampal volumetry method based on external landmarks is defective. A segmentation protocol with easily discernible landmarks for anterior and posterior borders can be expected to enhance reliability figures but the accuracy may nevertheless be poor [80]. However, the zero or even negative correlations found between part-volumes in the segmentation of the whole hippocampus indicate that marked differences in the anterior/posterior location of the hippocampus in relation to the landmarks were present. If a hippocampus is for example translated forwards, the volume anterior to the anterior landmark grows but the volume behind the posterior landmark shrinks. At the same time, the correlation between the “old” volumes and the cropped “new” (full) volumes was high, indicating a high reliability of the measurements.

The reliability of our full hippocampus segmentation method, as measured with absolute measure intraclass correlation in the MCI and low dose radiation studies, is not as high as one should desire. However, in both cases

there are known circumstances that lower the reliability. The method was being developed when the low dose radiation study was performed, and the second rater was relatively inexperienced when he segmented the radiation data. More importantly, there was a considerable systematic difference between the raters. Since the main conclusion of the paper is based on differences between study subjects, any systematic difference between raters should tend to cancel out in the calculations. In other words, consistency ICC or even simple correlation may be better measures of the relevant kind of reliability than absolute measurement ICC.

One of the highest reported reliability figures in recent years is from the study by Malykhin [187], surprisingly inter-rater ICC was excellent but intra-rater ICC was not. Intra-rater reliability can be expected to be higher than inter-rater reliability due to a smaller systematic bias between sessions for a rater than between raters. In our MCI study intra-rater reliability was marginally higher than inter-rater reliability. However, several studies report higher inter-rater than intra-rater reliability, which awaits an explanation [102].

In the MCI study, the material chosen for the reliability calculations included the baseline and follow-up scans of 14 patients who progressed to dementia. For both studies, the lack of really high reliability is to a considerable extent compensated by the agreement with external criteria (predictive power or criterion validity, see next section).

The reliability of our method for estimating ICV is excellent, and so is its correlation with a total segmentation. Its accuracy compared to other manual and automatic methods (cf. above) is an open issue that is being investigated in an ongoing study (see *Future directions*).

Power and sample size: some general considerations

Power calculations (see Introduction, section 6) are recommended before starting a scientific study in order to determine a suitable sample size, but no formal calculation was performed for any of the two hippocampal studies. However, there are other considerations determining sample size. At times where a limited but clinically important sample is available it can be wrong not to investigate it, especially if the study does not have any negative consequences for the involved subjects. In the case of Paper II, all available 0.5 T patient data from a large dataset of already performed MR investigations were used. Since the study got significant results its power does not matter for the interpretation of the results [188].

It is always more difficult to evaluate a study with a negative main finding and if its power is low the finding could very well be due to chance. Although retrospective power calculations should be avoided [188], the sample size in Paper I is arguably low. The sample size was to a large degree due to the strict exclusion criteria used, which ought to have considerably reduced true (as opposed to methodological) variability. The careful matching of patients and controls also probably improved power by lowering variability. Consistency ICC turned out to be close to excellent in spite of the second rater's relative inexperience. The best argument that the negative main result is not due to chance is, however, the significant and expected findings concerning sex, age and laterality. These findings show the criterion validity of the method, i.e. its agreement with external measures. The same set of findings in the MCI hippocampus study (Paper II) also strengthen the probability that the main result of that study is valid.

Clinical significance

In paper I the main finding was that no difference in hippocampal volume was found between the low dose irradiated group and the control group. Hence the reduced quality of life in the patient group was probably not due to a substantial volumetric reduction in hippocampal volume. Other possible kinds of damage to the hippocampus were not investigated. Yet another low dose radiation induced damage explaining the reduced quality of life might be leukoencephalopathy, mediated by damage to carotid artery or intracranial vessels [189]. Finally, psychogenetic mechanisms cannot be excluded, but findings from other cancer research tell against such an explanation; see [178] for a full discussion of this.

In paper II the main finding was that left hippocampal volume best predicts which MCI patients will subsequently convert to dementia. This seems to be the typical pattern in AD but with subsequently reduced left right hippocampal differences in disease progression [190]. MCI patients converting to non-AD dementia also had significantly smaller left hippocampal volumes than the stable MCI patients. Altogether the left hippocampus may be an early predictor of dementia regardless of dementia type.

Volumetry of the hippocampus together with volumetry of other MTL structures and/or other biomarkers has a potential to become a sensitive detector of the initial and intermediate neuropathological stages of Alzheimer's disease. In a study from our group, hippocampal volumetry together with other biomarkers has been shown to be a more sensitive predictor than the biomarkers alone [191]. According to Devanand et al. entorhinal and hippocampal volumetry both contribute to the prediction of dementia in MCI patients [139]. Entorhinal but not hippocampal volumes

have in two studies been found to differ significantly between controls and MCI subjects, but both entorhinal and hippocampal volumes differed between MCI and AD patients [192,193]. This indicates entorhinal volumetry as an earlier biomarker than hippocampal volumetry which is in agreement with the Braak stages of the major pathogenesis in AD [33]. In a recent study right entorhinal cortical thickness assessment was found to be the best single predictor of MCI to AD conversion among neurochemical, psychometric and imaging biomarkers [194].

3.2 Discussion of the WML study

Reliability

Conventional inter-rater reliabilities as measured with Spearman's rho were acceptable for the Fazekas rating and the manual MRIcron volumetry. The inter-method reliability for the manual MRIcron and automatic FreeSurfer methods was also acceptable. It is common to evaluate reliability in WML research with intraclass correlation (ICC) with excellent results e.g. in Gao et al. and Smith et al. [165,176]. However, it is misleading to use such an analysis in a data structure where the distribution is skewed, with very sparse data points in the upper third of the volume range, as was the case with WML in the present sample. While showing an excellent reliability for the whole sample, like previous studies, the ICC in the lower aggregated tertiles for the manual method was non-significant (Table 3 in Paper III). The density of low burden WML (Figure 2, Paper III) probably represents a very common clinical distribution and reliability analysis must under these conditions be performed and interpreted with caution.

The finding in the present study of lower reliability in the Fazekas rating with lower WML burden is congruent with the finding by Wardlaw et al. [195] where cohorts with lower WML burden showed lower reliability. The higher reliability for high WML burden has been considered as a ceiling effect but there might as well be floor effects in visual rating [196]. Presumed ceiling and floor effects do not disqualify visual rating as a candidate for clinical use but such effects would in general lower the usefulness for WML research, e.g. the possibility to find valid correlations with psychometrics.

The non-significant Spearman correlation in the aggregated lower tertiles between manual MRIcron volumetry raters could possibly be due to the high density of data affecting the rank order of the ratings. It is unclear if the high Spearman correlation for high burden WML implies an overestimation of the reliability due to the sparse data density (leading to less error with ranked data) or if there really was a higher reliability as the BA plot of fractional differences seems to imply (Figure 3 in Paper III, right). In our opinion the low reliability in low burden WML is most probably not due to the combination of high density of data and the nature of the Spearman correlation. Rather it may be a real rater- and inter-method variation as is visible in the BA plots, which seem to indicate an increased variation in the lowest quarter of the volume range.

