



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.



**Multicentre Evaluation of MRI Variability in
the Quantification of Infarct Size in
Experimental Focal Cerebral Ischaemia**

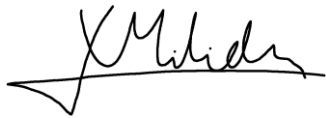
Xenios Milidonis

Doctor of Philosophy
The University of Edinburgh

2016

Declaration

I declare that this thesis has been composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified. Parts of Chapters 3 to 6 have been published or presented as specified in the first page of each chapter.

A handwritten signature in black ink, appearing to read 'X Milidonis', with a long horizontal stroke extending to the left.

Xenios Milidonis

October 2016

Abstract

Ischaemic stroke is a leading cause of death and disability in the developed world. Despite that considerable advances in experimental research enabled understanding of the pathophysiology of the disease and identified hundreds of potential neuroprotective drugs for treatment, no such drug has shown efficacy in humans. The failure in the translation from bench to bedside has been partially attributed to the poor quality and rigour of animal studies. Recently, it has been suggested that multicentre animal studies imitating the design of randomised clinical trials could improve the translation of experimental research. Magnetic resonance imaging (MRI) could be pivotal in such studies due to its non-invasive nature and its high sensitivity to ischaemic lesions, but its accuracy and concordance across centres has not yet been evaluated.

This thesis focussed on the use of MRI for the assessment of late infarct size, the primary outcome used in stroke models. Initially, a systematic review revealed that a plethora of imaging protocols and data analysis methods are used for this purpose. Using meta-analysis techniques, it was determined that T₂-weighted imaging (T2WI) was best correlated with gold standard histology for the measurement of infarct-based treatment effects. Then, geometric accuracy in six different preclinical MRI scanners was assessed using structural phantoms and automated data analysis tools developed in-house. It was found that geometric accuracy varies between scanners, particularly when centre-specific T2WI protocols are used instead of a standardised protocol, though longitudinal stability over six months is high. Finally, a simulation study suggested that the measured geometric errors and the different protocols are sufficient to render infarct volumes and related group comparisons across centres incomparable. The variability increases when both factors are taken into account and when infarct volume is expressed as a relative estimate. Data in this study were analysed using a custom-made semi-automated tool that was faster and more reliable in repeated analyses than manual analysis.

Findings of this thesis support the implementation of standardised methods for the assessment and optimisation of geometric accuracy in MRI scanners, as well as image acquisition and analysis of in vivo data for the measurement of infarct size in multicentre animal studies. Tools and techniques developed as part of the thesis show great promise in the analysis of phantom and in vivo data and could be a step towards this endeavour.

Lay Summary

Stroke is one of the deadliest diseases and very few options are available for treating affected patients. The identification of new treatments begins from animal studies that mimic the pathophysiology of human stroke. However, these studies lack the necessary quality to indicate which treatments are truly effective. In the future, animal studies could potentially involve participation of several research centres, which will bring improvements in their quality and reduce the number of animals used in research. One of the most important markers of the efficacy of treatments is their ability to reduce the size of stroke lesions. This can be measured using magnetic resonance imaging (MRI). This thesis aimed to examine how the size of lesions varies between different MRI scanners and determine which factors contribute mostly to this variation. It was found that different researchers use different MRI methods to evaluate lesions. At the same time, scanners that are routinely used in animal studies have different levels of accuracy, the variability in which is increased by the use of different MRI methods. Consequently, the observed variability in MRI causes a similar heterogeneity in the size of lesions across centres. The thesis proposes ways to improve the accuracy of MRI scanners and the concordance in the size of stroke lesions, which will permit implementation of studies involving several centres. Such studies will help identify new and exciting treatments for the millions of stroke sufferers across the world.

Acknowledgements

First and foremost, I would like to thank my supervisors Prof. Ian Marshall, Prof. Malcolm Macleod and Dr Emily Sena for their time and continuous guidance and support throughout the course of this project. I further thank Ian for providing all the necessary bits and pieces for carrying out the experiments and for helping me identify sources of funding for attending conferences, Malcolm for his many insightful comments and ideas and Emily for sharing her expertise in systematic review and meta-analysis. Their invaluable contribution has helped me enhance and refine my knowledge in basic and translational research.

I would like to thank the staff of Edinburgh Preclinical Imaging (EPI) facility—Dr Maurits Jansen and Ross Lennen—for the countless of hours we spent scanning and especially Ross Lennen for kindly assisting with the validation of the developed tools. I am also indebted to the staff of the rest of the imaging centres who participated in the multicentre study and contributed in the interpretation of findings, and Dr Chris McCabe (University of Glasgow) for providing training in the analysis of stroke data.

Special thanks go to Moira Henderson for showing extraordinary kindness and patience while dealing with all my requests during the last three years, and to all the other people at the Centre for Clinical Brain Sciences (CCBS) for their willingness to help whenever asked to. I also thank my colleagues at the postgraduate office who shared stories and sweet treats and made this journey an experience to remember.

I am deeply grateful to the Multicentre Preclinical Animal Research Team (Multi-PART) and the Tony Watson Bequest (University of Edinburgh) for their financial support, as well as to the European Stroke Research Foundation (ESRF) and the Guarantors of Brain for awarding me travel grants.

Finally, I thank my parents, family and friends for their relentless understanding, encouragement and support throughout this venture, and to whom this thesis is dedicated.

Table of Contents

Declaration.....	iii
Abstract.....	iv
Lay Summary	vi
Acknowledgements.....	vii
List of Figures.....	xiii
List of Tables.....	xvi
List of Abbreviations	xvii
List of Symbols	xix
1 Introduction.....	1
1.1 Overview and Motivation.....	2
1.1.1 Translational failure in stroke research	3
1.1.2 Multicentre animal studies.....	5
1.2 Aims of this Thesis.....	6
1.3 Thesis Outline.....	7
2 Background.....	9
2.1 Stroke.....	10
2.1.1 Definition and classification.....	10
2.1.2 Pathophysiology	11
2.1.3 Rodent models of stroke.....	14
2.2 Magnetic Resonance Imaging in Stroke.....	16
2.2.1 Basic principles of MRI	17
2.2.1.1 Nuclear magnetic resonance.....	17
2.2.1.2 Relaxation.....	19
2.2.1.3 Spatial encoding and image formation	20
2.2.2 T ₁ - and T ₂ - weighted imaging	21
2.2.3 Diffusion- and perfusion-weighted imaging.....	23
2.2.4 Infarct size measurement.....	25

2.2.5	Limitations and sources of heterogeneity	27
2.2.5.1	Anatomy, pathophysiology and stroke models	28
2.2.5.2	MRI equipment and acquisition parameters	29
2.2.5.3	MRI quality assurance.....	30
2.2.5.4	Image analysis	31
3	Systematic Review and Meta-Analysis of MRI in Experimental Focal Cerebral Ischaemia.....	33
3.1	Introduction.....	34
3.2	Methods	36
3.2.1	CAMARADES database	36
3.2.2	Identification of relevant studies.....	36
3.2.3	Data extraction.....	37
3.2.4	Data analysis	39
3.3	Results	40
3.3.1	Study quality.....	41
3.3.2	MRI acquisition.....	42
3.3.3	MRI post-processing	47
3.3.4	Meta-analysis	48
3.4	Discussion.....	52
4	Construction of an MRI Phantom and Development of a Data Analysis Tool.....	57
4.1	Introduction.....	58
4.2	Quality Assurance Protocol	60
4.2.1	Phantom.....	60
4.2.2	Quality assurance criteria.....	62
4.2.3	MRI scanner and equipment	63
4.2.4	T ₂ relaxometry.....	65
4.2.5	Quality assurance scanning protocol	66
4.3	Automated Image Analysis Tool.....	67

4.3.1	Analysis pipeline	68
4.3.2	Segmentation methods	70
4.3.2.1	Simple thresholding.....	70
4.3.2.2	k-means clustering	71
4.3.2.3	Canny edge detection	72
4.3.3	Graphical user interface.....	74
4.3.4	Validation dataset.....	76
4.3.5	Statistical analysis.....	78
4.4	Results.....	80
4.4.1	Phantom's T ₂ relaxation time	80
4.4.2	Intra- and inter-rater agreement.....	81
4.4.3	Performance of automated analysis tool.....	83
4.4.3.1	Selection of Canny edge detection parameters	83
4.4.3.2	Comparison of segmentation methods	88
4.5	Discussion	91
5	Development of a Semi-Automated Infarct Segmentation Tool	95
5.1	Introduction.....	96
5.2	Methods.....	98
5.2.1	Analysis pipeline	98
5.2.2	Graphical user interface.....	102
5.2.3	Validation dataset.....	104
5.2.4	Comparison with stereological point counting.....	104
5.2.5	Measurement of corrected infarct volume	107
5.2.6	Statistical analysis.....	108
5.3	Results.....	110
5.3.1	Selection of parameters for semi-automated segmentation	110
5.3.2	Selection of parameters for stereological point counting.....	114
5.3.3	Comparison of methods	114
5.3.4	Intra- and inter-rater agreement.....	118

5.4	Discussion.....	120
6	Between- and Within-Scanner Variability.....	125
6.1	Introduction.....	126
6.2	Methods	128
6.2.1	Between-scanner variability	128
6.2.2	Within-scanner variability	130
6.2.2.1	Gradient calibration.....	130
6.2.2.2	Imaging.....	131
6.2.3	Post hoc imaging	132
6.2.3.1	Comparison of imaging coils	132
6.2.3.2	Influence of magnetic susceptibility.....	132
6.2.4	Simulations.....	133
6.2.4.1	Reference data	133
6.2.4.2	Scaling using 3D affine geometric transformations	133
6.2.4.3	Estimation of scaling factors.....	136
6.2.5	Image analysis	139
6.2.5.1	Measurement of quality assurance criteria	139
6.2.5.2	Deformation mapping.....	140
6.2.5.3	Measurement of infarct volume in simulated data	142
6.2.6	Statistical analysis.....	143
6.3	Results	145
6.3.1	Between-scanner variability	145
6.3.2	Within-scanner variability	149
6.3.3	Post hoc imaging	152
6.3.3.1	Comparison of imaging coils	152
6.3.3.2	Influence of magnetic susceptibility.....	153
6.3.4	Simulations.....	154
6.3.4.1	Scaling factors.....	154
6.3.4.2	Infarct volume between scanners	156

6.3.4.3	Group comparisons.....	160
6.4	Discussion	166
7	Discussion and Conclusions	175
7.1	Contributions.....	176
7.1.1	Tools	178
7.1.2	Findings.....	179
7.2	Recommendations.....	183
7.3	Limitations and Further Considerations	188
7.3.1	Systematic review and meta-analysis.....	188
7.3.2	Multicentre quality assurance.....	188
7.3.3	Image processing	190
7.3.4	Statistics.....	192
7.4	Future Work.....	193
7.5	Conclusions.....	195
	References	197
A	Meta-Analysis Methodology.....	217
A.1	Estimating Effect Size	218
A.2	Pooling Effect Sizes.....	219
A.3	Meta-Regression.....	223
B	Characteristics of Systematic Review Articles	225
B.1	Extracted Information	226
B.2	Reviewed Articles References	230
C	Statistical Tests.....	237
D	Phantom Scanning Form.....	247

List of Figures

Figure 2.1	Time course of ischaemic cascade.....	12
Figure 2.2	Spin excitation and relaxation.	18
Figure 2.3	Imaging planes and anatomical terms of location in brain MRI.	21
Figure 2.4	Coronal brain MRI images in a rodent model of stroke.	22
Figure 3.1	Flow chart demonstrating the selection process for relevant articles in the systematic review and meta-analysis.....	41
Figure 3.2	Study quality criteria met by studies, sorted by overall prevalence....	42
Figure 3.3	Prevalence of imaging items, sorted by overall prevalence.	43
Figure 3.4	Meta-regression of corresponding histological and MRI effect sizes. .	51
Figure 4.1	A simple structural phantom for assessing geometric accuracy in preclinical MRI scanners.	61
Figure 4.2	The preclinical MRI scanner and equipment used for phantom imaging.	64
Figure 4.3	Automated pipeline for the analysis of phantom data.	69
Figure 4.4	A graphical user interface for the analysis of phantom data.	75
Figure 4.5	The Rician noise in MRI images.	77
Figure 4.6	Measurement of phantom's T ₂ relaxation time.	80
Figure 4.7	Bland-Altman plots assessing intra-rater agreement.	81
Figure 4.8	Bland-Altman plots assessing inter-rater agreement.	82
Figure 4.9	Bland-Altman plots assessing the performance of Canny edge detection for measuring the phantom's dimensions for various thresholds.	84
Figure 4.10	Bland-Altman plots assessing the performance of Canny edge detection for measuring the volume of the phantom's cylindrical compartment for various thresholds.	85
Figure 4.11	Bland-Altman plots assessing the performance of Canny edge detection for measuring the phantom's dimensions for various Gaussian filters.	86

Figure 4.12	Bland-Altman plots assessing the performance of Canny edge detection for measuring the volume of the phantom's cylindrical compartment for various Gaussian filters.....	87
Figure 4.13	Dice coefficient for various Canny thresholds and Gaussian filters. ...	88
Figure 4.14	Bland-Altman plots comparing different methods for measuring the phantom's dimensions.	89
Figure 4.15	Bland-Altman plots comparing different methods for measuring the volume of the phantom's cylindrical compartment.	90
Figure 4.16	Dice coefficient and analysis time for different phantom data analysis methods.....	91
Figure 5.1	Semi-automated pipeline for the measurement of infarct size.	99
Figure 5.2	A graphical user interface for the measurement of infarct size.	103
Figure 5.3	Measurement of infarct volume using stereological point counting..	105
Figure 5.4	Comparison of bias field correction methods.....	110
Figure 5.5	Dice coefficient for various parameters in the semi-automated analysis pipeline.....	112
Figure 5.6	Infarct for various parameters in the semi-automated analysis pipeline.	113
Figure 5.7	Bland-Altman plots comparing methods for measuring infarct volume.	115
Figure 5.8	Dice coefficient, analysis time and infarct volume for all methods....	116
Figure 5.9	Infarct in two rats identified by semi-automated segmentation.....	117
Figure 5.10	Bland-Altman plots assessing intra- and inter-rater agreement.	119
Figure 5.11	Dice coefficient comparing repeated analyses.....	120
Figure 5.12	T ₂ relaxation map in the rat brain following stroke.	123
Figure 6.1	The digital eddy current compensation tool in VnmrJ software.	131
Figure 6.2	Scaling scans across the slice direction.	135
Figure 6.3	Generating 2D geometric deformation maps.	141
Figure 6.4	Between-scanner variability in scaling and volumetric error.	146
Figure 6.5	Influence of SNR and slice thickness in volumetric errors.	147

Figure 6.6	Sample images acquired using six different MRI scanners.....	148
Figure 6.7	Deformation in images acquired with six different MRI scanners.....	149
Figure 6.8	Within-scanner variability in scaling and volumetric error.....	150
Figure 6.9	Deformation in repeated MRI.....	151
Figure 6.10	Effect of imaging coil on geometric distortion.....	152
Figure 6.11	Magnetic susceptibility effects on geometric distortion.....	153
Figure 6.12	Reference and scaled datasets.....	155
Figure 6.13	Infarct volume corrected based on all slices and expressed as an absolute or relative estimate.....	158
Figure 6.14	Infarct volume corrected based on slices including the infarct and expressed as an absolute or relative estimate.....	159
Figure 6.15	Infarct volume in male and female rats corrected based on all slices and expressed as an absolute or relative estimate.....	164
Figure 6.16	Infarct volume in male and female rats corrected based on slices including the infarct and expressed as an absolute or relative estimate.	165

List of Tables

Table 3.1	MRI acquisition parameters for 45 T2WI protocols.	45
Table 3.2	MRI acquisition parameters for 23 DWI protocols.	46
Table 3.3	Characteristics of studies included in meta-analysis.	49
Table 3.4	Combined effect sizes based on MRI and histology infarct size.	50
Table 3.5	Meta-regression analysis of corresponding MRI and histology effect sizes.	51
Table 4.1	Parameters of the phantom’s T ₂ relaxometry pulse sequence.	65
Table 4.2	Parameters of the standardised QA pulse sequence.	67
Table 5.1	Optimising the performance of the infarct segmentation tool.	101
Table 6.1	Details of scanners and imaging coils.	129
Table 6.2	Parameters of standardised and stroke pulse sequences.	130
Table 6.3	Estimating pooled scaling factors for data simulation.	138
Table 6.4	Pooled scaling factors for simulating datasets.	154
Table 6.5	Repeated measures analyses of variance between scanners.	157
Table 6.6	Percentage median differences in infarct volume between male and female rats.	161
Table 6.7	Mann–Whitney <i>U</i> tests comparing infarct volume between male and female rats.	163
Table 7.1	Recommendations regarding the use of MRI for the assessment of infarct size in multicentre studies in rodent models of stroke.	187
Table B.1	Basic characteristics of articles included in the systematic review.	226
Table B.2	Reporting of study quality items.	227
Table B.3	Reporting of basic MRI parameters per type.	229
Table C.1	Shapiro–Wilk test for normality.	238
Table C.2	Mauchly’s sphericity test.	242
Table C.3	<i>p</i> -values for Bonferroni pairwise comparisons.	243
Table C.4	<i>U</i> statistics for Mann–Whitney <i>U</i> tests.	246

List of Abbreviations

2D	2-dimensional
3D	3-dimensional
ABS	Acrylonitrile butadiene styrene
ADC	Apparent diffusion coefficient
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BBB	Blood-brain barrier
BW	Bandwidth
CAMARADES	Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies
CBF	Cerebral blood flow
CE	Coefficient of error
CI	Confidence interval
CSF	Cerebrospinal fluid
CT	Computed tomography
DC	Dice coefficient
DWI	Diffusion-weighted imaging
ETL	Echo train length
ES	Effect size
FLAIR	Fluid-attenuated inversion recovery
FOV	Field of view
FSE	Fast spin echo
FWHM	Full width at half maximum
Gd	Gadolinium
GUI	Graphical user interface
H&E	Haematoxylin and eosin
ID	Inner diameter
IQR	Interquartile range

MCA	Middle cerebral artery
MCAO	Middle cerebral artery occlusion
MRI	Magnetic resonance imaging
MSME	Multi-slice multi-echo
Multi-PART	Multicentre Preclinical Animal Research Team
NMD	Normalised mean difference
NSA	Number of sample averages
PET	Positron emission tomography
PVE	Partial volume effect
PWI	Perfusion-weighted imaging
QA	Quality assurance
RARE	Rapid acquisition with relaxation enhancement
rBW	Receiver bandwidth
RF	Radiofrequency
ROI	Region of interest
SD	Standard deviation
SE	Spin echo
SEM	Standard error of the mean
sMRI	Structural magnetic resonance imaging
SNR	Signal-to-noise ratio
STAIR	Stroke Therapy Academic Industry Roundtable
T1WI	T ₁ -weighted imaging
T2WI	T ₂ -weighted imaging
TE	Echo time
TE _{eff}	Effective echo time
TR	Repetition time
TTC	2,3,5-triphenyl-2H-tetrazolium chloride

List of Symbols

B_0	Main magnetic field (direction or strength)
f_{SD}	Factor adjusting standard deviation of Canny edge detector's Gaussian filter
f_T	Factor adjusting Canny edge detector's thresholds
f_{T_I}	Factor adjusting infarct measurement threshold
G_x	Frequency encoding gradient
G_z	Slice selection gradient
I^2	Between-study heterogeneity in effect size
M_0	Equilibrium magnetisation
M_{xy}	Transverse magnetisation
\bar{R}^2	Adjusted coefficient of determination
s_x	Scaling factor in x direction
s_y	Scaling factor in y direction
s_z	Scaling factor in z direction
V_{CH}	Volume of contralateral brain hemisphere
V_{IH}	Volume of ipsilateral brain hemisphere
V_{Ic}	Volume of infarct corrected for brain swelling
V_{Iu}	Volume of infarct uncorrected for brain swelling
γ	Gyromagnetic ratio
η^2	Partial eta-squared, the effect size for analysis of variance
χ^2	Chi-squared (distribution or test)
ω_0	Larmor frequency

Chapter 1

Introduction

Despite recent advances and discoveries in experimental research, stroke is one of the deadliest diseases in the world and limited options are available for its treatment. Multicentre animal studies utilising the diagnostic and prognostic strength of magnetic resonance imaging (MRI) could be the new paradigm, improving the understanding of stroke and shaping the future of its treatment. In this chapter, the motivation behind this thesis is elaborated and specific aims are listed.

1.1 Overview and Motivation

As a disease affecting a large part of the human population, stroke has been a priority for many scientists in academia and a considerable research effort has enabled understanding of the underlying mechanisms and related pathophysiological changes during the progress of cerebral ischaemia. It was known since the late 1970s that the treatment of acute stroke is based on the rescue of the tissue at risk before the ischaemic cascade reaches its terminal event, the death of neurons and the development of the infarct. This turned the attention of stroke research into the development of drugs that target specific mechanisms of the cascade to protect neurons from dying. However, no neuroprotective drugs have been proven effective in the clinic, despite the fact that over half of 1026 drugs that were tested in animals showed efficacy (O'Collins et al., 2006). Today, only a handful of treatment regimens are of proven benefit in humans (Donnan, Fisher, Macleod, & Davis, 2008), but these are available only for a very small number of patients. It is estimated that the only effective pharmacological treatment, intravenous thrombolysis, reaches at most 7% of patients admitted to the hospital (Barber, Zhang, Demchuk, Hill, & Buchan, 2001; Schwamm et al., 2013), because it can only be administered within the first few hours after stroke onset and is contraindicated in those with a high risk of secondary complications (Davis & Donnan, 2009; Wardlaw, 2010). There is an urgent need for developing new and effective strategies for the treatment of acute stroke.

Despite that the funds spent on stroke research saw a modest increase recently in the developed countries, the proportion of healthcare costs associated with the disease is still higher than the proportion of funds used for stroke research (Donnan et al., 2008; Luengo-Fernandez, Leal, & Gray, 2015; Rothwell, 2001). In 2012 in the UK, for example, the disease accounted for 13% of the combined costs of cancer, coronary heart disease, dementia and stroke management, but only 7% of corresponding funding was spent towards stroke research (Luengo-Fernandez et al., 2015). In contrast, cancer accounted for about 23% of healthcare costs but received almost two thirds of the available research funds.

It is believed that this decline of interest towards stroke resulted from a series of failed efforts to reproduce the positive results of experimental studies in clinical trials, which gradually dissuaded funding agencies and pharmaceutical companies from developing new therapeutic strategies.

1.1.1 Translational failure in stroke research

The bench to bedside failure in stroke has been attributed to a range of translational roadblocks, ranging from the inadequacy of animal models of stroke to mimic the human condition to poor experimental methodological quality and rigour and various forms of study and publication bias that falsely inflated the efficacy of interventions in animals, or understated their efficacy in clinical trials (Dirnagl, 2006; Landis et al., 2012; Sena, van der Worp, Howells, & Macleod, 2007; van der Worp & Macleod, 2011).

Animal models of stroke have intrinsic differences from human stroke, mostly related to the timing of phases of the ischaemic cascade and the duration of pathophysiological changes. While the onset of ischaemia can be precisely defined in animal models, this is not a straightforward task in humans. In addition, the site of occlusion and the extent of collateral supply between animals used in experiments are often homogeneous, but they vary greatly between patients. As of today, there is no animal model available that can mimic the complexity and dynamics of human stroke in an accurate and consistent manner (Canazza, Minati, Boffano, Parati, & Binks, 2014; Howells et al., 2010), therefore imaging and other biomarkers assessed in animals might not reflect similar biomarkers in humans. Despite the countless efforts to define a clinically relevant and predictive biomarker for tissue fate and late outcome, none is widely accepted as a gold standard. What is more, many preclinical trials focus on early evaluation of effects and ignore late outcome, which is more relevant to clinical investigation. On the other hand, some clinical trials have tested drugs under conditions where efficacy was not explicitly proved in animals, such as in comorbid patients or at a substantially longer time after the onset of stroke (Perel

et al., 2007).

Infarct size is the most widely assessed outcome in both clinical and preclinical studies and assertion of efficacy is often based on its reduction in the treated cohort (Macrae, 2011). Preclinical studies tend to measure it at a single time point that could be one day or several weeks following stroke. Earlier time points are preferred as a means to avoid sample size shrinking due to animal death, minimise disease complications and decrease study costs (Krongold et al., 2015). However, infarct progression is a spatiotemporal process and late recanalization could potentially affect the size of the infarct after it was assessed, thereby rendering measured effects invalid (Macrae, 2011; Zanier et al., 2013).

Preclinical studies of stroke have always been carried out locally and in controlled conditions, with the majority of them using young and healthy male rodents. Standardisation of the characteristics of animals used, housing and experimental methods may improve the homogeneity of outcome measures and thus the statistical power of a study, but can affect the reproducibility and the external validity of findings detrimentally (Richter, Garner, Auer, Kunert, & Wurbel, 2010; Richter, Garner, & Wurbel, 2009). In addition, systematic reviews showed that most animal studies use very small sample sizes and do not employ randomisation of treatment allocation and blinding of outcome assessment, techniques considered fundamental to internal validity and standard in clinical trials (Macleod et al., 2008; Perel et al., 2007; Vesterinen et al., 2011). These could explain why published animal studies are more likely to report positive treatment effects, when that effect is not actually present (Dirnagl, 2006; Sena et al., 2007). Additionally, the efficacy of many drugs tested in animals might have been further inflated by failing to publish negative findings (Sena, van der Worp, Bath, Howells, & Macleod, 2010).

1.1.2 Multicentre animal studies

It is evident that current practice in experimental stroke lacks the necessary quality and rigour to move the translation of stroke treatments forward, and amelioration demands radical changes. It has recently been suggested that multicentre preclinical stroke trials with a similar design to randomised clinical trials can lead to improvements in the validity and generalisability of experimental findings (Bath, Macleod, & Green, 2009; Dirnagl & Fisher, 2012; Dirnagl et al., 2013; Howells & Macleod, 2012). A large multicentre study will not replace studies investigating mechanisms of pathophysiology and drug action, but can replace the numerous single-centre studies that measure the efficacy of the exact same treatment in similar conditions (Dirnagl et al., 2013). Participation of several preclinical centres will allow efficient monitoring of animals by concurrent assessment of different cohorts, thus not increasing the study duration significantly. Through the inclusion of a large sample of animals, the statistical power of the multicentre study will be significantly larger and will allow deduction of valid statistical inferences. Decisions regarding the implementation of clinical trials will then be based on better grounds, while potentially reducing the total number of animals killed and saving large amounts of money in the long run (Howells, Sena, O'Collins, & Macleod, 2012; Ioannidis et al., 2014). Furthermore, multicentre animal studies can promote national or international collaboration and sharing of data and expertise between research groups, something considered common in clinical practice today (Dirnagl et al., 2013).

MRI has recently become a favourite modality for the assessment of lesions in experimental stroke, as it allows *in vivo* brain imaging with excellent soft tissue differentiation. However, animal MRI scanners are far less common than clinical scanners and collaborations between institutions are still rather scarce. As such, there is a wide diversity of hardware, software, scanning protocols and image processing techniques that could possibly affect the equivalence of outcomes between centres. As an outcome assessment method, MRI must be standardised within and between centres to improve the accuracy and comparability of measured effects; this is

customary for large clinical imaging collaborations (Jack et al., 2008; Wintermark et al., 2008), but preclinical research has yet to reach this milestone.

Identification of the best strategies for all aspects of a multicentre stroke study and development of relevant methods is expectedly a very challenging and arduous task. Recently, a consortium of neuroscientists and stroke experimentalists from around the world was set up with the aim of developing a framework for the design and implementation of multicentre stroke trials. This consortium, the Multicentre Preclinical Animal Research Team (Multi-PART; www.multi-part.org), took important steps towards sample selection, modelling of the disease, identification of relevant endpoints and determination of standardised procedures for behavioural, histological and statistical analyses. However, still no work has been done with the use of MRI for the assessment of infarct size.

1.2 Aims of this Thesis

In collaboration with Multi-PART, the main aim of this thesis is to examine the use of MRI for the measurement of infarct size in focal cerebral ischaemia and propose methods for improving its accuracy and reproducibility to facilitate multicentre animal stroke studies. Specific goals are to:

- Systematically review the literature to determine the variability in methods used for the acquisition and analysis of MRI data for the measurement of lesion size.
- Identify the MRI acquisition method providing the best estimate of infarct-based treatment effect by comparison against gold standard post-mortem histological assessment using meta-analysis techniques.
- Develop an MRI phantom and associated image analysis tools for the assessment of geometric accuracy in preclinical scanners.

- Develop a semi-automated tool for the assisted segmentation of the infarct in rodent MRI data and compare its performance against manual analysis and other common volumetric techniques.
- Design and execute a multicentre study to compare geometric accuracy between different preclinical MRI scanners using a standardised and centre-specific protocols, and record drifts in the performance over time.
- Simulate in vivo data modelling the geometric errors in participating scanners and the differences in stroke protocols to examine their impact on infarct size across centres and related group comparisons.

1.3 Thesis Outline

The content of each of the remaining chapters is summarised below:

Chapter 2 sets the scene by summarising relevant background information about focal cerebral ischaemia and the basic animal models used. The following section is devoted to the use of MRI in clinical and experimental stroke, with emphasis on methods used for the assessment of lesion size and particularly infarct size. The major challenges of infarct size quantification are finally described.

Chapter 3 presents a systematic review and meta-analysis of the use of MRI in experimental stroke. MRI acquisition and post-processing methods for the measurement of ischaemic lesions are recorded and their heterogeneity is evaluated. The reporting of common study quality criteria and imaging information is also assessed. Finally, the MRI contrast technique that is best associated with gold standard histology for the measurement of infarct-based treatment effects is identified using meta-regression.

Chapter 4 describes the procedure for the construction of a simple structural phantom and the development of accompanying phantom data analysis tools for evaluating

geometric accuracy in small animal MRI scanners. The performance of the tool is validated by analysis of scans simulating different linear scaling errors and levels of image quality.

Chapter 5 describes the development of a simple threshold-based semi-automated tool for the analysis of rodent MRI data for the measurement of infarct size. The performance of the tool is compared against gold standard manual tracing of the boundary of the lesion and stereological point counting.

Chapter 6 presents a multicentre study assessing geometric accuracy in six different scanners across Europe using the developed phantom and tools; this study was designed based on findings of the previous chapters. The short-term and long-term longitudinal variability in geometric accuracy over a period of six months in the worst-performing system following calibration is also evaluated. Finally, the impact of between-scanner variability on the quantification of infarct volume and infarct-based statistical group comparisons is examined. Images corresponding to each system were generated by transforming a reference dataset of MRI scans in a rat model of focal cerebral ischaemia and were analysed using the tool described in the previous chapter.

Chapter 7 summarises the main findings of this thesis and discusses their importance for animal stroke research. In addition, guidance to multicentre stroke studies regarding the use of MRI for the measurement of infarct volume is given and, finally, future directions are discussed.

Chapter 2

Background

This chapter presents the main types of stroke, the pathophysiological processes involved and the major animal models employed in the laboratory for its assessment. A brief introduction to the fundamentals of signal detection and image formation in MRI is then presented, along with a description of common imaging methods used for the assessment of lesion in stroke. Particular emphasis is given to the measurement of infarct size, the most commonly assessed outcome in preclinical stroke trials. The chapter closes with a description of the main challenges and sources of error behind such measurements, some of which are addressed in the studies described later in this thesis.

2.1 Stroke

Almost 17 million people suffer a stroke each year in the developed world and more than a third of them die, whilst many others are left with serious and long-term disability (Mozaffarian et al., 2015). This makes stroke the second most common cause of death after coronary artery disease, accounting for 11.8% of total deaths worldwide (Mozaffarian et al., 2015). Although stroke mortality varies between geographic regions and race or ethnicity, and a steady reduction has been observed in the western countries the last few decades, the global burden of the disease is rising detrimentally (Mathers & Loncar, 2006). This is due to the fact that widespread pathological conditions, such as hypertension, diabetes, obesity and smoking, are major risk factors for stroke and are also on the rise (Duong, 2012b; Mozaffarian et al., 2015). At the same time, the disease accounts for more than 4% of healthcare costs in these countries, with costs expected to increase multifold over the next few decades (Donnan et al., 2008; Heidenreich et al., 2011; Mozaffarian et al., 2015).

2.1.1 Definition and classification

Stroke is defined as the loss of cerebral function, focally or globally, due to a sudden reduction of cerebral blood flow (CBF) arising from vascular disturbances (Sacco et al., 2013). Based on the cause of CBF reduction, stroke is classified into two major types, haemorrhagic and ischaemic. Haemorrhagic stroke is characterised by the spontaneous rupture of a small vessel due to the presence of an arterial aneurysm or a vascular malformation and occasionally due to persistent hypertension, leading to cerebral bleeding (Warlow, Sudlow, Dennis, Wardlaw, & Sandercock, 2003). In most cases, haemorrhagic stroke results in intracerebral haemorrhage and in a quarter of them subarachnoid haemorrhage; both account for about 13% of all human strokes (Mozaffarian et al., 2015). The most prevalent type of stroke is ischaemic and accounts for 87% of all cases. It results when a brain vessel is occluded either by an atherosclerotic plaque (thrombus) originating from the carotid artery or by an embolic clot carried from the heart into the brain via the bloodstream, restricting the

blood supply to the underlying tissue. While the majority of ischaemic strokes affect one of the large cerebral arteries, a quarter of them are lacunar and are characterised by occlusion of one of the penetrating arteries supplying the brain's deep structures, often due to chronic hypertension (Warlow et al., 2003). Cerebral ischaemia is further classified by clinicians based on the duration of neurological deficits following occlusion. A transient ischaemic attack is followed by restoration of normal function before the first 24 hours post-stroke, whilst a permanent ischaemic stroke is characterised by symptoms persisting for longer than 24 hours (Donnan et al., 2008). The lack of accurate and reliable evidence renders this distinction rather arbitrary, but helps guide patient management in day-to-day clinical practice and categorise patients in epidemiological studies (Donnan et al., 2008; Warlow et al., 2003).