Intensity variations

The higher rater variation in low WML may be due to difficulties in the handling of intensity distortion affecting the thresholding step in the manual volumetry. Intensity inhomogeneities in MR images and variation in grayscale level between scan series result in inconsistent classifications of hyperintensities that in turn introduce measurement errors in the manual

WML assessment. In order to assess WML volumes as accurately as possible under conditions with varying intensity levels through the image slices of a subject, a methodology was chosen where the thresholding was adjusted as a compromise to best fit the visible WML volume through all image slices of a subject's brain. Since the localization and extent of WML vary, the accuracy of the thresholding can be expected to be decreased by these intensity distortions. A particular shift in grayscale level was observed in the anterior part of all FLAIR image series (Paper III, Figure 4), which may be due to gradient eddy currents or cross talk between 2D FLAIR slices [16,17]. The unreliability for low WML seems to emanate mainly from the anterior half of the image data and the intensity shift in the anterior part may well be a reason for more complex considerations in the thresholding of low WML burden cases. For higher WML volumes the thresholding step in general only affects the amount of WML selected, but for lower volumes the thresholding more often affected the presence or absence of WML in a slice. The thresholding at higher WML volumes could therefore be expected to be more straightforward. In short, increased complexity in the thresholding may be the main cause of an increased rater variation for low WML volumes.

Differences between FreeSurfer and manual volumetry

The FreeSurfer suite includes intensity normalization steps which to some extent will limit intensity distortion but that also may extinguish some of the hypointensities. Further, the T1 MR sequence used for the FreeSurfer volumetry does in general have lower WML definition and somewhat smaller areas of WML hypointensities compared to the hyperintensities in the FLAIR sequence. A visual inspection of the T1 weighted images and the FreeSurfer segmentations confirmed that the substantial deviations between the

volumetric methods in the detection of WML, in various locations might mainly be due to the lower definition in the T1 weighted images. The FreeSurfer segmentation often omitted large amounts of deep WML seen in the FLAIR images (Paper III, Figure 5) and on a few occasions classified sulcal cortical voxels as WML. Occasionally FreeSurfer also detected more WML in the periventricular region than visible in the T1 weighted images. Compared to WML seen in FLAIR images the amount of WML found by FreeSurfer still was lower even in the periventricular regions. The T1 weighted MR images in Paper III, Figures 5 and 6 are examples showing that the noise level makes the detection of punctate WML a difficult task, nevertheless FreeSurfer detects several punctate WML patches in this region. The single punctate WML patch in the FLAIR slice in Figure 6 is hardly visible in the T1 slice and it conceivable that no punctate WML patch in this location is detected in FreeSurfer. In some cases FreeSurfer detected small punctate WML not detected in the manual volumetry (Figure 6 middle row), but often it detected less punctate WML than the manual volumetry (Figure 6 lower row). Hence FreeSurfer seems to detect punctate WML in a somewhat random way. In short, the FreeSurfer volumetry generally detects less WML, which is likely due to a combination of the method, the characteristics of WML and the visibility of WML in T1-weighted images, and leads to larger differences between the methods when measuring larger WML volumes (Paper III, Figure 1).

4 CONCLUSIONS

Among the various sources of variation and differences in structural MRI assessment some known systematic differences could be used to confirm an interpretation of the result that low dose radiation gave no volumetrically discernible damage to the hippocampus.

Hippocampal atrophy is present in a variety of diseases and hippocampal volumetry can be apprehended as an unspecific disease marker. Especially left hippocampus seems to be an early dementia biomarker but with the largest effect size in AD.

If ICV is estimated with a reliable and valid method, ICV normalization can greatly reduce overall volumetric variance and contribute significantly to the discrimination between groups.

WML assessment in clinical samples stands in need of refined methodology and reliability analysis. Intraclass correlation should not be used in strongly skewed samples and the kind of ICC used should always be fully specified.

5 FUTURE DIRECTIONS

The author intends to further develop the methods for hippocampal volumetry and WML estimation in cooperation with several groups of researchers. Within the Gothenburg MCI study three upcoming substudies, lead by the author of the thesis, are under progress. One of these substudies are almost completed (Study 1) and a pilot study has been performed in yet another (Study 3). In collaboration with other research groups a number of studies (Studies 4-7) are ongoing. In these, hippocampus and MTL volumetry are part of the design. One of these studies (Study 4) is also near completion. After describing all separate ongoing or planned studies below, a section will follow that summarizes some methodological developments involved in them.

Study 1: The diagnostic value of WML estimation methods in suspected dementia

The WML estimation methods discussed in this thesis have been related to the different diagnostic groups involved in the sample and their discriminative power compared. The results from this comparison are planned to be published in a follow-up article to the one presented here (Paper III).

Study 2: Optimizing manual WML volumetry

Diffusion tensor imaging is a promising tool for WML assessment but is not a universal solution. The study of Zhan et al. of non-demented elderly, indicates difficulties for diffusion imaging to detect pathological tissue with high lesion severity as assessed by FLAIR WML intensities. This finding is congruent with the proposal of Zhang et al. to use multimodal MRI assessment for improved differential dementia diagnoses [197] and the suggestion by Cherubini et al. [23] that DTI and T2* relaxation together would better characterize myelin structure in white matter [23]. In the light of this and the fact that DTI is not always available it is important to optimize the methods for WML assessment on structural MRI. The planned study will involve updating MIST with an intensity normalization algorithm and algorithms for evaluating WML of different kinds and at different locations. The work will be done in cooperation with Niklas Klasson.

Study 3: MTL volumetry in 1.5T MR images of MCI patients

The MCI hippocampus study will continue with manual hippocampal volumetry on 1.5 T MR data from more than 200 patients and controls. The material overlaps with that of the WML study presented in this thesis. The volumetric study of these MR data will also explore other structures in the medial temporal lobe. A pilot study of the transentorhinal cortex volumetry has recently been performed on 48 subjects (20 patients and 28 controls).

Study 4: The m44 hippocampus study

m44 is an epidemiological and longitudinal study of 141 men born in 1944, focusing on the metabolic syndrome, which has a possible connection with chronic stress. From the m44 subjects, two smaller groups ($n = 17$ and $n = 20$) were selected from the lowest and the highest quartile, regarding serum testosterone levels, respectively. All participants from the two smaller groups underwent an MRI examination of the brain on a Philips Gyroscan Intera 1.5 T scanner. The hippocampi have been segmented and preliminary analyses have been performed.

Study 5: The physical training study

MRI (3 T) images from around 80 healthy 12-year-old school children, half of whom received extra physical training in school, are being analyzed using FreeSurfer. ICV has been manually segmented using MIST. MTL segmentation will also be performed with our custom software. The main study hypothesis is that the group with extra training will have larger hippocampi. The study will also offer a possibility to compare an automatic and a manual method. It is done in cooperation with the Center for Brain Repair at Sahlgrenska Academy.

Study 6: The child brain tumor study

A longitudinal study of hippocampal volume changes in children treated for brain tumors with or without radiation therapy is ongoing since two years and until now, data from around 50 patients and comprising over 200 scan occasions have been collected. The main hypothesis to be tested in the study is that the hippocampal growth curves in children who have received radiation therapy, especially children who received a high radiation dose to the hippocampus, are shifted significantly downwards compared to children who have only received chemotherapy, and even more compared to normal children [108]. Shielding techniques in radiotherapy and proton therapy will be developed in the near future and hippocampal volumetry can be used to evaluate the effects. The study is done in cooperation with the section for Pediatric Oncology at Sahlgrenska Academy.