For the clarity of this thesis, the term 'stroke' will refer to focal cerebral ischaemia hereafter and, therefore, any information presented may not be relevant or applicable to other forms of stroke.

2.1.2 Pathophysiology

The onset of stroke marks the beginning of a long sequence of molecular and cellular events taking place within the brain territory whose blood supply has been compromised, called the ischaemic cascade (Figure 2.1).

As soon as CBF reduces critically, oxygen and glucose levels at the affected area fall and energy generation via production of adenosine triphosphate (ATP) becomes insufficient. At this acute stage, the depletion of energy results in dysfunction of ATP-reliant ion pumps across the membranes of neurons and glia. Consequently, membrane potentials are lost and ion homeostasis is disrupted, causing an elevation of sodium, calcium and chloride ions in the intracellular space. Damage to the cells is further catalysed by the elevated calcium in the cytosol, by stimulating degradation of cytoskeletal proteins and release of enzymes, such as phospholipases which decompose membrane phospholipids.

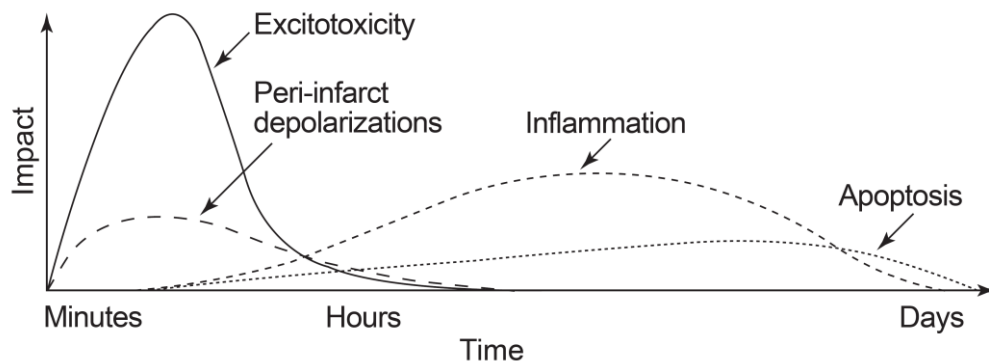


Figure 2.1 Time course of ischaemic cascade.

Various pathophysiological mechanisms follow cerebral ischaemia, initiating with excitotoxicity that damages neurons and glia, and triggers peri-infarct depolarisations, inflammation and apoptosis that lead to tissue death. Reprinted from Dirnagl, Iadecola, and Moskowitz (1999), with permission from Elsevier.

At the same time, the excitatory neurotransmitter glutamate is excessively released by calcium-mediated exocytosis from presynaptic cells into the synaptic clefts that accelerates the movement of calcium into the cells (Dirnagl et al., 1999). The pathological process impairing neuron potentials by excessive stimulation of glutamate and enzymes triggered by energy depletion is called excitotoxicity. Accumulation of ions in the cells results in their swelling and the formation of cytotoxic oedema, an almost immediate aftermath of cerebral ischaemia (Durukan & Tatlisumak, 2007). Excess calcium also causes the generation of free radicals via phospholipases' activity, which initiate lipid peroxidation in mitochondria and may cause deleterious alterations in deoxyribonucleic and ribonucleic acids (Brouns & De Deyn, 2009). These changes, along with cytoskeletal breakdown, make the cell membranes more permeable to the harmful molecules and gradually lead to cell necrosis and the formation of infarct (necrotic tissue). Cell death in less hypoperfused neighbouring tissue is mediated primarily via apoptotic pathways acting in a chain reaction. As experiments in rats revealed, apoptosis is also propagated in surrounding tissue by spreading peri-infarct depolarisations that occur spontaneously over the course of ischaemia (Dijkhuizen et al., 1999; Dirnagl et al., 1999).

Within a few hours following the development of the infarct, the endothelial junctions

that make up the brain's blood-brain barrier (BBB) are disrupted and plasma proteins leak out of brain capillaries. This fluid penetrates into the parenchymal extracellular space leading to accumulation of vasogenic oedema and swelling of the brain, which marks the beginning of the subacute stage of stroke. While cytotoxic oedema is dominant the first few hours following stroke, it soon gives way to vasogenic oedema that can take up to 30% of the actual infarct size and can persist for several days (Gerriets et al., 2004; Swanson & Sharp, 1994).

In the clinic, by the time a stroke patient is admitted for treatment, cell death may have already taken place at the end artery territory. The focus of clinicians is then turned onto salvaging the surrounding tissue that has been compromised, which will expectedly transit into a non-salvageable infarct within hours following stroke. Towards the differentiation of these two areas, Astrup and his colleagues were the first to introduce the concepts of ischaemic core and penumbra (Astrup, Siesjo, & Symon, 1981; Astrup, Symon, Branston, & Lassen, 1977). Using electrophysiological techniques, they examined somatosensory evoked responses to determine thresholds in CBF at which electrical activity was abolished in the baboon cortex. Their experiments suggested the existence of two thresholds; one for morphological damage and a less severe threshold for functional impairment. A CBF less than about 15 ml/100g/min renders neurons non-functional and permanent damage occurs within minutes after stroke; this was referred to as the ischaemic core. At 20 ml/100g/min, neuronal functionality is maintained, but evoked responses decline gradually when flow is reduced below this value and therefore marks the threshold for functional disruption. This region surrounds the ischaemic core and was defined as the ischaemic penumbra; while neurons in this region are functionally impaired, they are structurally intact and an increase in CBF can recover their signalling and normalise extracellular ionic activity (Astrup et al., 1981; Astrup et al., 1977). The limited perfusion of this region is maintained by small peripheral vessels and is enough to keep the cell membrane ion channels working to preserve the homeostasis needed for survival for several hours following stroke. Neuronal death in penumbral

tissue initiates if the blood flow decreases critically or if it is kept at insufficient levels for a prolonged period of time. As the penumbra might retrieve its normal functioning if CBF is restored, it is sometimes referred to as the 'tissue at-risk' and is today a well-established concept in stroke often used by researchers as a target for therapeutic interventions. Its accurate measurement is important in identifying patients who may benefit from thrombolytic treatment and in designing clinical trials (Campbell & Macrae, 2015; Duong, 2013).

2.1.3 Rodent models of stroke

The translation of stroke treatments to clinical trials begins from experimental studies of cultured cells and organotypic brain slices *in vitro* or animal models *in vivo* (Macrae, 2011). The study of stroke using animal models constitutes the gold standard for the understanding of the pathophysiology of the disease and the development of novel therapies and biomarkers (Sicard & Fisher, 2009). Highly controlled animal experiments ensure consistency in the site of arterial occlusion, the severity and duration of ischaemia, and pre- and post-ischaemia monitoring of outcomes using a plethora of assessment methods, either *in vivo* or post-mortem histological techniques. This comes in contrast with human stroke, where the delayed introduction of the patient into the clinic allows insufficient investigation of the disease during the acute phase, with assessment restricted to the use of *in vivo* methods.

In animal models, stroke is artificially induced by occluding a cerebral artery and causing ischaemia in the underlying brain territory. Several species are used in preclinical stroke studies, including rodents, cats, rabbits and non-human primates. The choice is primarily based on the correlation between animal and human pathophysiology, but it is always driven by the availability of the animals, the ease of handling and by ethical and financial considerations. Consequently, while non-human primates have gyrencephalic brains and great behavioural similarity to humans, they are very rarely used (Carmichael, 2005). The vast majority of

experimentalists perform their studies on mice and rats, which have low maintenance costs, husbandry and handling, can be bred in dedicated facilities for ensuring genetic consistency and develop equivalent patterns of infarct and functional deficits after stroke (Durukan & Tatlisumak, 2007; Howells et al., 2010). Mice offer superior genetic manipulation which is useful for studying the molecular pathophysiology. Rats, on the other hand, are better suited for modelling focal cerebral ischaemia because their size enables practical in vivo or ex vivo assessment, they possess a circle of Willis similar to that of humans and most rat models of stroke have small intra-strain variability allowing reproducible experiments (Canazza et al., 2014; Carmichael, 2005; Durukan & Tatlisumak, 2007; Sicard & Fisher, 2009).

The cerebral artery most prone to occlusion in humans is the middle cerebral artery (MCA), accounting for 80% of cases in Caucasians (Warlow et al., 2003). For this reason, most rodent stroke models involve permanent or transient occlusion of this artery. The most commonly used rat model is the intraluminal filament method (Howells et al., 2010; Sicard & Fisher, 2009), firstly described by Koizumi, Yoshida, Nakazawa, and Ooneda (1986) and later modified by Longa, Weinstein, Carlson, and Cummins (1989). The procedure involves the introduction of a silicon rubber-coated monofilament into the external carotid artery, which is then inserted into the internal carotid artery and advanced to occlude the origin of the MCA. The filament can be kept in place to permanently occlude the artery, or carefully retracted to allow reperfusion of the ischaemic area, which is useful for modelling different durations of ischaemia and studying the effects of early or late reperfusion (Howells et al., 2010). This model does not require a craniotomy, but the filament can cause endothelial ruptures that lead to small haemorrhages or microbleeds, produced infarcts can have a highly variable size and associated animal mortality is often higher than other MCA occlusion (MCAO) models (Canazza et al., 2014).

A range of other MCAO models are also available, offering different options regarding the duration of occlusion, the location and reproducibility of ischaemic injury; these include electrocoagulation of the proximal end of the artery (Robinson,

Shoemaker, Schlumpf, Valk, & Bloom, 1975; Tamura, Graham, McCulloch, & Teasdale, 1981), injection of clot-like particles to mimic embolic stroke (Kudo, Aoyama, Ichimori, & Fukunaga, 1982), and topical application of vasoconstrictor endothelin-1 to induce dose-dependent local ischaemia (Macrae, Robinson, Graham, Reid, & McCulloch, 1993; Sharkey, Ritchie, & Kelly, 1993).

2.2 Magnetic Resonance Imaging in Stroke

The findings of the early animal studies by Astrup and colleagues (Astrup et al., 1981; Astrup et al., 1977) gave birth to the idea that human stroke is characterised by cellular alterations and corresponding blood flow thresholds. This idea was confirmed when positron emission tomography (PET) was later used to measure blood perfusion and oxygen levels following cerebral ischaemia in humans (Baron et al., 1981; Marchal et al., 1993). Similar successes in mammalian stroke studies followed, which served as the catalyst that helped establish medical imaging into a necessity in preclinical and clinical studies to quantify the extent of infarcted and salvageable tissue and other physiological parameters, which are important for the testing of new interventions or guiding treatment procedures (Wardlaw, 2010).

Until the 1990s only PET and computed tomography (CT) were utilised to visualise stroke lesion. PET scanners, however, are not readily available and related image processing is time-consuming, whilst the more widespread CT is not sensitive to the early ischaemic cascade and delivers a considerable radiation dose to patients (Ebinger et al., 2009). While MRI was relatively common by then, it took a few years for its potential in stroke to become apparent. The breakthrough came with the invention of more advanced techniques with clinical utility, namely diffusion-weighted (DWI) and perfusion-weighted imaging (PWI).

Today, MRI is a widely used diagnostic modality as it is non-invasive, radiation-free and offers excellent soft tissue differentiation. Brain imaging is one of its most popular

applications, where MRI protocols allow measurement of structural, metabolic, haemodynamic and functional characteristics of the brain during the progression of the disease. MRI findings are used to correlate structural alterations in the brain with late functional impairment due to stroke to aid the diagnosis of complications and the prognosis of tissue fate and stroke recurrence (Duong, 2013; Redgrave, Coutts, Schulz, Briley, & Rothwell, 2007; Wintermark et al., 2008).

2.2.1 Basic principles of MRI

2.2.1.1 Nuclear magnetic resonance

MRI relies on the electromagnetic properties of atomic nuclei and the phenomenon of nuclear magnetic resonance (McRobbie, Moore, Graves, & Prince, 2006). Most MRI applications exploit the presence of hydrogen nuclei, the most abundant in the body. Each hydrogen nucleus consists of a single positively charged proton, which rotates around its own axis due to its spin, an intrinsic property of all subatomic particles. Therefore, the proton generates its own magnetic field called magnetic moment, forcing it to act as a tiny magnet when placed in an external magnetic field.

An MRI scanner uses a large superconducting electromagnet which produces a very strong homogeneous magnetic field inside its bore. When a subject is placed in the scanner, its protons are aligned or anti-aligned with the field's magnetic lines (Figure 2.2). The direction of alignment with the magnetic field (B_0) is energetically beneficial and therefore a slightly larger number of protons are aligned than anti-aligned, which is enough to produce a net longitudinal magnetization pointing in the direction of B_0 (equilibrium magnetisation, M_0). The magnetic field also exerts torque on the protons, causing them to precess around the field's axis with a frequency called the Larmor frequency, ω_0 . This is determined by the gyromagnetic ratio, γ , which is unique for

each nucleus, and the strength of the field, B_0^1 , using the Larmor equation:

$$\omega_0 = \gamma B_0 \quad (2.1)$$

Most modern human scanners generate magnetic fields with a strength of 1.5 or 3 T and animal scanners 7 T or higher. For the hydrogen nucleus, $\gamma = 42.6 \text{ MHz/T}$, thus the Larmor frequency falls within the radiofrequency (RF) range.

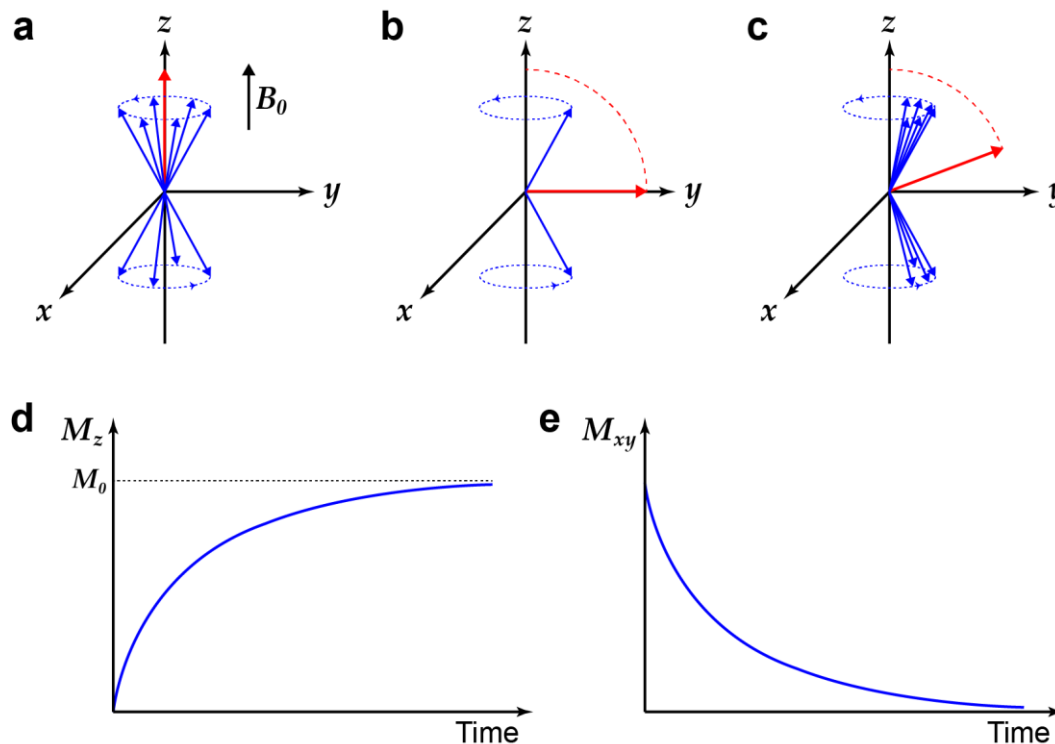


Figure 2.2 Spin excitation and relaxation.

MRI uses a Cartesian coordinate system of reference where the z axis is the longitudinal direction along the scanner's bore and is perpendicular to the x - y plane (transverse plane). (a) At equilibrium the spins (blue arrows) precess randomly around z axis in one of two possible directions, but there is a net longitudinal magnetisation M_0 (red arrow) towards the direction of the scanner's magnetic field B_0 . (b) When excited with a radiofrequency pulse, the spins momentarily gain phase coherence and transverse magnetisation M_{xy} . (c) Following excitation, longitudinal magnetisation M_z is slowly restored via T_1 relaxation, while M_{xy} is reduced via T_2 relaxation (spin dephasing). Both of these mechanisms of relaxation follow exponential curves, as shown in (d) and (e) respectively.

¹ B_0 typically refers to both the magnetic field direction and strength (in units of magnetic flux density: Tesla, T).

A magnetic resonance signal is then produced by exciting the protons using an RF coil emitting pulses with the Larmor frequency to achieve resonance condition. This flips the net magnetization vector of protons by usually 90° to gain transverse magnetisation. Excitation also induces phase coherence to protons, meaning that their magnetic moments are pointing in the same direction and they precess in synchrony. As soon as the excitation pulse is over, the protons slowly realign with B_0 while losing phase coherence. During this process protons emit a weak radio signal that can be amplified and measured using an RF receive coil at various time points before they reach magnetic equilibrium.

2.2.1.2 Relaxation

The time protons need to reach equilibrium is known as relaxation time and is dependent on two major processes: (1) the spin-lattice interaction by which protons restore longitudinal magnetisation by dispersing energy into the surrounding atoms or molecules (termed T_1 or longitudinal relaxation), and (2) the spin-spin interaction which dephases the spins as energy is dissipated by interactions between neighbouring protons (termed T_2 or transverse relaxation) (Figure 2.2). T_2 relaxation occurs faster than T_1 relaxation and is enhanced by additional but less severe dephasing caused by field inhomogeneities, known as T_2^* relaxation.

Hydrogen nuclei in the human body are found mostly in water and, therefore, the water content of each tissue type affects the rate of these relaxation mechanisms; this rate is determined by measurable relaxation times T_1 and T_2 that are specific to tissues and depend on B_0 . The difference in relaxation times between tissues allows the creation of images with a different intensity for each type. MRI acquisition can be weighted so that one of the two main relaxation processes determines the contrast between tissues; if T_1 relaxation prevails, T_1 -weighted images are acquired, otherwise T_2 -weighted images are produced.

2.2.1.3 Spatial encoding and image formation

The signal recorded by the RF receive coil comes from the entire sample. To localise the excited protons and produce an image, the signal is spatially encoded in each orthogonal direction using three additional coils surrounding the main magnet in the MRI machine, called gradient coils. For encoding in the slice direction, one of these coils applies a linearly varying magnetic field that when added to the main field B_0 produces a distribution of frequencies along this direction centred around ω_0 . Since the sample has now a linearly varying resonance frequency, an RF pulse can be used to excite each slice of the sample at a time. The slice thickness is increased by increasing the bandwidth (BW) of the transmitted RF pulse or decreasing the strength of the slice selection gradient (typically G_z). After a slice has been selected, the other two gradient coils are used to encode in-plane information. The frequency encoding gradient (typically G_x) coil applies a linear magnetic field gradient to produce a similar variation in Larmor frequency as G_z , whereas the phase encoding gradient (typically G_y) coil applies a brief pulse in the perpendicular direction to induce a linear variation in the phase of protons. G_x is applied during reception of the MRI signal, hence it is also called the readout gradient. Scanning can be performed in any orientation within the scanner and each of the three gradient coils is used to perform one of these functions accordingly. The application of the three gradients assigns unique information to protons in each voxel (a contraction of 'volume' and 'element'), which is spatially decoded to produce the image; for this, signals are digitised and recorded in the spatial frequency domain in a complex mathematical matrix called k -space, which is converted into an intensity image using the Fourier transform. Basic imaging planes and anatomical terms of location in brain MRI, as referred to in this thesis, are shown in Figure 2.3.

A series of RF pulses and gradients is applied to a sample to generate an image; this is known as a pulse sequence. The duration and weighting of a pulse sequence is determined primarily by the time between the excitation and sampling (echo time, TE) and the time between successive excitations (repetition time, TR). Dozens of such

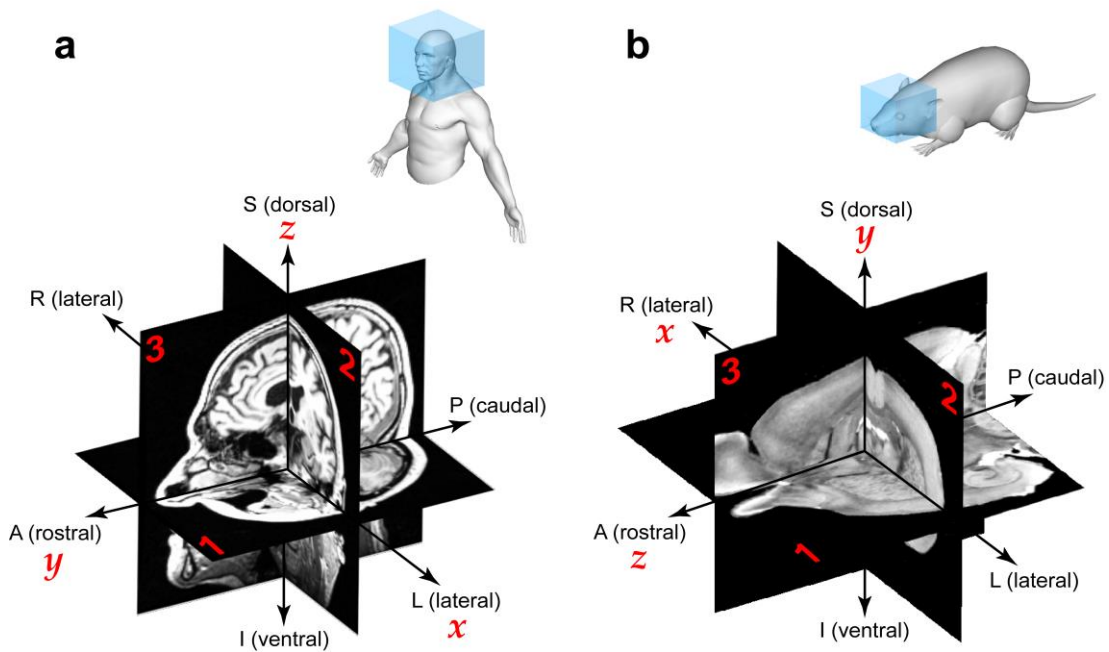


Figure 2.3 Imaging planes and anatomical terms of location in brain MRI. These are cross sections through 3-dimensional structural MRI images of (a) a human and (b) a rat, showing the (1) axial, (2) coronal and (3) sagittal imaging planes and anatomical terms of location. Due to differences in anatomy and positioning in the scanner (humans in the supine position and rodents in the prone position), the scanner's orthogonal axes correspond to different anatomical directions. The figure shows the convention used in this thesis, where x , y and z indicate the positive direction. A indicates anterior; P, posterior; S, superior; I, inferior; R, right; L, left.

sequences have been developed over the years; a routine MRI brain examination for stroke commonly consists of a battery of sequences for assessing various aspects of the disease. Common MRI techniques used for the assessment of lesion size are described in the following sections.

2.2.2 T₁- and T₂- weighted imaging

Structural MRI (sMRI) is by far the most widely used MRI technique, producing high resolution images in a short period of time for examining the anatomy and pathophysiology of the brain. T₁-weighted (T1WI) or T₂-weighted imaging (T2WI) are basic structural methods that can be used to differentiate between haemorrhagic and ischaemic strokes and facilitate treatment decisions in the clinic (Dijkhuizen & Nicolay, 2003; Duong, 2013). T1WI is often used along with the injection of

gadolinium (Gd) chelate-based paramagnetic contrast agents to assess BBB permeability. These agents accumulate in areas where BBB integrity has been compromised and enhance the T₁-weighted signal by shortening the T₁ relaxation time of these areas; this effect correlates with subsequent haemorrhagic transformation (Knight et al., 1998). In contrast, T₂WI is sensitive to the presence of vasogenic oedema at 12 hours or later after occlusion of the artery (Figure 2.4). The oedema has a higher water content compared to surrounding normal tissue and its T₂ relaxation time is longer; as such, it looks more hyperintense (brighter) in the T₂-weighted image. Therefore, the technique is often employed for the visualisation of the oedema, which is a surrogate for necrotic tissue and can be used to indirectly measure infarct size in both humans and animals (Dijkhuizen & Nicolay, 2003; Farr & Wegener, 2010). Nevertheless, both T₁WI and T₂WI are avoided at the acute phase of stroke due to their poor sensitivity to early ischaemic injury (Wardlaw, 2010).

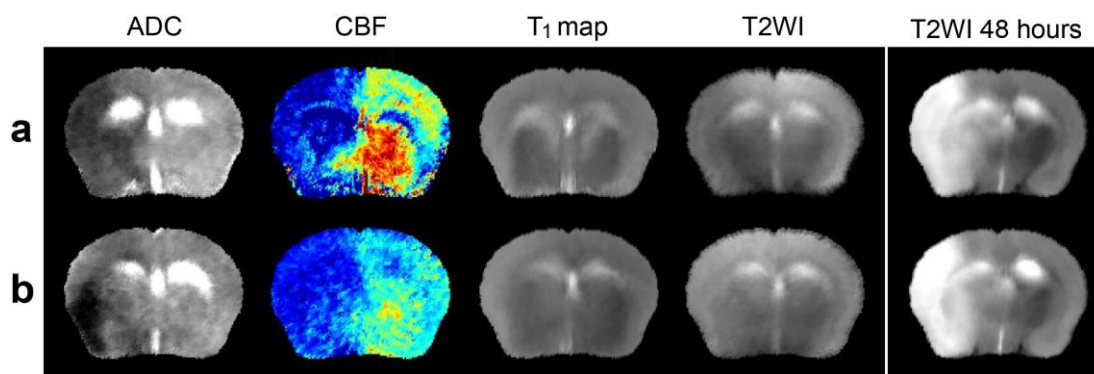


Figure 2.4 Coronal brain MRI images in a rodent model of stroke.

These are average MRI images for two different groups of mice that were subjected to 90 minutes of middle cerebral artery occlusion and scanned (a) during artery occlusion and (b) 3 hours later for apparent diffusion coefficient (ADC) quantification using diffusion imaging, cerebral blood flow (CBF) quantification using perfusion imaging, quantitative T₁-weighted imaging (T₁WI) and qualitative T₂-weighted imaging (T₂WI); mice were scanned again at 48 hours after stroke for T₂WI alone (column on the right). ADC and CBF maps can identify reduced diffusion of water and perfusion of blood within minutes after stroke, whereas T₁WI and T₂WI are insensitive to early ischaemic changes. Late T₂WI demarcates the infarct (bright), which is often larger than the ischaemic core on early ADC maps (dark). The shift of the midline of the brain towards the normal contralateral hemisphere in T₂WI at 48 hours is caused by swelling of the ipsilateral hemisphere. Adapted from Leithner et al. (2015), with permission from Wolters Kluwer.

Specific structural sequences produce quantitative MRI data that can be converted to maps quantifying T1 or T2 relaxation time in a voxel-wise manner, instead of showing intensity in arbitrary units. As these relaxation times are properties inherent to tissue, quantitative MRI maps are useful for identifying biomarkers that are comparable across centres. Alteration of T2 relaxation time throughout the course of ischaemia was shown to have predictive utility and can inform on changes in tissue status (Wegener et al., 2006). Quantitative maps are also used to derive thresholds of abnormality (Pham et al., 2010).

2.2.3 Diffusion- and perfusion-weighted imaging

In normal tissue, water molecules can diffuse relatively freely. In contrast, in the ischaemic region during the acute phase of stroke the diffusion of water is restricted as it shifts from the extracellular space into the cells due to disruption of ion homeostasis. DWI can detect areas where the movement of water is reduced or blocked within minutes and up to 6 hours following stroke, effectively providing a means for confirmation of occlusion (Dijkhuizen & Nicolay, 2003; Farr & Wegener, 2010). Tissue with reduced diffusion at the acute stage of stroke indicates the presence of cytotoxic oedema and appears hyperintense in the diffusion-weighted image. Studies suggest that this lesion represents the non-salvageable ischaemic core (Albers & Fisher, 2011; Duong, 2013; Moseley et al., 1990). In the subacute and chronic phases of stroke the cytotoxic oedema subsides due to cell lysis and the DWI lesion becomes less hyperintense. Therefore, while the accuracy of the technique for depicting the extent of the infarct reduces with time, it can effectively be used to distinguish the acute from the later stages of stroke.

From the raw DWI data, the apparent diffusion coefficient (ADC) map can be produced where the intensity of each voxel is relative to the vector sum of the phases of all diffusing molecules in the voxel, which represents the rate of diffusion. Since in the ischaemic area the diffusion of water in the extracellular space is restricted, it appears hypointense (dark) in the ADC map (Figure 2.4). The ADC map is generally

preferred to DWI for assessing early lesions as it is independent of the T₂ shine-through effect, a phenomenon that causes areas with long T₂ values to appear bright in DWI images, even if they do not restrict diffusion.

Several studies have established ADC thresholds for defining the tissue at risk and final infarct size in rodents (Meng, Fisher, Shen, Sotak, & Duong, 2004; Reid et al., 2012; Shen, Meng, Fisher, Sotak, & Duong, 2003), but there is no consensus on the optimal values among the research community; imprecise thresholds may result in either inclusion of undamaged tissue or exclusion of infarcted areas. In addition, the hyperintense acute DWI lesion is more likely to represent glial rather than neuronal damage, since glial cells have a better chance of recovery and reopening of the occluded artery partially salvages the DWI lesion (Rivers & Wardlaw, 2005). This view is supported by clinical DWI studies, indicating that the technique provides inconsistent identification of infarct core (Kranz & Eastwood, 2009; Rivers et al., 2007).

PWI is a different technique that exploits the magnetic properties of water in blood to detect its passage through the tissue. A common PWI method is arterial spin labelling, in which the blood in a specific brain area is magnetically 'labelled' and the adjacent slice encompassing the tissue of interest is imaged soon afterwards to detect the flow of the labelled spins. This technique is often used in rodent models of stroke (McCabe et al., 2009; Meng et al., 2004), as it is non-invasive and provides a direct assessment of CBF. Contrast bolus tracking is another PWI method used mostly in humans, where an exogenous Gd-based contrast agent is given to the subject intravenously to enhance the visualisation of the flowing blood in cerebral arteries (Wardlaw, 2010). A signal-time curve is obtained for each voxel while the contrast agent first passes through the brain; this can be processed to generate maps quantifying different parameters including CBF (Figure 2.4) and cerebral blood volume, among others (Kane et al., 2007).

In spite of the versatility of perfusion imaging, Dani and colleagues state that the technique is highly variable between human subjects and would probably remain as

such even if it was standardised (Dani et al., 2011; Dani et al., 2012). In addition, it is not uncommon that the perfusion lesion includes areas of benign oligoemia that usually present no risk of infarction (Butcher et al., 2008). Identification of the most clinically useful PWI parameter and appropriate threshold for its definition is a subject of long and ongoing research (Albers & Fisher, 2011; Bandera et al., 2006; Bardutzky et al., 2005; Dani et al., 2011; Kane et al., 2007).

The perfusion lesion is generally larger than the acute diffusion lesion. Areas of the perfusion lesion that look normal in the diffusion image form what is called the perfusion-diffusion (PWI/DWI) mismatch. This region is hypoperfused due to cerebral vascular occlusion (typically visualised by PWI-derived CBF maps) but excludes irreversibly damaged cells where neuronal signalling is disrupted and cytotoxic oedema has already been developed (typically indicated by DWI-derived ADC maps) (Campbell & Macrae, 2015; Duong, 2013). Such a tissue state coincides with the definition of the ischaemic penumbra by Astrup and colleagues (1981), thus the PWI/DWI mismatch is considered to be a promising non-invasive indicator of this region. Its feasibility as a possible target area for testing the efficacy of interventions has been demonstrated in various animal species (McCabe et al., 2009; Meng et al., 2004). As such, the PWI/DWI mismatch early after stroke was successfully used in major clinical trials for the selection of subjects that were most likely to benefit from treatment (Albers et al., 2006; Davis et al., 2008), though its benefit was not shown in another trial (Hacke et al., 2009). In general, the lack of accurate definitions for both underlying MRI methods poses serious doubts about the use of the mismatch to guide treatment in routine clinical practice (Bandera et al., 2006; Butcher et al., 2008).