Study 7: The CogThyr study

In this study, which is ongoing since around one year, 50 women with newly discovered Graves' disease will be scanned with structural MRI of the hippocampus, whole brain and orbitae. Scans are done before, during (only orbital scans) and after treatment. Fifty matched healthy controls will be included. The subjects also undergo neuropsychological testing and neuropsychiatric interviews. The MRI results will be analyzed using FreeSurfer and our custom software. The main aim of the study is to investigate the causes of remaining cognitive symptoms after completed therapy and the main hypothesis is that women with worse cognitive symptoms will have lower hippocampal volumes. Another aim is to find MRI criteria of incipient TAO (thyroid associated ophthalmopathy). The study is done in cooperation with the Department for Endocrinology at Sahlgrenska Academy.

Beside these 7 studies, the optimization of ICV segmentation methodology is the topic of an ongoing series of studies headed by Niklas Klasson. Our custom method for ICV estimation will be further evaluated as part of this study.

Methodological developments involved in the ongoing and planned studies

Developments of WML methodology

We plan to develop a new integrated method based on our custom software MIST. Intensity normalization will be included and several alternative algorithms for this (cf. above) will be tested, since normal inter- and intra-individual differences in true MRI white matter intensity must not be leveled out. We also intend to develop an algorithm for scoring and evaluating different kinds and locations of WML. The motivation for this is that different kinds of WML probably have very different diagnostic significance. The work will be done together with Niklas Klasson.

Developments in manual MTL volumetry

In MIST we have recently implemented a procedure developed by Taylor et al. [198] for measuring the volume of the transentorhinal cortex, which is one of the regions affected in the earliest stages of AD.

We use ICV normalization in connection with hippocampal volumetry and the results of the ICV studies mentioned above will be integrated with future versions of MIST.

Comparisons of manual and automatic volumetry

In all ongoing or planned volumetric studies except the volumetric m44 study, FreeSurfer is used alongside with manual volumetry. The primary motive for this is to pick up possible changes in other structures than MTL subregions. Changes seen with FreeSurfer can then be verified using manual segmentation with MIST. Our use of FreeSurfer also offers an opportunity for systematic comparisons between it and the manual method. Such a comparison is already being performed by Niklas Klasson concerning ICV segmentation but it will be extended to hippocampal volumetry as well. This will be an especially exciting task in the physical exercise and child brain tumor studies since it is not known with certainty how well FreeSurfer performs in children. At the moment the default statistical atlas used by the FreeSurfer algorithm stems from MRI segmentations on adults. Since a child atlas is not used one might suspect that the segmentations by FreeSurfer will not be as good as manual volumetry and even worse than in adult materials, but this is yet to find out. Also, it is not well known how the method performs when the brain is grossly damaged.

6 REFERENCES

1. Starkman MN, Giordani B, Gebarski SS, Schteingart DE (2003) Improvement in learning associated with increase in hippocampal formation volume. *Biol Psychiatry* 53: 233-238.
2. Geuze E, Vermetten E, Bremner JD (2005) MR-based in vivo hippocampal volumetrics: 2. Findings in neuropsychiatric disorders. *Mol Psychiatry* 10: 160-184.
3. Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, et al. (2007) Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet neurology* 6: 734-746.
4. Visser PJ, Vos S, van Rossum I, Scheltens P (2012) Comparison of International Working Group criteria and National Institute on Aging-Alzheimer's Association criteria for Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 8: 560-563.
5. Weiner MW, Aisen PS, Jack CR, Jr., Jagust WJ, Trojanowski JQ, et al. (2010) The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement* 6: 202-211 e207.
6. Sullivan EV, Marsh L, Pfefferbaum A (2005) Preservation of hippocampal volume throughout adulthood in healthy men and women. *Neurobiol Aging* 26: 1093-1098.
7. Walhovd KB, Westlye LT, Amlie I, Espeseth T, Reinvang I, et al. (2011) Consistent neuroanatomical age-related volume differences across multiple samples. *Neurobiology of aging* 32: 916-932.
8. Csernansky JG, Hamstra J, Wang L, McKeel D, Price JL, et al. (2004) Correlations between antemortem hippocampal volume and postmortem neuropathology in AD subjects. *Alzheimer Dis Assoc Disord* 18: 190-195.
9. Bombois S, Debette S, Bruandet A, Delbeuck X, Delmaire C, et al. (2008) Vascular subcortical hyperintensities predict conversion to vascular and mixed dementia in MCI patients. *Stroke; a journal of cerebral circulation* 39: 2046-2051.
10. Staekenborg SS, Koedam EL, Henneman WJ, Stokman P, Barkhof F, et al. (2009) Progression of mild cognitive impairment to dementia: contribution of cerebrovascular disease compared with medial temporal lobe atrophy. *Stroke; a journal of cerebral circulation* 40: 1269-1274.

11. Brickman AM, Provenzano FA, Muraskin J, Manly JJ, Blum S, et al. (2012) Regional White Matter Hyperintensity Volume, Not Hippocampal Atrophy, Predicts Incident Alzheimer Disease in the Community. *Archives of neurology*: 1-7.
12. Kruggel F, Turner J, Muftuler LT (2010) Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. *NeuroImage* 49: 2123-2133.
13. Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, et al. (2009) MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *Neuroimage* 46: 177-192.
14. Huppertz HJ, Kroll-Seger J, Kloppel S, Ganz RE, Kassubek J (2010) Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. *NeuroImage* 49: 2216-2224.
15. Clark KA, Woods RP, Rottenberg DA, Toga AW, Mazziotta JC (2006) Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. *NeuroImage* 29: 185-202.
16. Simmons A, Tofts PS, Barker GJ, Arridge SR (1994) Sources of intensity nonuniformity in spin echo images at 1.5 T. *Magnetic Resonance in Medicine* 32: 121-128.
17. Wicks DAG, Barker GJ, Tofts PS (1993) Correction of intensity nonuniformity in MR images of any orientation. *Magnetic Resonance Imaging* 11: 183-196.
18. Deoni SC, Rutt BK, Peters TM (2006) Synthetic T1-weighted brain image generation with incorporated coil intensity correction using DESPOT1. *Magn Reson Imaging* 24: 1241-1248.
19. Warntjes JB, Dahlqvist O, Lundberg P (2007) Novel method for rapid, simultaneous T1, T*2, and proton density quantification. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 57: 528-537.
20. Blystad I, Warntjes J, Smedby O, Landtblom AM, Lundberg P, et al. (2012) Synthetic MRI of the brain in a clinical setting. *Acta Radiol.*
21. Jack CR, Jr., Bernstein MA, Fox NC, Thompson P, Alexander G, et al. (2008) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of magnetic resonance imaging : JMRI* 27: 685-691.
22. Stevenson VL, Parker GJ, Barker GJ, Birnie K, Tofts PS, et al. (2000) Variations in T1 and T2 relaxation times of normal appearing white matter and lesions in multiple sclerosis. *Journal of the neurological sciences* 178: 81-87.
23. Cherubini A, Péran P, Hagberg GE, Varsi AE, Luccichenti G, et al. (2009) Characterization of white matter fiber bundles with T2*