2.2.4 Infarct size measurement

Despite many attempts to develop advanced prognostic MRI biomarkers utilising multi-modal information for assessing the efficacy of potential therapeutics for stroke (Duong, 2012a), infarct size remains the primary outcome used (Farr & Wegener, 2010; Macrae, 2011). It is known to be an important biomarker of the progression of

ischaemia and its measurement in the clinic is encouraged by current imaging recommendations (Wintermark et al., 2013). In both clinical and preclinical research, infarct size is often measured in parallel with functional and behavioural scores to assess early or late outcome (Macrae, 2011; Saunders, Clifton, & Brown, 1995; Schiemanck, Kwakkel, Post, & Prevo, 2006).

Traditionally, histological and immunohistochemical methods have been employed to quantify the extent of the infarct in animals, by staining brain slices *ex vivo* using dyes, such as 2,3,5-triphenyl-2H-tetrazolium chloride (TTC), haematoxylin and eosin (H&E) or cresyl violet (Bederson et al., 1986; Liu, Zhen, Meloni, Campbell, & Winn, 2009; Sommer, 2010). These approaches offer unparalleled structural detail and specificity, but involve time-consuming procedures and—as post-mortem analyses—do not allow longitudinal assessment, which is important in stroke as ischaemia and infarction evolve substantially over time. While histology is still regarded as the gold standard for measuring infarct size (Csonka et al., 2010; Liu et al., 2009), the advent of high resolution animal MRI scanners for non-invasive *in vivo* assessment increased the popularity of MRI across preclinical centres. Studies reported excellent association between MRI and histology, with the former being able to predict histologically determined chronic infarct size when used in the first few days after stroke (Barone, Clark, Feuerstein, Lenkinski, & Sarkar, 1991; Peeling et al., 2001).

As described earlier, the sensitivity of DWI to early cytotoxic effects allows its use in the acute period for imaging the ischaemic core, but this lesion size could be later reduced due to early reperfusion or increased due to prolonged ischaemia. Thus, the method of choice for the visualisation of the late infarct is T2WI. However, infarct size on T₂-weighted data is confounded by the presence of vasogenic oedema, which accounts for about 30% of the total apparent hyperintensity and can lead to false-positive results if not taken into account (Swanson et al., 1990; Swanson & Sharp, 1994). Oedema causes swelling of the ipsilateral brain hemisphere; this peaks during the first 48 hours and can persist for up to seven days in rat MCAO models (Macrae, 2011). A number of formulae have been devised and used over the years to correct

for this effect. The simplest of these is to normalise the infarct size with the ratio of the area of the contralateral hemisphere to that of the ipsilateral hemisphere (Leach, Swan, Eisenthal, Dopson, & Nobbs, 1993). Another method was devised by Swanson et al. (1990) and was later adapted by Lin, He, Wu, Khan, and Hsu (1993); it suggests that the size of the infarct can be approximated by subtraction of the healthy area in the ipsilateral hemisphere from the area of the whole contralateral hemisphere. Both of these methods assume that the oedema is present only within the infarcted tissue and were proposed for use on TTC-stained rodent brains, but since then they have been regularly applied on T2WI data as well (Barone et al., 1991). Gerriets et al. (2004) later developed a formula for correcting infarct volume extracted from MRI data specifically by making a number of assumptions based on in vivo observations. Accurate quantification of infarct size requires acquisition of a sufficient number of slices throughout the MCA territory, typically in the coronal plane, and the visible lesion area is traced in each slice. Formulae for correcting for brain swelling require measurement of the volume of both brain hemispheres as well.

Infarct size is reported in published articles either as an absolute estimate, such as in units of volume (mm^3 or μL), or as a relative estimate, for example the percentage of brain or contralateral hemisphere. Absolute estimates are useful for determining the precise infarct size variation between treated cohorts, but relative values permit straightforward comparison between studies using animals with different characteristics (Liu et al., 2009; Macrae, 2011).

2.2.5 Limitations and sources of heterogeneity

Heterogeneity in the anatomy and pathophysiology between animals, as well as the experimental procedures for modelling stroke and assessing outcome, are thought to be the primary contributors to lesion size variability between animals and centres (Liu et al., 2009). While scientists tend to use standardised protocols for these criteria, it has been argued that heterogeneity in some aspects of animal experimentation might help in mimicking the complex human condition as much as possible and improve

the generalisability of results (Richter et al., 2010; Richter et al., 2009). On the other hand, outcome assessment methods including MRI, can introduce a moderate but systematic bias in measurements if not properly calibrated or carried out. In a multicentre study where absolute measurements are primarily combined and the number of animals per group and per centre can be highly variable, such systematic differences between centres could render combined effects and, subsequently, conclusions inaccurate.

Thus far, the use of MRI in experimental stroke has been highly heterogeneous between centres. Standardised guidelines to inform the conduct of imaging in human stroke clinical trials are available (Wintermark et al., 2008; Wintermark et al., 2013), but no such guidelines exist for experimental stroke. Below, important factors that could influence the measurement of infarct size using MRI are described.

2.2.5.1 Anatomy, pathophysiology and stroke models

For answering a research question, experimenters can choose the animal species and their strain, sex, age, size and comorbidity from a wide range of options; these can greatly influence the results of their study. When combined with inconsistencies in the conduct of experimental procedures, including surgery and occlusion of the artery, this heterogeneity can produce conspicuous variability in lesion sizes between preclinical studies and research centres (Liu et al., 2009; Walberer et al., 2006).

In MRI, blood flow fluctuations and animal motion are known to introduce physiological noise that degrades the quality of data. During scanning, motion is adequately compensated by anaesthetising the animal and often by using dedicated animal cradles equipped with nose cones or stereotaxic ear bars to hold the animal's head in place. However, it is known that most anaesthetics used during scanning have intrinsic neuroprotective effects themselves and could further affect lesion evolution and measured outcomes (Howells et al., 2010; Kawaguchi, Furuya, & Patel, 2005; Macleod et al., 2009).

The sinuses, nasal cavity and ear canals of the head may introduce significant artefacts in MRI data. These air-filled cavities located rostral and ventral to the rodent brain give rise to local magnetic susceptibility effects that induce magnetic field inhomogeneities around the borders with tissue (air-tissue interface) and cause spatial distortions and 'blooming' of surrounding tissue, the extent of which depends on the type of pulse sequence.

Quantification with basic T₂-weighted sequences, including spin echo (SE) and fast spin echo (FSE), is also affected by the cerebrospinal fluid (CSF). Due to its high water content, CSF appears hyperintense similarly to the oedema and may confound the estimation of infarct size when the ventricles are adjacent to the lesion (see T2WI at 48 hours following stroke in Figure 2.4). Fluid-attenuated inversion recovery (FLAIR) is a pulse sequence that suppresses the signal from CSF, but is generally a slower technique and its T₂ signal change is poorly understood (Leigh & Krakauer, 2014), thus is not often used in experimental stroke.

2.2.5.2 MRI equipment and acquisition parameters

Preclinical MRI scanners make use of large magnets producing higher field strengths than clinical scanners. A higher field strength provides a better signal-to-noise ratio (SNR) in acquired data, which is important when scanning small animals as it permits the use of smaller voxels for high resolution imaging, but magnifies magnetic susceptibility artefacts.

Gradient coil nonlinearities and RF coil loading effects can induce inhomogeneities in the magnetic field, which can produce spatial distortions and intensity variations in the images. Various types of coils exist for signal detection, which influence image quality in additional ways. Volume coils offer homogeneous intensity across the scanned region and are often used for human brain MRI, but are rarely used for imaging small animals because are impractical and introduce large amounts of noise. On the other hand, surface coils that sit on the top of the animal's head are easier to use and have higher sensitivity to tissue that lies close to the coil, improving the SNR

within the brain. This sensitivity, however, compromises global intensity homogeneity by introducing a steep radial intensity gradient across the image.

The heterogeneity in imaging protocols for the assessment of infarct size in experimental stroke has not been evaluated thoroughly, but some studies suggest that it could be very high. A systematic review of 141 diffusion MRI studies in stroke models found that only a quarter of them provided adequate information regarding imaging, the conduct and findings of which differed significantly (Rivers & Wardlaw, 2005). Similar variability was observed in later systematic reviews of clinical multimodal MRI (Schiemanck et al., 2006), diffusion (Kranz & Eastwood, 2009) and perfusion imaging (Dani et al., 2011; Dani et al., 2012).

It is reasonable to assume that this heterogeneity holds true for other MRI methods as well. Variation in the imaging plane and pulse sequence parameters affect image resolution, SNR and the contrast between normal and damaged tissue. Another important adverse result would be the inconsistency in partial volume effect (PVE). This artefact arises from the use of large voxel sizes (low resolution of acquired images); the larger the voxel the more tissue types it may represent. In the final image, the signal coming from contributing tissues is averaged, resulting in loss of detail and blurring which, consequently, could cause ambiguity in measurements (Ren, Shen, Bardutzky, Fisher, & Duong, 2004). High-resolution MRI can minimise tissue differentiation errors, but at the expense of SNR and speed of scanning.

2.2.5.3 MRI quality assurance

In order for MRI systems to perform optimally and acquired images to be of any clinical or preclinical use, scanners and coils must be subject to frequent and rigorous quality assurance (QA). This aims to identify sources of errors and measure their effects by scanning temporally stable objects called phantoms; measured effects are then compensated at regular intervals to avoid any drifts in the system's performance (Firbank, Harrison, Williams, & Coulthard, 2000; Price et al., 1990). QA with daily and weekly or monthly tests has been the standard in clinical practice for years, with

various protocols being used to assess a number of important performance characteristics of the scanners by comparison with well-established criteria (American College of Radiology, 2015). However, the conduct of QA in preclinical centres is still very scarce and dedicated protocols are absent, with the majority of centres relying on annual maintenance of the systems by external service engineers. As a consequence, imaging performance between scanners may vary significantly and any changes in MRI endpoints assessed in longitudinal studies spanning over several months may be attributed, at least partially, to drifts in scanner performance, unbeknown to the experimenter.

2.2.5.4 Image analysis

MRI intensity variations described earlier are collectively known as bias field; this often rules out the use of automated tools for image analysis that rely on assumptions of global intensity homogeneity. For this reason, researchers prefer to measure lesion size by manually tracing the lesion in the images. Manual analysis itself is a laborious method with proven poor reliability and reproducibility in many applications, with variability between or within raters being common and significant (Ay et al., 2008; Galinovic et al., 2014). Amongst other factors, the experience and the subjectivity of judgement of the rater contribute mostly to infarct size variability; this effect is called observer bias. Simple methods to reduce it exist but were found to be rarely performed for the measurement of various outcomes (van der Worp et al., 2010; Vesterinen et al., 2011); these include blinding of analysis to group assignment and repetition by the same and at least another independent rater to confirm that measurements are repeatable and reproducible, respectively.

Another potential limitation of current practice is inherent to the mathematical procedures used to estimate infarct volume, which lack specific guidelines. While the standard method for correcting for brain swelling or calculating relative estimates of infarct volume is by comparing it against the volume of the whole brain or a hemisphere (Lin et al., 1993), some choose to use the brain or hemispheric volume

measured only in slices where the infarct is present (Lestro Henriques et al., 2015). A possible incentive for this is to avoid the negating effect of formulae used to correct for brain swelling, when this is significant (McBride, Klebe, Tang, & Zhang, 2015). Inconsistencies in calculations could contribute to variability in reported infarct volume estimates and associated group comparisons, even when all other study characteristics are identical.

Identification and complete elimination of the effect of all aforementioned sources of error is difficult—if not impossible—due to the fact that most of them appear concurrently and produce complex and unpredictable variation. Rather, it could be useful to alleviate those effects that arise from known sources and can be measured using simple and systematic approaches. The following chapters of this thesis will measure the variability in experimental stroke MRI and address some of the persisting contributing factors, to improve the reliability in the measurement of infarct size across centres.

Chapter 3

Systematic Review and Meta-Analysis of MRI in Experimental Focal Cerebral Ischaemia

This chapter aims to assess previously published studies in animal models of stroke to record methods for MRI data acquisition and analysis for the quantification of ischaemic lesions. The heterogeneity and adequacy of reporting basic imaging parameters is evaluated, as well as the reporting of basic study quality criteria. Finally, via means of meta-analysis the MRI contrast technique that best correlates with gold standard histology for estimating infarct-based treatment effects is identified; this technique will form the basis of the studies described in later chapters.

Parts of this chapter have been published or presented as:

Milidonis, X., Marshall, I., Macleod, M. R., & Sena, E. S. (2015). Magnetic resonance imaging in experimental stroke and comparison with histology: systematic review and meta-analysis. *Stroke*, 46(3), 843-851. doi:10.1161/STROKEAHA.114.007560

Milidonis, X., Marshall, I., Macleod, M. R., & Sena, E. S. (2015). Magnetic resonance imaging for the assessment of lesion size in experimental stroke: systematic review and meta-analysis. *Presented at the 7th Annual Scientific Meeting of the Scottish Imaging Network: A Platform for Scientific Excellence (SINAPSE)*. Aberdeen, UK.

3.1 Introduction

MRI is an invaluable and versatile imaging tool used in both clinical and preclinical research. While its variability in clinical research was extensively documented and largely addressed, the heterogeneity of MRI in preclinical studies is not fully explored. As described in the previous chapter, many MRI techniques are sensitive to the consequences of early or late ischaemia and can provide information helpful for confirmation of occlusion, localization and tracking of the evolving ischaemia, or prognosis of late outcome (Dijkhuizen & Nicolay, 2003; Farr & Wegener, 2010; Wardlaw, 2010). Nevertheless, the traditional method for measuring the primary outcome in stroke, infarct volume, is post-mortem histological analysis of brain slices (Liu et al., 2009; Macrae, 2011; Sommer, 2010). This approach requires that a subgroup of animals be killed at an arbitrary time point after the induction of ischaemia; the choice of this time point is critical, as brain injury post-stroke is a spatiotemporally evolving process (Wegener et al., 2006). MRI, in contrast, is an alternative method allowing longitudinal *in vivo* assessment of the developing infarct with often faster preparation procedures and less-demanding measurement techniques. These factors could make MRI the preferred method in a multicentre setting where hundreds of animals could be used, but participating scanners and the techniques employed at each centre should be accurate enough to allow efficient comparison and pooling of measurements.

As an initial step towards this endeavour, this chapter aims to determine the extent of the variability in MRI data acquisition and post-processing methods for the estimation of lesion size through a systematic review of published literature. Systematic review is a method used for the identification of all relevant studies in order to answer a particular research question (Sena, Currie, McCann, Macleod, & Howells, 2014). Critical appraisal of these studies is performed in a systematic and unbiased manner, in order to produce more meaningful summaries of current evidence than simple narrative reviews. As such, systematic review findings could generate hypotheses for, or drive the conduct of similar studies in the future. In

stroke, systematic reviews of data arising from preclinical studies may provide insight into the failure of translation by identifying potential gaps in knowledge and sources of error in methodology and reporting of findings.

The second objective of this chapter is to examine the performance of recorded MRI methods for the measurement of infarct volume via comparison with gold standard histology. While several studies showed equivalence between MRI and histological infarct size measurements in rodents post-stroke (Barone et al., 1991; Gerriets et al., 2004; Peeling et al., 2001), an alternative assessment by combining lesion-based treatment effects from multiple studies will provide more robust and accurate estimates of correlation. For this purpose, meta-analysis techniques are employed in this chapter. Meta-analysis uses data recorded in a systematic review to calculate weighted average estimates of effects and their confidence intervals (CI). Its potential in stroke is obvious; it may be used to give a clear view of the true efficacy of a treatment intervention and allow informed decisions to be made regarding whether this intervention should be taken forward into clinical trials (Sena et al., 2014). When used on data grouped according to various aspects of study design, such as the imaging method or the time of outcome assessment following stroke, meta-analysis can indicate how each of these factors influences average estimates. Finally, associations between two or more variables recorded in the systematic review can be examined using meta-regression (Vesterinen et al., 2014); this technique is used here for the comparison of MRI and histology.

3.2 Methods

3.2.1 CAMARADES database

The Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES; www.camarades.info) group is an international consortium of experts in the field of experimental and translational medicine aiming to provide a framework for systematic review and meta-analysis of treatment interventions in animal models, following a similar standard to the one enforced on clinical reviews by the Cochrane Collaboration. Since its establishment in 2004, CAMARADES has been conducting systematic reviews of interventions tested in animal models of stroke, and was more recently expanded into other neurological diseases. Records of papers and extracted information are saved and analysed in a Microsoft® Access® data manager (2003, Microsoft Corp., Redmond, Washington, USA), which is accessed by all CAMARADES collaborators. At the commencement of this review (April 2014), this database held details of 1824 publications describing 5329 experiments testing the efficacy of candidate drugs in focal cerebral ischaemia. The study described in this chapter was performed in collaboration with this group, based at the University of Edinburgh.