- relaxometry and diffusion tensor imaging. *Magnetic Resonance in Medicine* 61: 1066-1072.
24. Pinero DJ, Connor JR (2000) Iron in the brain: An important contributor in normal and diseased states. *Neuroscientist* 6: 435-453.
 25. Li TQ, Yao B, Van Gelderen P, Merkle H, Dodd S, et al. (2009) Characterization of T 2* heterogeneity in human brain white matter. *Magnetic Resonance in Medicine* 62: 1652-1657.
 26. Mitsumori F, Watanabe H, Takaya N, Garwood M (2007) Apparent transverse relaxation rate in human brain varies linearly with tissue iron concentration at 4.7 T. *Magnetic Resonance in Medicine* 58: 1054-1060.
 27. Panizzon MS, Fennema-Notestine C, Kubarych TS, Chen C-H, Eyler LT, et al. (2012) Genetic and environmental influences of white and gray matter signal contrast: A new phenotype for imaging genetics? *NeuroImage* 60: 1686-1695.
 28. Wang J, Shaffer ML, Eslinger PJ, Sun X, Weitekamp CW, et al. (2012) Maturation and aging effects on human brain apparent transverse relaxation. *PLoS One* 7: e31907.
 29. Oakden W, Lobaugh N, Black S, Stanisz GJ (2007) Quantitative T2 Relaxation of White Matter Hyperintensities in Probable Alzheimer Patients. *Proc Intl Soc Mag Reson Med* 15.
 30. Westlye LT, Walhovd KB, Dale AM, Espeseth T, Reinvang I, et al. (2009) Increased sensitivity to effects of normal aging and Alzheimer's disease on cortical thickness by adjustment for local variability in gray/white contrast: a multi-sample MRI study. *Neuroimage* 47: 1545-1557.
 31. Bartzokis G, Cummings JL, Sultzer D, Henderson VW, Nuechterlein KH, et al. (2003) White matter structural integrity in healthy aging adults and patients with Alzheimer disease: a magnetic resonance imaging study. *Arch Neurol* 60: 393-398.
 32. Salat DH, Chen JJ, van der Kouwe AJ, Greve DN, Fischl B, et al. (2011) Hippocampal degeneration is associated with temporal and limbic gray matter/white matter tissue contrast in Alzheimer's disease. *NeuroImage* 54: 1795-1802.
 33. Braak H, Braak E (1997) Staging of Alzheimer-related cortical destruction. *Int Psychogeriatr* 9 Suppl 1: 257-261; discussion 269-272.
 34. Sled JG, Zijdenbos AP, Evans AC (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 17: 87-97.
 35. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, et al. (2010) N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* 29: 1310-1320.

36. Hadjidemetriou S, Studholme C, Mueller S, Weiner M, Schuff N (2009) Restoration of MRI data for intensity non-uniformities using local high order intensity statistics. *Med Image Anal* 13: 36-48.
37. Robitaille N, Mouiha A, Crepeault B, Valdivia F, Duchesne S, et al. (2012) Tissue-based MRI intensity standardization: application to multicentric datasets. *Int J Biomed Imaging* 2012: 347120.
38. Vovk A, Cox RW, Stare J, Suput D, Saad ZS (2011) Segmentation priors from local image properties: without using bias field correction, location-based templates, or registration. *NeuroImage* 55: 142-152.
39. Li YJ, Ga SN, Huo Y, Li SY, Gao XG (2007) Characteristics of hippocampal volumes in healthy Chinese from MRI. *Neurol Res* 29: 803-806.
40. Brickman AM, Schupf N, Manly JJ, Luchsinger JA, Andrews H, et al. (2008) Brain morphology in older African Americans, Caribbean Hispanics, and whites from northern Manhattan. *Arch Neurol* 65: 1053-1061.
41. Utsunomiya H, Takano K, Okazaki M, Mitsudome A (1999) Development of the temporal lobe in infants and children: analysis by MR-based volumetry. *AJNR Am J Neuroradiol* 20: 717-723.
42. Mulani SJ, Kothare SV, Patkar DP (2005) Magnetic resonance volumetric analysis of hippocampi in children in the age group of 6-to-12 years: a pilot study. *Neuroradiology* 47: 552-557.
43. Szabo CA, Xiong J, Lancaster JL, Rainey L, Fox P (2001) Amygdalar and Hippocampal Volumetry in Control Participants: Differences Regarding Handedness. *AJNR Am J Neuroradiol* 22: 1342-1345.
44. Anstey KJ, Maller JJ, Meslin C, Christensen H, Jorm AF, et al. (2004) Hippocampal and amygdalar volumes in relation to handedness in adults aged 60-64. *Neuroreport* 15: 2825-2829.
45. Dellatolas G, Luciani S, Castresana A, Remy C, Jallon P, et al. (1993) Pathological left-handedness. Left-handedness correlatives in adult epileptics. *Brain* 116: 1565-1574.
46. Feczko E, Augustinack JC, Fischl B, Dickerson BC (2009) An MRI-based method for measuring volume, thickness and surface area of entorhinal, perirhinal, and posterior parahippocampal cortex. *Neurobiol Aging* 30: 420-431.
47. Huntgeburth SC, Petrides M (2012) Morphological patterns of the collateral sulcus in the human brain. *The European journal of neuroscience* 35: 1295-1311.
48. Liu T, Wen W, Zhu W, Kochan NA, Trollor JN, et al. (2011) The relationship between cortical sulcal variability and cognitive performance in the elderly. *NeuroImage* 56: 865-873.
49. Panizzon MS, Fennema-Notestine C, Eyler LT, Jernigan TL, Prom-Wormley E, et al. (2009) Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral cortex* 19: 2728-2735.

50. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, et al. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489: 391-399.
51. Eyer LT, Prom-Wormley E, Panizzon MS, Kaup AR, Fennema-Notestine C, et al. (2011) Genetic and environmental contributions to regional cortical surface area in humans: a magnetic resonance imaging twin study. *Cerebral cortex* 21: 2313-2321.
52. Peper JS, Brouwer RM, Boomsma DI, Kahn RS, Hulshoff Pol HE (2007) Genetic influences on human brain structure: a review of brain imaging studies in twins. *Human brain mapping* 28: 464-473.
53. Szeszko PR, Lipsky R, Mentschel C, Robinson D, Gunduz-Bruce H, et al. (2005) Brain-derived neurotrophic factor val66met polymorphism and volume of the hippocampal formation. *Molecular psychiatry* 10: 631-636.
54. Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, et al. (2012) Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 44: 552-561.
55. Woollett K, Maguire EA (2011) Acquiring "the Knowledge" of London's layout drives structural brain changes. *Curr Biol* 21: 2109-2114.
56. Maller JJ, Reglade-Meslin C, Anstey KJ, Sachdev P (2006) Sex and symmetry differences in hippocampal volumetrics: before and beyond the opening of the crus of the fornix. *Hippocampus* 16: 80-90.
57. Del Tredici K, Braak H (2008) Neurofibrillary changes of the Alzheimer type in very elderly individuals: neither inevitable nor benign: Commentary on "No disease in the brain of a 115-year-old woman". *Neurobiology of aging* 29: 1133-1136.
58. Raz N, Lindenberger U, Rodrigue KM, Kennedy KM, Head D, et al. (2005) Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cerebral cortex* 15: 1676-1689.
59. Malykhin NV, Bouchard TP, Camicioli R, Coupland NJ (2008) Aging hippocampus and amygdala. *Neuroreport* 19: 543-547.
60. Fanselow MS, Dong HW (2010) Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron* 65: 7-19.
61. Ta AT, Huang SE, Chiu MJ, Hua MS, Tseng WY, et al. (2012) Age-related vulnerabilities along the hippocampal longitudinal axis. *Human brain mapping* 33: 2415-2427.
62. Hayashi T, Wada A, Uchida N, Kitagaki H (2009) Enlargement of the hippocampal angle: a new index of Alzheimer disease. *Magn Reson Med Sci* 8: 33-38.
63. Adachi M, Kawakatsu S, Sato T, Ohshima F (2012) Correlation between volume and morphological changes in the hippocampal formation in Alzheimer's disease: Rounding of the outline of the hippocampal body on coronal MR images. *Neuroradiology* 54: 1079-1087.

64. Dawe RJ, Bennett DA, Schneider JA, Arfanakis K (2011) Neuropathologic correlates of hippocampal atrophy in the elderly: a clinical, pathologic, postmortem MRI study. *PLoS One* 6: e26286.
65. Barnes J, Ridgway GR, Bartlett J, Henley SM, Lehmann M, et al. (2010) Head size, age and gender adjustment in MRI studies: a necessary nuisance? *NeuroImage* 53: 1244-1255.
66. Mathalon DH, Sullivan EV, Rawles JM, Pfefferbaum A (1993) Correction for head size in brain-imaging measurements. *Psychiatry Res* 50: 121-139.
67. Szabo CA, Lancaster JL, Xiong J, Cook C, Fox P (2003) MR Imaging Volumetry of Subcortical Structures and Cerebellar Hemispheres in Normal Persons. *AJNR Am J Neuroradiol* 24: 644-647.
68. Brickman AM, Siedlecki KL, Muraskin J, Manly JJ, Luchsinger JA, et al. (2009) White matter hyperintensities and cognition: Testing the reserve hypothesis. *Neurobiol Aging*.
69. Skoog I, Olesen PJ, Blennow K, Palmertz B, Johnson SC, et al. (2012) Head size may modify the impact of white matter lesions on dementia. *Neurobiology of aging* 33: 1186-1193.
70. Jack CR, Jr., Bentley MD, Twomey CK, Zinsmeister AR (1990) MR imaging-based volume measurements of the hippocampal formation and anterior temporal lobe: validation studies. *Radiology* 176: 205-209.
71. Free SL, Bergin PS, Fish DR, Cook MJ, Shorvon SD, et al. (1995) Methods for normalization of hippocampal volumes measured with MR. *AJNR Am J Neuroradiol* 16: 637-643.
72. Bigler ED, Tate DF (2001) Brain volume, intracranial volume, and dementia. *Invest Radiol* 36: 539-546.
73. Haug H (1986) History of neuromorphometry. *J Neurosci Methods* 18: 1-17.
74. Grieve SM, Clark CR, Williams LM, Peduto AJ, Gordon E (2005) Preservation of limbic and paralimbic structures in aging. *Hum Brain Mapp* 25: 391-401.
75. Sanfilipo MP, Benedict RH, Zivadinov R, Bakshi R (2004) Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. *Neuroimage* 22: 1732-1743.
76. Obenaus A, Yong-Hing CJ, Tong KA, Sarty GE (2001) A reliable method for measurement and normalization of pediatric hippocampal volumes. *Pediatr Res* 50: 124-132.
77. Marsh L, Morrell MJ, Shear PK, Sullivan EV, Freeman H, et al. (1997) Cortical and hippocampal volume deficits in temporal lobe epilepsy. *Epilepsia* 38: 576-587.
78. Ferguson KJ, Wardlaw JM, Edmond CL, Deary IJ, MacLulich AM (2005) Intracranial area: a validated method for estimating intracranial volume. *J Neuroimaging* 15: 76-78.

79. Eritaia J, Wood SJ, Stuart GW, Bridle N, Dudgeon P, et al. (2000) An optimized method for estimating intracranial volume from magnetic resonance images. *Magn Reson Med* 44: 973-977.
80. Geuze E, Vermetten E, Bremner JD (2005) MR-based in vivo hippocampal volumetrics: 1. Review of methodologies currently employed. *Mol Psychiatry* 10: 147-159.
81. Lee N, Tien RD, Lewis DV, Friedman AH, Felsberg GJ, et al. (1995) Fast spin-echo, magnetic resonance imaging-measured hippocampal volume: correlation with neuronal density in anterior temporal lobectomy patients. *Epilepsia* 36: 899-904.
82. Bobinski M, de Leon MJ, Wegiel J, Desanti S, Convit A, et al. (2000) The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease. *Neuroscience* 95: 721-725.
83. Pfefferbaum A, Sullivan EV, Adalsteinsson E, Garrick T, Harper C (2004) Postmortem MR imaging of formalin-fixed human brain. *NeuroImage* 21: 1585-1595.
84. Van Duijn S, Nabuurs RJA, Van Rooden S, Maat-Schieman MLC, Van Duinen SG, et al. (2011) MRI artifacts in human brain tissue after prolonged formalin storage. *Magnetic Resonance in Medicine* 65: 1750-1758.
85. Donnelly LF, O'Brien KJ, Dardzinski BJ, Poe SA, Bean JA, et al. (2003) Using a phantom to compare MR techniques for determining the ratio of intraabdominal to subcutaneous adipose tissue. *AJR Am J Roentgenol* 180: 993-998.
86. Gunter JL, Bernstein MA, Borowski BJ, Ward CP, Britson PJ, et al. (2009) Measurement of MRI scanner performance with the ADNI phantom. *Med Phys* 36: 2193-2205.
87. Mori K, Hagino H, Saitou O, Yotsutsuji T, Tonami S, et al. (2002) [Effects of the volume and shape of voxels on the measurement of phantom volume using three-dimensional magnetic resonance imaging]. *Nippon Hoshasen Gijutsu Gakkai Zasshi* 58: 88-93.
88. Luft AR, Skalej M, Welte D, Kolb R, Klose U (1996) Reliability and exactness of MRI-based volumetry: a phantom study. *J Magn Reson Imaging* 6: 700-704.
89. Gouttard S, Styner M, Prastawa M, Piven J, Gerig G (2008) Assessment of reliability of multi-site neuroimaging via traveling phantom study. *Med Image Comput Comput Assist Interv* 11: 263-270.
90. Aubert-Broche B, Griffin M, Pike GB, Evans AC, Collins DL (2006) Twenty new digital brain phantoms for creation of validation image data bases. *IEEE Trans Med Imaging* 25: 1410-1416.
91. Geremia E, Clatz O, Menze BH, Konukoglu E, Criminisi A, et al. (2011) Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage* 57: 378-390.