3.2.2 Identification of relevant studies

For systematic review, publications held in the CAMARADES database were searched to identify those that fulfilled pre-defined inclusion criteria: controlled studies of interventions in animal models of stroke where MRI was the outcome measurement method and lesion size the measured outcome. The full text of identified papers was then screened to select only the studies which determined lesion size in vivo and reported measurements in text or graphically. Articles published as abstracts or in languages other than English were also considered; the latter were translated.

For meta-analysis, the papers included in the systematic review were further screened to keep only those which reported use of post-mortem histological assessment of infarct size in addition to MRI. Studies were retained if histology was used immediately after MRI in the same cohort of animals, and measurements for treatment and control groups were provided for both methods.

3.2.3 Data extraction

To examine the heterogeneity in the conduct of MRI in the systematic review, a range of relevant parameters were extracted from all included articles, partially based on generic contemporary recommendations by Stout et al. (2013): the method of positioning the animal in the scanner, the anaesthetic used for scanning, the physiological parameters monitored during scanning, the field strength and manufacturer of the scanner, the type of RF coil, the type of pulse sequence and relevant parameters (TR, TE, field of view (FOV), matrix size, slice thickness, number of slices, number of signal averages (NSA)), the scan duration, the software used for image analysis, the method used for delineating the boundary of the lesion, the formula used for correcting for the presence of oedema (applicable to T2WI) and how the infarct size is reported (for example mm³ or percentage of brain volume). Other pulse sequence parameters were considered where appropriate, such as echo train length (ETL) for FSE sequence, or details of diffusion-sensitising gradients in DWI. The quality of image analysis was assessed by examining whether raters were blinded to group allocation and if analysis was repeated by multiple independent raters. Other meta-data were extracted, including the first author's name, the year of publication, the journal and details regarding the animals and stroke model used: the animal species, the type of ischaemia, the intervention tested, the ischaemic side of brain and the cerebral artery that was occluded for ischaemia.

As suggested by systematic review and meta-analysis guidelines (Sena et al., 2014), the quality of studies was also assessed; here, a published study quality checklist was used (Macleod, O'Collins, Howells, & Donnan, 2004), comprising the following 10

items: (1) publication in a peer-reviewed journal, (2) control of temperature during surgery and induction of ischaemia, (3) random allocation to treatment or control, (4) blinded induction of ischaemia, (5) blinded assessment of outcome, (6) use of anaesthetic without significant intrinsic neuroprotective activity, (7) use of comorbid animals (for example aged, diabetic or hypertensive), (8) sample size calculation, (9) compliance with animal welfare regulations and (10) statement of potential conflict of interests. Publications were awarded a single point for each of the items that they met, for a maximum score of 10.

For meta-analysis, infarct size measurements for MRI and histology group comparisons were extracted from relevant papers. In this case, a 'comparison' is defined as the assessment of outcome in a treatment group as compared to a control group, which are given a dose of a drug or vehicle, respectively, at a given time before or after the induction of ischemia (Sena, 2010). For performing a comparison, the mean infarct volume, the standard deviation (SD) or standard error of the mean (SEM), and the number of animals per group were recorded. Where data were not given in the text, values were extracted from graphs using the graphical tool Paint (6.3, Microsoft Corp., Redmond, Washington, USA), based on provided scales or the values of the infarct size axis. Finally, the time point of assessing infarct size after occlusion of the artery was also noted.

MRI scanner and sequence parameters described above were extracted once by myself and were randomly double-checked for confirmation. All other parameters were extracted previously by at least one CAMARADES reviewer, and were independently extracted again by myself for confirmation.

3.2.4 Data analysis

For systematic review, the percentage of studies reporting each study design and quality parameter was calculated. Several articles described the use of more than one MRI method (for example, T2WI, DWI or PWI); where relevant, the percentage of reported imaging parameters was calculated against the total number of MRI methods and not the total number of included articles. Since the choice of many imaging parameters is often influenced by the animal species used, the range and the median value of continuous parameters (scan duration and pulse sequence parameters, where applicable) were calculated separately for each species. Where values for the same parameter were reported in different units, they were converted to the same scale before pooling.

In meta-analysis, if studies used a single control group and multiple treatment groups, the number of animals in the control group was divided by the number of treatment groups it served for performing appropriate comparisons (Vesterinen et al., 2014). Then, the treatment effects were measured for both MRI and histology infarct size estimates as a normalised mean difference (NMD) (Vesterinen et al., 2014), representing the percentage improvement in the treated cohort (known as effect size (ES)). The DerSimonian and Laird random effects meta-analysis (DerSimonian & Laird, 1986) was used to pool treatment effects because substantial heterogeneity between studies was anticipated. The significance of heterogeneity between k groups/comparisons was assessed using the chi-squared (χ^2) distribution with $k - 1$ degrees of freedom (df) (Vesterinen et al., 2014). The CAMARADES database and Microsoft Excel® (2013, Microsoft Corp., Redmond, Washington, USA) were used for calculations. Meta-regression was performed in Stata® (13, StataCorp, College Station, Texas, USA) using function `metareg` to assess the relationship between ESs determined histologically and by each MRI method (Harbord & Higgins, 2008). The proportion of the variance in the dependent variable (histological outcome) that is explained by variation in the independent variable (MRI) is indicated by an adjusted coefficient of determination, \bar{R}^2 . The time of outcome assessment after stroke was

then included as a regression covariate to examine whether it accounts for some between-study heterogeneity and improves the model's predictability. Statistical significance was set at $p < 0.05$. For a detailed description of the meta-analysis techniques and related formulae used in this study, see Appendix A.

3.3 Results

Figure 3.1 demonstrates the selection process for both systematic review and meta-analysis. Fifty-seven studies using MRI in animal models of stroke were initially identified, two of which did not report measurement of lesion size and one performed *ex vivo* MRI; data were extracted from the remaining 54 studies, published between 1994 and 2010 (references for all included papers are listed in Appendix B). One study was published in Chinese and was translated into English (2%). Two studies were published in abstract only (4%); to avoid data duplication, all articles were cross-checked to verify that these two studies were not published in full as well. Most studies used rats (42, 78%), 11 used mice (20%) and one used baboons (2%), while all studies used MCAO models of focal cerebral ischaemia. Forty-five studies used T2WI (83%), 23 used DWI (43%), four used T1WI (7%), three used PWI (6%) and two used 'plasma volume imaging' (4%), a type of perfusion imaging involving injection of a superparamagnetic contrast agent; one study did not state the method employed (2%). Overall, there were 78 different scanning protocols, with 19 articles (35%) reporting more than one method for assessing lesion size. Basic information regarding each study is given in Table B.1 in Appendix B.

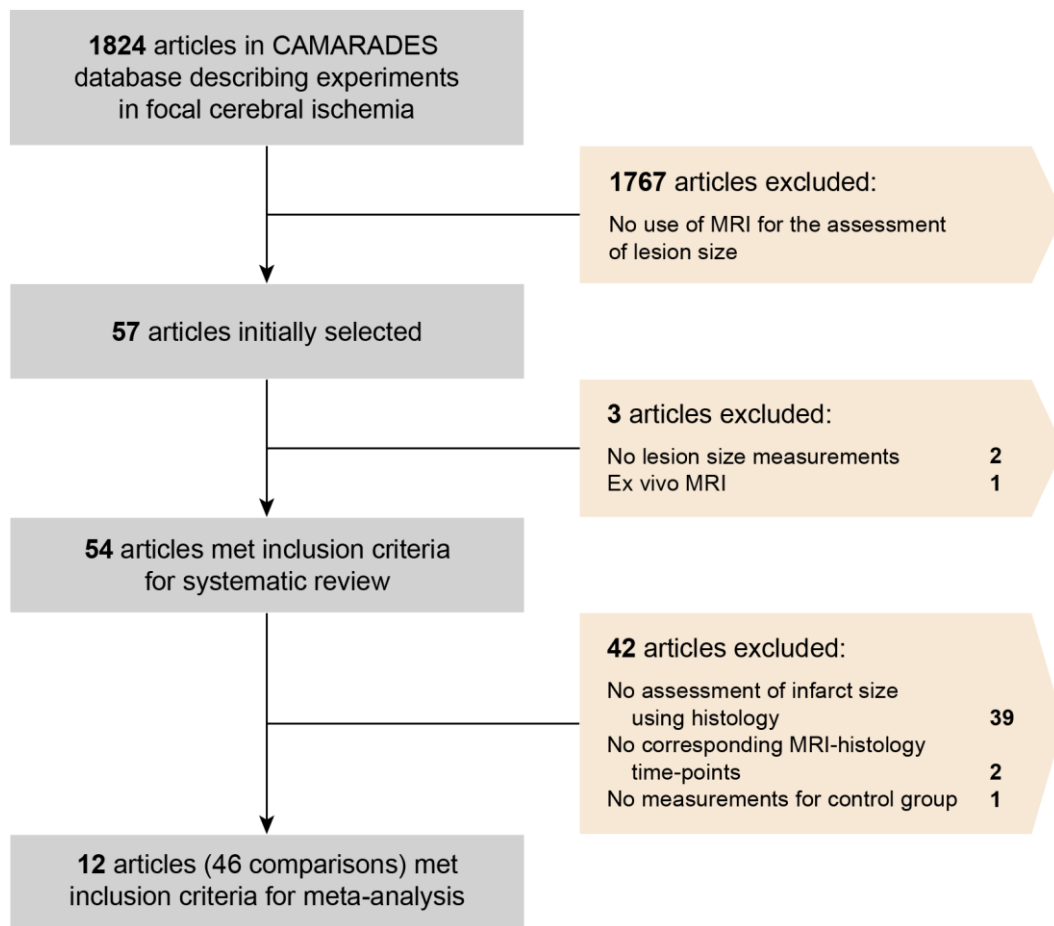


Figure 3.1 Flow chart demonstrating the selection process for relevant articles in the systematic review and meta-analysis.

3.3.1 Study quality

The median number of study quality items scored was 4 and the range was 0–6. Three studies gave a statement of potential conflicts of interest (6%), two reported blinding of the induction of ischaemia (4%), two the use of animals with comorbidities (4%) and no study reported a sample size calculation (Figure 3.2). Random allocation to group was reported in 18 studies (33%) and blinded assessment of lesion size in just 9 studies (17%), despite the fact that some studies reported blinding of the assessment of outcomes other than MRI lesions; overall, 17 studies mentioned blinding of the assessment of at least one outcome (31%). Control of animals' temperature during surgery, compliance with animal welfare regulations and use of an anaesthetic without known neuroprotective activity were reported in the majority of reviewed

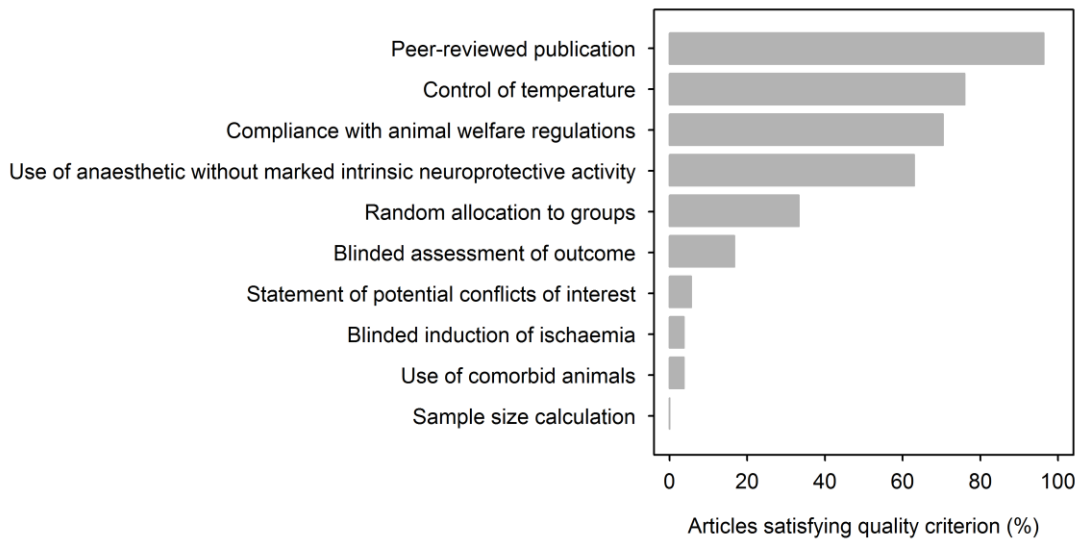


Figure 3.2 Study quality criteria met by studies, sorted by overall prevalence.

articles (41 articles, 76%; 38 articles, 70%; and 34 articles, 63%, respectively). A full report on study quality per article is given in Table B.2 in Appendix B.

3.3.2 MRI acquisition

Figure 3.3 shows the prevalence of various items related to MRI methodology in reviewed articles. Parameters were assessed for reporting either against the total number of articles (54) or against the total number of scanning protocols (78), where relevant. For this reason, it was assumed that if a study reported a scanner and specific equipment for imaging, a software for image analysis and information regarding animal positioning, anaesthetisation and monitoring of physiological parameters, these were applicable to all scanning protocols used in that study.

Articles tended to report pulse sequence parameters more often than information regarding imaging equipment and procedures (Figure 3.3). However, the use of MRI between studies was highly heterogeneous, with a wide range of anaesthetics, scanner equipment manufacturers, scanner field strengths and types of RF coils reported. Surprisingly, 3 articles did not mention the scanner's field strength (6%); the studies reporting the field strength used 1.5 T to 9.4 T magnets, with 4.7 T and 7

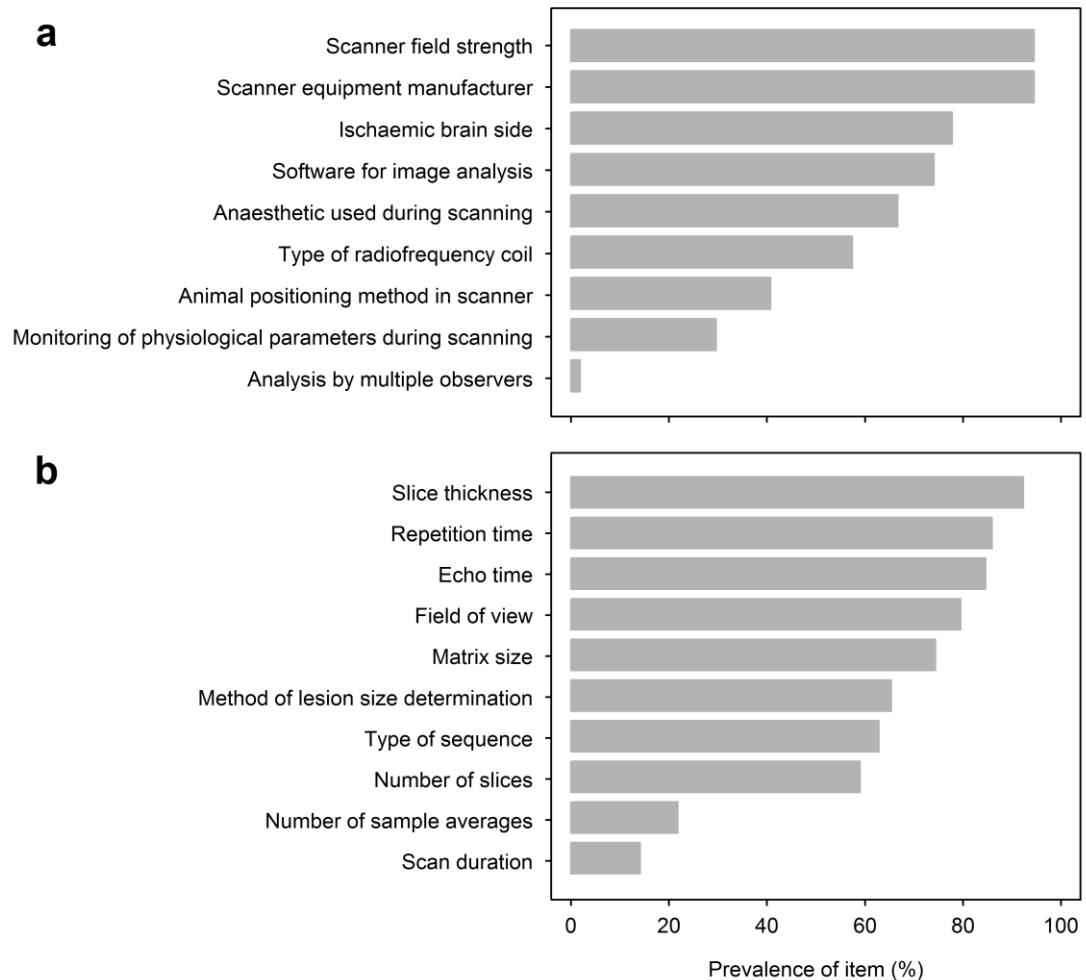


Figure 3.3 Prevalence of imaging items, sorted by overall prevalence. Items were assessed for reporting either (a) against the total number of articles or (b) against the total number of scanning protocols, where relevant.