92. Chincarini A, Bosco P, Calvini P, Gemme G, Esposito M, et al. (2011) Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *NeuroImage* 58: 469-480.
93. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, et al. (2011) Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res Notes* 4: 299.
94. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, et al. (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33: 341-355.
95. Fischl B, Salat DH, van der Kouwe AJ, Makris N, Segonne F, et al. (2004) Sequence-independent segmentation of magnetic resonance images. *NeuroImage* 23 Suppl 1: S69-84.
96. Klauschen F, Goldman A, Barra V, Meyer-Lindenberg A, Lundervold A (2009) Evaluation of automated brain MR image segmentation and volumetry methods. *Human brain mapping* 30: 1310-1327.
97. Morey RA, Petty CM, Xu Y, Hayes JP, Wagner HR, 2nd, et al. (2009) A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage* 45: 855-866.
98. Tae WS, Kim SS, Lee KU, Nam EC, 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.
99. Hasan KM (2009) A questionable gold standard for hippocampus volume and asymmetry. *Neuroradiology* 51: 201-202; author reply 203-204.
100. Augustinack JC, Huber KE, Stevens AA, Roy M, Frosch MP, et al. (2013) Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. *NeuroImage* 64: 32-42.
101. Entis JJ, Doerga P, Barrett LF, Dickerson BC (2012) A reliable protocol for the manual segmentation of the human amygdala and its subregions using ultra-high resolution MRI. *NeuroImage* 60: 1226-1235.
102. Boccardi M, Ganzola R, Bocchetta M, Pievani M, Redolfi A, et al. (2011) Survey of protocols for the manual segmentation of the hippocampus: preparatory steps towards a joint EADC-ADNI harmonized protocol. *J Alzheimers Dis* 26 Suppl 3: 61-75.
103. Pruessner JC, Kohler S, Crane J, Pruessner M, Lord C, et al. (2002) Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. *Cereb Cortex* 12: 1342-1353.

104. Pruessner JC, Li LM, Serles W, Pruessner M, Collins DL, et al. (2000) Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. *Cereb Cortex* 10: 433-442.
105. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, et al. (2012) The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 8: S1-68.
106. Wyman BT, Harvey DJ, Crawford K, Bernstein MA, Carmichael O, et al. (2012) Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimer's & dementia : the journal of the Alzheimer's Association*.
107. Shattuck DW, Prasad G, Mirza M, Narr KL, Toga AW (2009) Online resource for validation of brain segmentation methods. *Neuroimage* 45: 431-439.
108. Sanchez CE, Richards JE, Almli CR (2012) Age-specific MRI templates for pediatric neuroimaging. *Dev Neuropsychol* 37: 379-399.
109. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, et al. (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud* 48: 661-671.
110. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. (2003) The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Clinical chemistry* 49: 7-18.
111. Galea S, Tracy M (2007) Participation rates in epidemiologic studies. *Ann Epidemiol* 17: 643-653.
112. Wylie GR, Genova H, Deluca J, Chiaravalloti N, Sumowski JF (2012) Functional magnetic resonance imaging movers and shakers: Does subject-movement cause sampling bias. *Human brain mapping*.
113. Apostolova LG, Morra JH, Green AE, Hwang KS, Avedissian C, et al. (2010) Automated 3D mapping of baseline and 12-month associations between three verbal memory measures and hippocampal atrophy in 490 ADNI subjects. *NeuroImage* 51: 488-499.
114. Ward A, Arrighi HM, Michels S, Cedarbaum JM (2012) Mild cognitive impairment: disparity of incidence and prevalence estimates. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 8: 14-21.
115. Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, et al. (2009) Automated 3D mapping of hippocampal atrophy and its clinical correlates in 400 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. *Hum Brain Mapp* 30: 2766-2788.
116. Ard MC, Edland SD (2011) Power calculations for clinical trials in Alzheimer's disease. *J Alzheimers Dis* 26 Suppl 3: 369-377.

117. Perkins DO, Wyatt RJ, Bartko JJ (2000) Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. *Biological psychiatry* 47: 762-766.
118. Olsson E, Eckerström C, Berg G, Borga M, Ekholm S, et al. (2012) Hippocampal volumes in patients exposed to low-dose radiation to the basal brain. A case-control study in long-term survivors from cancer in the head and neck region. *Radiat Oncol* 7.
119. Eckerström C, Olsson E, Borga M, Ekholm S, Ribbelin S, et al. (2008) Small baseline volume of left hippocampus is associated with subsequent conversion of MCI into dementia: the Goteborg MCI study. *J Neurol Sci* 272: 48-59.
120. Olsson E, Klasson N, Berge J, Eckerström C, Edman Å, et al. (in press) White matter lesion assessment in patients with cognitive impairment and healthy controls: reliability comparisons between visual rating, a manual and an automatic volumetrical MRI method - the Gothenburg MCI Study. *J Aging Research*.
121. Darzy KH, Shalet SM (2009) Hypopituitarism following radiotherapy. *Pituitary* 12: 40-50.
122. Ricard D, Taillia H, Renard JL (2009) Brain damage from anticancer treatments in adults. *Curr Opin Oncol* 21: 559-565.
123. Balu DT, Lucki I (2009) Adult hippocampal neurogenesis: Regulation, functional implications, and contribution to disease pathology. *Neuroscience & Biobehavioral Reviews* 33: 232-252.
124. Fukuda A, Fukuda H, Swanpalmer J, Hertzman S, Lannering B, et al. (2005) Age-dependent sensitivity of the developing brain to irradiation is correlated with the number and vulnerability of progenitor cells. *J Neurochem* 92: 569-584.
125. Hellstrom NA, Bjork-Eriksson T, Blomgren K, Kuhn HG (2008) Differential recovery of neural stem cells in the subventricular zone and dentate gyrus after ionizing radiation. *Stem Cells*.
126. Nagel BJ, Delis DC, Palmer SL, Reeves C, Gajjar A, et al. (2006) Early patterns of verbal memory impairment in children treated for medulloblastoma. *Neuropsychology* 20: 105-112.
127. Nagel BJ, Palmer SL, Reddick WE, Glass JO, Helton KJ, et al. (2004) Abnormal hippocampal development in children with medulloblastoma treated with risk-adapted irradiation. *AJNR Am J Neuroradiol* 25: 1575-1582.
128. Monje ML, Vogel H, Masek M, Ligon KL, Fisher PG, et al. (2007) Impaired human hippocampal neurogenesis after treatment for central nervous system malignancies. *Ann Neurol* 62: 515-520.
129. Monje M (2008) Cranial radiation therapy and damage to hippocampal neurogenesis. *Dev Disabil Res Rev* 14: 238-242.
130. Hall P, Adami HO, Trichopoulos D, Pedersen NL, Lagiou P, et al. (2004) Effect of low doses of ionising radiation in infancy on

- cognitive function in adulthood: Swedish population based cohort study. *Bmj* 328: 19.
131. Okada M, Okabe A, Uchihori Y, Kitamura H, Sekine E, et al. (2007) Single extreme low dose/low dose rate irradiation causes alteration in lifespan and genome instability in primary human cells. *Br J Cancer* 96: 1707-1710.
 132. Löfdahl E, Berg G, Johansson K-A, Leonsson-Zachrisson M, Malmgren H, et al. Compromised quality of life in adult patients who have received a radiation dose towards the basal part of the brain. A case-control study in long-term survivors from cancer in the head and neck region. *Radiation Oncology, in press.*
 133. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, et al. (2006) Mild cognitive impairment. *Lancet* 367: 1262-1270.
 134. Nordlund A, Rolstad S, Hellstrom P, Sjogren M, Hansen S, et al. (2005) The Goteborg MCI study: mild cognitive impairment is a heterogeneous condition. *J Neurol Neurosurg Psychiatry* 76: 1485-1490.
 135. Wolf H, Jelic V, Gertz HJ, Nordberg A, Julin P, et al. (2003) A critical discussion of the role of neuroimaging in mild cognitive impairment. *Acta Neurol Scand Suppl* 179: 52-76.
 136. Anstey KJ, Maller JJ (2003) The role of volumetric MRI in understanding mild cognitive impairment and similar classifications. *Aging Ment Health* 7: 238-250.
 137. Chetelat G, Baron JC (2003) Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. *Neuroimage* 18: 525-541.
 138. de Leon MJ, Mosconi L, Blennow K, DeSanti S, Zinkowski R, et al. (2007) Imaging and CSF studies in the preclinical diagnosis of Alzheimer's disease. *Annals of the New York Academy of Sciences* 1097: 114-145.
 139. Devanand DP, Pradhaban G, Liu X, Khandji A, De Santi S, et al. (2007) Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology* 68: 828-836.
 140. Tapiola T, Pennanen C, Tapiola M, Tervo S, Kivipelto M, et al. (2008) MRI of hippocampus and entorhinal cortex in mild cognitive impairment: a follow-up study. *Neurobiol Aging* 29: 31-38.
 141. Roman GC, Erkinjuntti T, Wallin A, Pantoni L, Chui HC (2002) Subcortical ischaemic vascular dementia. *Lancet Neurol* 1: 426-436.
 142. Debette S, Markus HS (2010) The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ* 341: c3666.
 143. Inzitari D, Pracucci G, Poggesi A, Carlucci G, Barkhof F, et al. (2009) Changes in white matter as determinant of global functional decline in older independent outpatients: three year follow-up of LADIS (leukoaraiosis and disability) study cohort. *BMJ* 339: b2477.

144. Erkinjuntti T, Inzitari D, Pantoni L, Wallin A, Scheltens P, et al. (2000) Research criteria for subcortical vascular dementia in clinical trials. *J Neural Transm Suppl* 59: 23-30.
145. Levy-Cooperman N, Ramirez J, Lobaugh NJ, Black SE (2008) Misclassified tissue volumes in Alzheimer disease patients with white matter hyperintensities: importance of lesion segmentation procedures for volumetric analysis. *Stroke; a journal of cerebral circulation* 39: 1134-1141.
146. Schmahmann JD, Smith EE, Eichler FS, Filley CM (2008) Cerebral white matter: neuroanatomy, clinical neurology, and neurobehavioral correlates. *Annals of the New York Academy of Sciences* 1142: 266-309.
147. Eckerström C, Olsson E, Klasson N, Bjerke M, Gothlin M, et al. (2011) High white matter lesion load is associated with hippocampal atrophy in mild cognitive impairment. *Dementia and geriatric cognitive disorders* 31: 132-138.
148. Kim KW, MacFall JR, Payne ME (2008) Classification of white matter lesions on magnetic resonance imaging in elderly persons. *Biological psychiatry* 64: 273-280.
149. Debette S, Bombois S, Bruandet A, Delbeuck X, Lepoittevin S, et al. (2007) Subcortical hyperintensities are associated with cognitive decline in patients with mild cognitive impairment. *Stroke; a journal of cerebral circulation* 38: 2924-2930.
150. Marquine MJ, Attix DK, Goldstein LB, Samsa GP, Payne ME, et al. (2010) Differential patterns of cognitive decline in anterior and posterior white matter hyperintensity progression. *Stroke; a journal of cerebral circulation* 41: 1946-1950.
151. van Straaten EC, Harvey D, Scheltens P, Barkhof F, Petersen RC, et al. (2008) Periventricular white matter hyperintensities increase the likelihood of progression from amnesic mild cognitive impairment to dementia. *Journal of neurology* 255: 1302-1308.
152. Yoshita M, Fletcher E, Harvey D, Ortega M, Martinez O, et al. (2006) Extent and distribution of white matter hyperintensities in normal aging, MCI, and AD. *Neurology* 67: 2192-2198.
153. DeCarli C, Fletcher E, Ramey V, Harvey D, Jagust WJ (2005) Anatomical mapping of white matter hyperintensities (WMH): exploring the relationships between periventricular WMH, deep WMH, and total WMH burden. *Stroke; a journal of cerebral circulation* 36: 50-55.
154. Wakefield DB, Moscufo N, Guttmann CR, Kuchel GA, Kaplan RF, et al. (2010) White matter hyperintensities predict functional decline in voiding, mobility, and cognition in older adults. *Journal of the American Geriatrics Society* 58: 275-281.
155. Hachinski V, Iadecola C, Petersen RC, Breteler MM, Nyenhuis DL, et al. (2006) National Institute of Neurological Disorders and Stroke-

- Canadian Stroke Network vascular cognitive impairment harmonization standards. *Stroke; a journal of cerebral circulation* 37: 2220-2241.
156. Moorhouse P, Rockwood K (2008) Vascular cognitive impairment: current concepts and clinical developments. *Lancet neurology* 7: 246-255.
 157. Duvernoy H (2005) *The Human Hippocampus*: Springer.
 158. Wieshmann UC, Free SL, Stevens JM, Shorvon SD (1998) Image contrast and hippocampal volumetric measurements. *Magnetic Resonance Imaging* 16: 13-17.
 159. Borga M, Mamgren H, Knutsson H. FSED - feature selective edge detection; 2000; Barcelona, Spain. pp. 229-232.
 160. Lindqvist H (2005) Automatisk estimering av intrakraniell volym från MR-bilder. Linköping University.
 161. Pettersson J, Knutsson H, Malmgren H, Borga M (2006) Automatic segmentation of CT and MR Volume Data Using Non-Rigid Morphon Registration. In: Petterson J Automatic Generation of Patient Specific Models for Hip Surgery Simulation, Thesis, Linköping university.
 162. Starck G, Borga M, Friberg M, Olsson E, Ribbelin S, et al. (2002) Fully automatic segmentation of the hippocampus in MR images. *ESMRMB 2002: 19th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology*.
 163. Wrangsjö A. KH (2005) Morphons and Brains – Non-rigid registration for atlas-based segmentation of MRI volumes. *Proc 13:th Nordic Baltic Conference on Bio-Medical Engineering and Medical Physics*.
 164. Styner M, Smith R, Graves M, Mosconi M, Peterson S, et al. (2007) Asymmetric Bias in User Guided Segmentations of Brain Structures.
 165. Gao FQ, Swartz RH, Scheltens P, Leibovitch FS, Kiss A, et al. (2011) Complexity of MRI white matter hyperintensity assessments in relation to cognition in aging and dementia from the Sunnybrook Dementia Study. *J Alzheimers Dis* 26 Suppl 3: 379-388.
 166. Kapeller P, Barber R, Vermeulen RJ, Ader H, Scheltens P, et al. (2003) Visual rating of age-related white matter changes on magnetic resonance imaging: scale comparison, interrater agreement, and correlations with quantitative measurements. *Stroke; a journal of cerebral circulation* 34: 441-445.
 167. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA (1987) MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR American journal of roentgenology* 149: 351-356.
 168. Inzitari D, Simoni M, Pracucci G, Poggesi A, Basile AM, et al. (2007) Risk of rapid global functional decline in elderly patients with severe cerebral age-related white matter changes: the LADIS study. *Archives of internal medicine* 167: 81-88.