T being the most prevalent (15 of 51 studies, 29% each). Thirty-one articles reported the type of RF coil used (57%); from these, volume coils were the preferred type for both signal transmission and reception (17 studies, 55%), but surface coils or a combination of both types for signal transmission and reception were also commonly reported (7 studies, 23% each). Sixteen studies alone reported monitoring of physiological parameters during scanning (30%), with temperature being the most prevalent parameter (13 studies, 81%). Fifty-one articles reported the imaging plane or showed exemplar brain images (94%); this was the coronal plane for all (as shown in Figure 2.3). Interestingly, the side of the brain where ischaemia was induced was not mentioned in 12 studies (22%), while in some cases exemplar images shown in

the article conflicted with what it was reported in text. The prevalence for each different type of these MRI items is given in Table B.3 in Appendix B.

T2WI and DWI were the two most commonly used MRI methods and the heterogeneity in their scanning parameters was large, as Tables 3.1 and 3.2 suggest. Considering rat T2WI alone, the TE ranged from 30 ms to 120 ms, the number of slices from six to 35, the slice thickness from 0.35 mm to 3 mm, the FOV from 22×22 mm² to 120×120 mm² and the matrix size from 64×64 to 512×512. The range and median values for many corresponding parameters in mouse T2WI was slightly reduced, particularly TR, slice thickness, FOV and number of slices, as expected. Variation in DWI was similar to T2WI, and details of the diffusion-sensitising gradients were reported infrequently except for the *b*-factor; this ranged from 20 s/mm² to 1898 s/mm² for rat imaging and 70 s/mm² to 1224 s/mm² for mouse imaging. Probably the most important parameter regarding the scanning protocol, the type of pulse sequence, was not reported for many protocols; articles failed to include it for 12 T2WI protocols (27%) and 11 DWI protocols (48%). Finally, the duration of the scanning protocol per animal was infrequently reported (7 T2WI protocols, 16%; 2 DWI protocols, 9%); this varied significantly for rat imaging (T2WI lasted from 1 to 9 minutes; DWI often took longer, with a range of 7 to 15 minutes).

SE and FSE were the most commonly used types of pulse sequences for T2WI (14 of 45 protocols each, 31%), four studies used a multi-slice multi-echo (MSME) sequence (9%) and one study used FLAIR imaging (2%). DWI sequences included conventional SE (6 of 23 protocols, 26%), SE echo-planar imaging (4 protocols, 17%), stimulated echo (1 protocol, 4%), and FSE (1 protocol, 4%). Surprisingly, there was not an important difference in the range of values reported for most parameters between different types of pulse sequences, as examined separately for each species; the prevalence of reporting did not differ much either. It is useful to note that TE was generally shorter for higher field strengths, while the slice thickness was slightly reduced; to the extent permitted by the relatively small number of studies reviewed, other parameters were not found to be dependent on field strength.

Table 3.1 MRI acquisition parameters for 45 T2WI protocols.

Parameter	Total reported, n/N (%)	Reported per animal species, Species: n/N (%)	Details per animal species, range (median)
Type of pulse sequence	33/45 (73)	Rat: 23/34 (68) Mouse: 9/10 (90) Baboon: 1/1 (100)	– (–) – (–) – (–)
Scan duration (min)	7/45 (16)	Rat: 5/34 (15) Mouse: 2/10 (20) Baboon: 0/1 (0)	1-9 (7.5) 5 (5) N/G (N/G)
TE (ms)	43/45 (96)	Rat: 32/34 (94) Mouse: 10/10 (100) Baboon: 1/1 (100)	30-120 (60) [‡] 30-120 (66) 105 (105)
TR (ms)	43/45 (96)	Rat: 32/34 (94) Mouse: 10/10 (100) Baboon: 1/1 (100)	1750-8000 (3000) 1000-3000 (2550) 3000 (3000)
Number of slices	27/45 (60)	Rat: 18/34 (53) Mouse: 9/10 (90) Baboon: 0/1 (0)	6-35 (14) [‡] 5-25 (13) N/G (N/G)
Slice thickness (mm)	44/45 (98)	Rat: 33/34 (97) Mouse: 10/10 (100) Baboon: 1/1 (100)	0.35-3 (1.5) 0.5-1 (1) 3 (3)
FOV* (mm×mm)	38/45 (84)	Rat: 30/34 (88) Mouse: 7/10 (70) Baboon: 1/1 (100)	22×22-120×120 (30×30) 25.6×12.8-20×20 (20×20) 200×150 (200×150)
Matrix size* (pixels×pixels)	34/45 (76)	Rat: 28/34 (82) Mouse: 5/10 (50) Baboon: 1/1 (100)	64×64-512×512 (256×256) 128×64-256×256 (256×128) 256×192 (256×192)
NSA	12/45 (27)	Rat: 9/34 (26) Mouse: 2/10 (20) Baboon: 1/1 (100)	2-8 (4) 8-16 (12) 1 (1)
ETL for FSE [†]	5/14 (36)	Rat: 3/8 (38) Mouse: 1/5 (20) Baboon: 1/1 (100)	8-16 (8) 8 (8) 10 (10)
TI (ms) for FLAIR [†]	0/1 (0)	Rat: 0/1 (0)	N/G (N/G)

Dash (–) indicates that the value cannot be calculated; ETL, echo train length; FLAIR, fluid-attenuated inversion recovery; FOV, field of view; FSE, fast spin echo; n, number of protocols for which the parameter is reported; N, total number of protocols for which the parameter is relevant; NSA, number of sample averages; N/G, not given; TE, echo time; TI, inversion time; TR, repetition time.

*The range was estimated by sorting the area (product) defined by the two values, and the most frequently reported value is given instead of the median.

[†]Parameters applicable only to specific types of sequences.

[‡]Some articles reported a range of values only; the reported minimum and maximum values were not used for estimating the median, but were incorporated into the estimated range.

Table 3.2 MRI acquisition parameters for 23 DWI protocols.

Parameter	Total reported, n/N (%)	Reported per animal species, Species: n/N (%)	Details per animal species, range (median)
Type of pulse sequence	12/23 (52)	Rat: 11/22 (50) Mouse: 1/1 (100)	– (–) – (–)
Scan duration (min)	2/23 (9)	Rat: 2/22 (9) Mouse: 0/1 (0)	7-15 (11) N/G (N/G)
TE (ms)	17/23 (74)	Rat: 16/22 (73) Mouse: 1/1 (100)	33-100 (40) 30 (30)
TR (ms)	18/23 (78)	Rat: 17/22 (77) Mouse: 1/1 (100)	1000-4000 (3000) 1000-1000 (1000)
Number of slices	12/23 (52)	Rat: 11/22 (50) Mouse: 1/1 (100)	4-13 (8) [‡] 5 (5)
Slice thickness (mm)	19/23 (83)	Rat: 18/22 (82) Mouse: 1/1 (100)	1-2 (1.5) 0.5 (0.5)
FOV* (mm×mm)	17/23 (74)	Rat: 16/22 (73) Mouse: 1/1 (100)	22×22-80×80 (30×30) 22×16 (22×16)
Matrix size* (pixels×pixels)	17/23 (74)	Rat: 16/22 (73) Mouse: 1/1 (100)	64×64-256×256 (128×128) 128×64 (128×64)
NSA	4/23 (17)	Rat: 4/22 (18) Mouse: 0/1 (0)	2-4 (3) N/G (N/G)
ETL for FSE [†]	1/1 (100)	Mouse: 1/1 (100)	8 (8)
Diffusion sensitivity, <i>b</i> -factor (s/mm ²)	18/23 (78)	Rat: 17/22 (77) Mouse: 1/1 (100)	20-1898 (1000) [‡] 70-1224 (647) [‡]
Diffusion gradient strength, <i>G</i> (mT/m)	6/23 (26)	Rat: 5/22 (23) Mouse: 1/1 (100)	0-152 (80) [‡] 120 (120)
Diffusion gradient duration, δ (ms)	8/23 (35)	Rat: 7/22 (32) Mouse: 1/1 (100)	3.15-10 (10) 5 (5)
Diffusion gradient separation, Δ (ms)	6/23 (26)	Rat: 5/22 (23) Mouse: 1/1 (100)	13-50 (24.7) 17.5 (17.5)

Dash (–) indicates that the value cannot be calculated; ETL, echo train length; FOV, field of view; FSE, fast spin echo; n, number of protocols for which the parameter is reported; N, total number of protocols for which the parameter is relevant; NSA, number of sample averages; N/G, not given; TE, echo time; TR, repetition time.

*The range was estimated by sorting the area (product) defined by the two values, and the most frequently reported value is given instead of the median.

[†]Parameter applicable only to specific types of sequences.

[‡]Some articles reported a range of values only; the reported minimum and maximum values were not used for estimating the median, but were incorporated into the estimated range.

3.3.3 MRI post-processing

Forty articles reported the use of an image analysis software for the measurement of lesion size (74%; Figure 3.3); 17 of these are academic (43%), 16 are commercial (40%) and 7 are custom-made tools (17%). For nine of the 23 diffusion protocols, ADC maps were created for assessing lesions (39%), while CBF maps were generated in all three perfusion studies (100%). Out of a total of 78 different imaging protocols, 51 gave some information regarding the method employed to determine the boundary of lesions in MRI images (65%); for 39 protocols a type of thresholding was used (76%), followed by manual tracing in 10 protocols (20%) and an automated contour tracing algorithm in just two (4%). However, no study employing manual analysis provided specific selection criteria for determining lesions. A range of thresholds were reported for T2WI, DWI and PWI data analysis, including absolute diffusion or perfusion thresholds for the latter two methods, or thresholds based on the intensity in the contralateral healthy brain hemisphere for all three methods. Voxels in T2WI data were commonly selected if intensity was higher than 2 SDs above the mean intensity in the healthy contralateral hemisphere, or higher than 1.25 times this mean intensity; no study provided reasonable justification for their use. A reduction of ADC by 20% or 29% compared to the contralateral values was often classified as abnormal, but only those who used the latter threshold justified its use: the selected region correlates with post-mortem infarct volume. All studies which analysed T2WI data via thresholding used volume RF receive coils, except for three studies which used surface coils to acquire quantitative T₂ maps that are free of intensity inhomogeneities.

An important confounding parameter to the estimation of infarct size from T2WI data is brain swelling, arising due to vasogenic oedema in the subacute and chronic phases of stroke; 44 of the 45 imaging protocols were used at time points after MCAO when brain oedema might have been present. Nevertheless, only for 11 of these (25%) a method for correction was reported; all are intended for use in histological assessment of infarct size and not MRI. Nine studies (82%) applied the formula by Swanson et al. (1990), while the other two studies (18%) normalised the observed lesion size by the

ratio of hemispheric volumes according to Leach et al. (1993). Furthermore, the quality of data analysis was poor in the vast majority of studies; in addition to infrequent blinding of image analysis (Figure 3.2), only one article reported analysis by multiple independent raters (2%; Figure 3.3). Lesion size was usually reported in articles as an absolute estimate of volume (for example mm³; 45 protocols, 58%), followed by a range of relative estimates. Table B.3 in Appendix B provides further details for various MRI post-processing parameters.

3.3.4 Meta-analysis

To determine the relationship between various MRI methods and current gold standard post-mortem histology for the estimation of infarct size, studies which reported use of a histological method right after MRI in the same cohort of animals were further selected. Fifteen out of the 54 reviewed publications used histology (28%); two of these did not kill the animals for histological analysis right after MRI and another study did not provide infarct size measurements for a control group, thus were excluded (Figure 3.1). Characteristics of the remaining 12 studies are given in Table 3.3. Forty-six MRI and 46 histology comparisons between corresponding treatment and control groups are described in these studies, comprising of 428 animals that were assessed with MRI, 228 of which were then sacrificed for histology. Specifically, 11 studies performed T2WI (33 comparisons, 410 animals for T2WI and 210 for histology) and six performed DWI (11 comparisons, 147 animals for DWI and 105 for histology). Only one article reported T1WI and PWI (1 comparison each, 19 animals for MRI and histology); the relationship between histology and these specific imaging methods could not be examined. Histological assessment of infarct size was performed with TTC staining in 10 studies and H&E or cresyl violet staining in one study each. For meta-analysis, an ES was estimated for each comparison and then all effects were combined to estimate the pooled ES for each method. Meta-regression was used to assess the relationship between corresponding MRI and histology effects; detailed meta-analysis methodology is described in Appendix A.

Table 3.3 Characteristics of studies included in meta-analysis.

Study	Species	Histology method	MRI method	No. of comparisons	Time after stroke (hours)
Schabitz et al. (2001)	Rat	TTC	T2WI	1	168
			DWI	1	168
Goto et al. (2002)	Mouse	Cresyl violet	DWI	1	21
Nomura et al. (2005)	Rat	TTC	T2WI	3	168
Ding et al. (2006)	Rat	H&E	T2WI	1	48
			T1WI	1	48
			DWI	1	48
			PWI	1	48
Honma et al. (2006)	Rat	TTC	T2WI	4	168
Horita et al. (2006)	Rat	TTC	T2WI	3	168
			DWI	3	168
Liu et al. (2006)	Rat	TTC	T2WI	3	168
			DWI	3	168
Ukai et al. (2007)	Rat	TTC	T2WI	2	174
			DWI	2	174
Li et al. (2008)	Rat	TTC	T2WI	2	336
Omori et al. (2008)	Rat	TTC	T2WI	4	336
Onda et al. (2008)	Rat	TTC	T2WI	3	168
Toyama et al. (2009)	Rat	TTC	T2WI	7	168

DWI indicates diffusion-weighted imaging; H&E, haematoxylin and eosin; T2WI, T₂-weighted imaging; TTC, 2,3,5-triphenyl-2H-tetrazolium chloride.

Table 3.4 shows the combined estimate of efficacy as an NMD from all MRI comparisons, MRI comparisons that were nested for identical cohorts (see Appendix A), T2WI and DWI comparisons, along with the corresponding histological comparisons. Overall, the reduction in lesion size assessed by the first two MRI approaches is almost identical to the reduction observed histologically. T2WI gives a slightly more conservative estimate of efficacy (22.2%) compared to histology (25.5%), whereas the improvement in lesion size for DWI is 42.4%, noticeably larger than the 31.8% improvement determined by histology. The CIs of the T2WI estimate are narrower than histology, whereas DWI estimates vary more than histological estimates. The between-study heterogeneity in treatment effects is high, as suggested

by the significant p -values and large I^2 for all combined ESs in Table 3.4.

Meta-regression analysis is showcased in Table 3.5 and Figure 3.4. Overall, all imaging findings explain 56.6% of the observed variation in histological outcome, but nested MRI estimates in identical cohorts have a better correlation with histology. Nevertheless, the relationship is stronger for T2WI alone (explains 69.9% of variation in histology), whereas DWI performance is the poorest. Inclusion of the time of infarct assessment as an additional independent variable is not influencing the correlation between all MRI findings and histology, but improves the performance of T2WI and DWI meta-regression models.

Table 3.4 Combined effect sizes based on MRI and histology infarct size.

Method	No. of comparisons	No. of animals	Combined NMD [95% CI] (%)	p -value	I^2 (%)
All MRI	46	428	26.7 [22.6, 30.8]	< 0.001	92.9
Histology	46	228	27.5 [21.3, 33.7]	< 0.001	89.7
Nested MRI*	34	428	23.3 [18.8, 27.8]	< 0.001	93.2
Histology	34	228	26.0 [18.0, 33.9]	< 0.001	91.6
T2WI	33	410	22.2 [17.9, 26.5]	< 0.001	93.7
Histology	33	210	25.5 [17.5, 33.5]	< 0.001	91.8
DWI	11	147	42.4 [30.9, 54.0]	< 0.001	72.2
Histology	11	105	31.8 [24.0, 39.7]	0.002	64.7

CI indicates confidence interval; DWI, diffusion-weighted imaging; I^2 , between-study heterogeneity (equation A.26); NMD, normalised mean difference; T2WI, T₂-weighted imaging.

*If a study used various MRI methods to scan the same cohort of animals at the same time, effect sizes were pooled to create a nested effect size; the final combined estimate was estimated from nested and normal estimates from the rest of the studies (see Appendix A).

Table 3.5 Meta-regression analysis of corresponding MRI and histology effect sizes.