169. van Straaten EC, Fazekas F, Rostrup E, Scheltens P, Schmidt R, et al. (2006) Impact of white matter hyperintensities scoring method on correlations with clinical data: the LADIS study. *Stroke; a journal of cerebral circulation* 37: 836-840.
170. Gouw AA, Van der Flier WM, van Straaten EC, Barkhof F, Ferro JM, et al. (2006) Simple versus complex assessment of white matter hyperintensities in relation to physical performance and cognition: the LADIS study. *Journal of neurology* 253: 1189-1196.
171. Garrett D, Cohen R, Paul R, Moser D, Malloy P, et al. (2004) Computer-Mediated Measurement and Subjective Ratings of White Matter Hyperintensities in Vascular Dementia: Relationships to Neuropsychological Performance. *The Clinical Neuropsychologist* 18: 50-62.
172. Gurol ME, Irizarry MC, Smith EE, Raju S, Diaz-Arrastia R, et al. (2006) Plasma beta-amyloid and white matter lesions in AD, MCI, and cerebral amyloid angiopathy. *Neurology* 66: 23-29.
173. Klöppel S, Abdulkadir A, Hadjidemetriou S, Issleib S, Frings L, et al. (2011) A comparison of different automated methods for the detection of white matter lesions in MRI data. *NeuroImage* 57: 416-422.
174. Mortazavi D, Kouzani AZ, Soltanian-Zadeh H (2012) Segmentation of multiple sclerosis lesions in MR images: a review. *Neuroradiology* 54: 299-320.
175. Smart SD, Firbank MJ, O'Brien JT (2011) Validation of automated white matter hyperintensity segmentation. *J Aging Res* 2011: 391783.
176. Smith EE, Salat DH, Jeng J, McCreary CR, Fischl B, et al. (2011) Correlations between MRI white matter lesion location and executive function and episodic memory. *Neurology* 76: 1492-1499.
177. Fennema-Notestine C, Hagler DJ, Jr., McEvoy LK, Fleisher AS, Wu EH, et al. (2009) Structural MRI biomarkers for preclinical and mild Alzheimer's disease. *Human brain mapping* 30: 3238-3253.
178. Löfdahl E, Berg G, Johansson KA, Zachrisson ML, Malmgren H, et al. (2012) Compromised quality of life in adult patients who have received a radiation dose towards the basal part of the brain. A case-control study in long-term survivors from cancer in the head and neck region. *Radiat Oncol* 7: 179.
179. Reisberg B, Ferris SH, de Leon MJ, Crook T (1988) Global Deterioration Scale (GDS). *Psychopharmacol Bull* 24: 661-663.
180. Jack C, Jr, Twomey C, Zinsmeister A, Sharbrough F, Petersen R, et al. (1989) Anterior temporal lobes and hippocampal formations: normative volumetric measurements from MR images in young adults. *Radiology* 172: 549-554.

181. Cook MJ, Fish DR, Shorvon SD, Straughan K, Stevens JM (1992) Hippocampal volumetric and morphometric studies in frontal and temporal lobe epilepsy. *Brain* 115 (Pt 4): 1001-1015.
182. Wieshmann UC, Symms MR, Mottershead JP, MacManus DG, Barker GJ, et al. (1999) Hippocampal layers on high resolution magnetic resonance images: real or imaginary? *J Anat* 195 (Pt 1): 131-135.
183. Pantel J, O'Leary DS, Cretsingher K, Bockholt HJ, Keefe H, et al. (2000) A new method for the in vivo volumetric measurement of the human hippocampus with high neuroanatomical accuracy. *Hippocampus* 10: 752-758.
184. Pantel J, Cretsingher K, Keefe H (1998) Hippocampus Tracing Guidelines. Iowa; University of Iowa Mental Health Clinical Research Center (MHCRC).
185. Convit A, De Leon MJ, Tarshish C, De Santi S, Tsui W, et al. (1997) Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. *Neurobiol Aging* 18: 131-138.
186. Fazekas F, Kleinert R, Offenbacher H, Schmidt R, Kleinert G, et al. (1993) Pathologic correlates of incidental MRI white matter signal hyperintensities. *Neurology* 43: 1683-1689.
187. Malykhin NV, Bouchard TP, Ogilvie CJ, Coupland NJ, Seres P, et al. (2007) Three-dimensional volumetric analysis and reconstruction of amygdala and hippocampal head, body and tail. *Psychiatry research* 155: 155-165.
188. Lenth RV (2001) Some Practical Guidelines for Effective Sample Size Determination. *The American Statistician* 55: 187-193.
189. Soussain C, Ricard D, Fike JR, Mazon JJ, Psimaras D, et al. (2009) CNS complications of radiotherapy and chemotherapy. *Lancet* 374: 1639-1651.
190. 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: 1055-1064.
191. Eckerström C, Andreasson U, Olsson E, Rolstad S, Blennow K, et al. (2010) Combination of Hippocampal Volume and Cerebrospinal Fluid Biomarkers Improves Predictive Value in Mild Cognitive Impairment. *Dement Geriatr Cogn Disord* 29: 294-300.
192. Jauhiainen AM, Pihlajamaki M, Tervo S, Niskanen E, Tanila H, et al. (2009) Discriminating accuracy of medial temporal lobe volumetry and fMRI in mild cognitive impairment. *Hippocampus* 19: 166-175.
193. Pennanen C, Kivipelto M, Tuomainen S, Hartikainen P, Hanninen T, et al. (2004) Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiol Aging* 25: 303-310.
194. Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, et al. (2012) Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and

- neuropsychological test performance. *Neurobiology of aging* 33: 1203-1214.
195. Wardlaw JM, Ferguson KJ, Graham C (2004) White matter hyperintensities and rating scales-observer reliability varies with lesion load. *Journal of neurology* 251: 584-590.
 196. Mantyla R, Erkinjuntti T, Salonen O, Aronen HJ, Peltonen T, et al. (1997) Variable agreement between visual rating scales for white matter hyperintensities on MRI. Comparison of 13 rating scales in a poststroke cohort. *Stroke; a journal of cerebral circulation* 28: 1614-1623.
 197. Zhang Y, Schuff N, Ching C, Tosun D, Zhan W, et al. (2011) Joint assessment of structural, perfusion, and diffusion MRI in Alzheimer's disease and frontotemporal dementia. *International journal of Alzheimer's disease* 2011: 546871.
 198. Taylor KI, Probst A (2008) Anatomic localization of the transentorhinal region of the perirhinal cortex. *Neurobiol Aging* 29: 1591-1596.