Method	Time not a covariate			Time included as a covariate		
	Coefficient [95% CI]	<i>p</i> -value	\bar{R}^2	Coefficient [95% CI]	<i>p</i> -value	\bar{R}^2
All MRI	0.76 [0.52, 1.00]	< 0.001	0.566	0.76 [0.52, 1.00]	< 0.001	0.554
Nested MRI*	1.04 [0.74, 1.34]	< 0.001	0.672	1.07 [0.77, 1.37]	< 0.001	0.680
T2WI	1.08 [0.77, 1.38]	< 0.001	0.699	1.13 [0.83, 1.44]	< 0.001	0.724
DWI	0.31 [-0.12, 0.73]	0.136	0.433	0.36 [-0.05, 0.78]	0.079	0.534

Histology was used as the dependent variable. CI indicates confidence interval; DWI, diffusion-weighted imaging; \bar{R}^2 , adjusted coefficient of determination; T2WI, T₂-weighted imaging.

*If a study used various MRI methods to scan the same cohort of animals at the same time, effect sizes were pooled to create a nested effect size (see Appendix A).

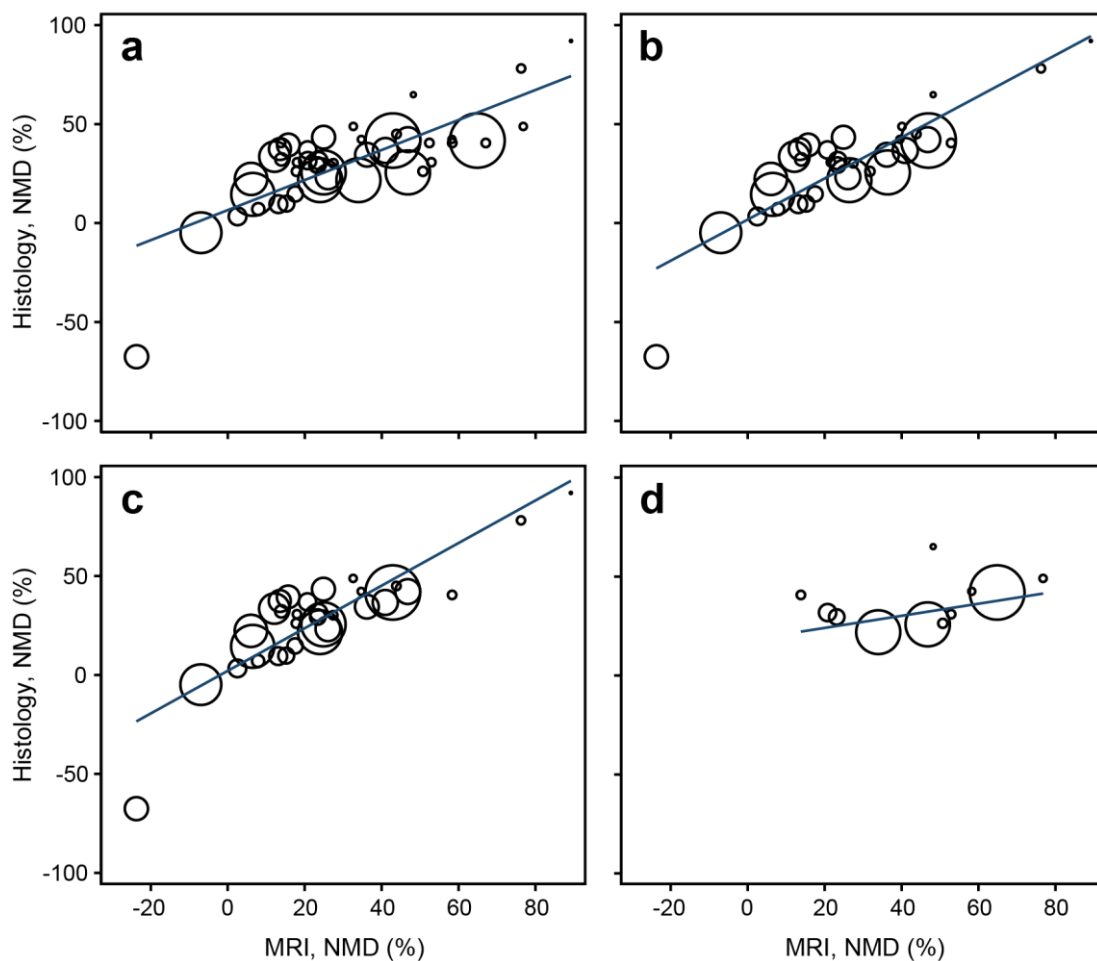


Figure 3.4 Meta-regression of corresponding histological and MRI effect sizes. (a) All MRI, (b) nested MRI, (c) T2WI, and (d) DWI effect sizes were used as the independent variables. The size of the circles reflects the precision of each estimate; large circles indicate small standard error and thus larger weight in the regression. The coefficients of regression and determination are given in Table 3.5. NMD indicates normalised mean difference.

3.4 Discussion

This systematic review of studies in animal models of stroke reveals great variability in MRI scanning parameters and post-processing methods used to assess lesion size. Reporting of methodology was assessed against several important items proposed by generic recommendations (Stout et al., 2013) and other items applied specifically to stroke; these covered the animal handling procedure for scanning, pulse sequences used and image analysis methods. Information about scanning equipment and basic pulse sequence parameters, such as TR, TE and slice thickness, were adequately reported, but information regarding the animal positioning, the physiological parameters monitored and methods for ensuring accurate and reproducible image analysis was limited. The heterogeneity in imaging systems and software for analysis is not in itself a cause for criticism and would not pose serious threats to the internal validity in single centre studies, but limits their external validity and may affect comparisons in multicentre studies, as suggested by various clinical MRI studies (Kivrak et al., 2013; Takao, Hayashi, & Ohtomo, 2013; Wurnig et al., 2013). However, detection bias in analyses carried out in the reviewed studies was potentially large, as only in one study image analysis was repeated by two independent raters and in only nine studies blinding of this procedure to group allocation was performed, introducing dubiety to reported measurements. Study quality was assessed using a standard approach (Macleod et al., 2004); most items were often not reported, despite the availability of recommendations before publication of the majority of articles (Macleod et al., 2009; Stroke Therapy Academic Industry Roundtable (STAIR), 1999). It is known that experimental stroke studies which do not report these study quality items may overstate treatment effects (Macleod et al., 2008).

Meta-analysis aimed to determine the reliability of MRI methods for determining infarct size, by examining the correlation of treatment effects based on infarct sizes measured by MRI and gold standard histology. Study characteristics and design were matched across the examined comparisons as much as possible by considering measurements performed in the same cohort of animals at the same time using both

methods. In addition, it was assumed that the various histological methods used provide equivalent estimates of infarct size (Barone et al., 1991; Bederson et al., 1986; Tureyen, Vemuganti, Sailor, & Dempsey, 2004). Excess heterogeneity between studies was accounted for by using random effects meta-analysis (DerSimonian & Laird, 1986). Due to the small number of studies satisfying the meta-analysis inclusion criteria, assessment of infarct size at various time points following stroke was considered (Table 3.3). This is useful in capturing possible temporal changes in outcome; indeed, the inclusion of time in the T2WI and DWI meta-regression models improved their predictability, suggesting that infarct-based treatment effects vary during the first two weeks after the onset of ischaemia. This finding emphasises that the selection of appropriate time points for assessing outcomes is a crucial element of the design of animal studies. While early time points help minimise attrition and reduce overall study costs, treatment effects measured early might be of poor validity and usefulness. The pronounced change in the adjusted coefficient of determination in the DWI meta-regression model could imply that this method is a less reliable indicator of the effectiveness of tested interventions, but it might have arisen from the small number of comparisons used, which precludes derivation of safe conclusions. Predictability of nested MRI comparisons was superior to that of all MRI comparisons, possibly due to the absorption of DWI effects by the more accurate T2WI for studies that used both.

Nevertheless, meta-analysis suggests that T2WI can be used effectively as an *in vivo* alternative to histology for the assessment of infarct size at least up to two weeks following stroke, despite the fact that T2WI-based measurements of absolute infarct volume in primary articles were often larger than corresponding histology estimates. This could be attributed to the observation that none of the included studies applied a correction to take into account the increased vasogenic oedema, which is not present in post-mortem histological slices as fixation may shrink the lesioned tissue (Barone et al., 1991; Gerriets et al., 2004). The narrower confidence limits of the combined T2WI-based effects suggest that this method may also have greater statistical power

than histology, and hence require fewer animals.

In contrast, DWI lesion volume did not correlate well with histological infarct volume, giving larger estimates of ES. This method is sensitive to the restricted diffusion of extracellular water due to the abnormal swelling of cells (cytotoxic oedema) at the early stages of stroke progression (Campbell & Macrae, 2015; Dijkhuizen & Nicolay, 2003; Farr & Wegener, 2010; Wardlaw, 2010). According to the meta-analysed studies, DWI lesions were larger than corresponding histology lesions at 21 hours and smaller in the chronic phase of stroke, agreeing with a previous systematic review suggesting that the extent of the infarct generally increases despite early DWI lesion normalisation (Rivers & Wardlaw, 2005).

This study intended to use stratified meta-analysis to examine the impact of individual imaging parameters on meta-regression models to refine the relationship between MRI and histology. The small number of studies included in meta-analysis and the variety of experimental methods used (related to species, stroke model and MRI methods) precluded such analysis, since stratification of studies by various criteria often led to groups consisting of a single study. The significant variability in MRI methodology raises questions about the best way to apply each MRI contrast technique, and whether measurements from different studies are comparable. It should be acknowledged, however, that the optimal scanning protocol must be a compromise between resolution, anatomical coverage, contrast between normal and injured tissue, and speed, in order to allow accurate quantification of lesion size while minimizing the duration of anaesthesia received by animals for MRI. As shown by this systematic review, the values of many imaging parameters should be chosen at least according to the animal species and the field strength of the scanner used.

This might be the first systematic review and meta-analysis of MRI in animal models of stroke, but it comes with a few limitations. Firstly, there is a lack of specificity by including studies utilizing any MRI technique for assessing various types of lesions at various time points. The primary objective of the study was to provide an overview

of the use of the modality in preclinical stroke research and examine whether this could be a limiting factor in data pooling in meta-analyses or multicentre animal studies; findings clearly support this hypothesis. Future studies should focus on individual MRI techniques or outcomes. Rivers and Wardlaw (2005) reviewed the use of DWI in animal stroke studies and identified a similarly incomplete reporting of experimental details. They also examined whether DWI could provide information regarding tissue viability by comparison with various histopathological markers, but their assessment was limited in description of findings reported in reviewed papers. Dani et al. (2012) assessed the reporting of PWI in human stroke and found poor consistency between studies. Perfusion imaging co-registered with diffusion imaging is a technique for identifying the salvageable ischaemic penumbra (represented by the mismatch of the two images), which provides a hypothetical target for reperfusion therapies. This concept has been subject to significant scrutiny in both clinical and preclinical research but no widely accepted definition exists (Dani et al., 2011). This could partially explain the small number of PWI protocols recorded in this review. There have been calls for consensus in the use of DWI and PWI for defining the penumbra in humans and in animals (Campbell & Macrae, 2015; Rivers & Wardlaw, 2005); however, there is a more pressing case, demonstrated here, for consensus in the use of more routine MRI techniques.

Secondly, assessed studies were identified in a single database comprising a non-exhaustive list of studies in animal models of stroke. However, these were previously identified and included in this database based on criteria that did not relate to the imaging methodology, and therefore the articles examined here can be considered representative of experimental stroke imaging studies in the literature. Few experiments assessed infarct size using both histology and MRI at corresponding time points (46 of 5329 comparisons), and this did not allow exploration of the impact imaging parameters have on the relationship between MRI and histological outcome measures. A systematic review seeking to identify all such studies would enable such analysis, but is likely to require screening of a large number of publications.

This study suggests that the variability between studies in the use of MRI for the assessment of lesion size is significant. A strong correlation between histology and T2WI was found for estimating treatment effects based on infarct size, suggesting that utilising a standardised T2WI scanning protocol and associated data post-processing techniques for quantification of infarct size is feasible and should be considered in future multicentre animal studies. Acknowledging the insufficient methodological description in imaging papers, Stout et al. (2013) recently published generic guidelines for the reporting of imaging information in preclinical papers. However, specific guidelines for the use and reporting of MRI methodology in experimental stroke are needed, similar to those developed for the conduct of human stroke imaging (Wintermark et al., 2008; Wintermark et al., 2013). This would permit exact replication of the experiments and analyses and would support the use of imaging endpoints in multicentre animal studies.

Chapter 4

Construction of an MRI Phantom and Development of a Data Analysis Tool

This chapter describes the development of tools for the assessment of geometric accuracy in preclinical MRI scanners in the context of stroke imaging. Specifically, the chapter presents the procedures regarding the construction of a small sMRI phantom and the development of an automated method for the analysis of related MRI images. Phantom scanning is performed using a standardised pulse sequence based on rat T2WI, as recorded in the preceding systematic review. The tools described here are used in Chapter 6 to perform a multicentre MRI study.

Parts of this chapter have been presented as:

Milidonis, X., Lennen, R. J., Jansen, M. A., & Marshall, I. (2014). Gradient scaling on preclinical magnetic resonance imaging scanners assessed with a dedicated phantom. *Presented at the 20th Annual Scientific Meeting of the British Chapter of the International Society for Magnetic Resonance in Medicine (ISMRM)*. Edinburgh, UK.

Milidonis, X., Lennen, R. J., Jansen, M. A., & Marshall, I. (2014). Gradient scaling on preclinical magnetic resonance imaging scanners assessed with a dedicated phantom. *Presented at the 6th Annual Scientific Meeting of the Scottish Imaging Network: A Platform for Scientific Excellence (SINAPSE)*. Edinburgh, UK.

4.1 Introduction

Chapter 3 revealed a large heterogeneity between studies in the use of MRI in experimental stroke. However, the true impact of this heterogeneity on the assessment of infarct size or any other morphological outcome can only be assessed by scanning the same subjects using different scanners or protocols. One way to perform this is by scanning a group of animals that have been subjected to cerebral ischaemia. However, cerebral ischaemic damage is a spatiotemporally evolving process happening at a fast rate during the first few days post-onset (Loubinoux et al., 1997; Neumann-Haefelin et al., 2000; Wegener et al., 2006; Zanier et al., 2013), while many animals die in later stages of the disease; such variability does not allow reproducible scanning of the same animal. In this case, a QA approach by scanning an MRI phantom would be more suitable; a phantom is a temporally stable object with known characteristics that permits reliable assessment of outcomes between or within scanners.

Despite that animal studies in stroke often measure morphological outcomes in sMRI data to evaluate the efficacy of tested interventions or examine new concepts of pathophysiology (Farr & Wegener, 2010; Zille et al., 2012), standardised QA approaches employing phantoms for monitoring and optimising the performance of scanners on a routine basis have not yet been developed. QA has almost always been an invaluable tool in clinical MRI practice, with performance tests carried out daily (American College of Radiology, 2015; Gunter et al., 2009). In preclinical centres, such tests are rarely performed by the personnel and most rely on annual maintenance of the systems by external engineers. As a result, systematic drifts in scanner performance during a lengthy stroke experiment may not be detected and accounted for. Such drifts could subsequently cause geometric distortions in images and further degradation of image quality, amongst others, with direct effect on the measurement of morphological outcomes. When these effects are combined with the large variability in scanning protocols and image processing methods used across centres, an effective comparison of outcomes across centres becomes almost impossible. The

impact of these shortfalls in experimental stroke will become more apparent in the future, when multicentre studies might be the new paradigm (Bath et al., 2009; Dirnagl & Fisher, 2012; Dirnagl et al., 2013). Multicentre studies have high demands for accurate and reproducible measurements in all participating centres for efficient pooling of data and derivation of valid conclusions, but unless the use of MRI is of a sufficient standard these requirements cannot be met.

The main objective of this thesis is to examine how geometric accuracy in preclinical scanners affects the quantification of infarct size and relevant comparisons between scanners. Towards this endeavour, a multicentre MRI QA study examining the between- and within-scanner variability was designed and carried out (Chapter 6); this chapter aims to describe the development of tools and methods that were central to this study. A simple structural phantom and accompanying methods for the acquisition and unbiased analysis of phantom data are presented, with the assessment of geometric accuracy being based on the measurement of two QA criteria: dimensions and volume. The development of the phantom scanning protocol was driven by findings of the systematic review and meta-analysis presented in Chapter 3; specifically, the phantom resembles the size and T_2 relaxation time of the rat MCA territory—the most widely used species in experimental stroke—and the QA pulse sequence is based on recorded T2WI parameters in this species (Table 3.1). The work presented here could encourage the introduction of simple and rigorous QA protocols in preclinical imaging centres and facilitate future multicentre stroke studies.

4.2 Quality Assurance Protocol

4.2.1 Phantom

Ideally, assessment of between- and within-scanner variability in a multicentre QA study requires construction of an identical phantom for each centre; based on this, a number of phantoms were constructed according to the following criteria:

- high inter-phantom precision to allow comparable assessment of geometric accuracy
- size similar to that of the rat MCA territory and small enough to be used with any type and size of RF coil for rat MRI
- durable and with minimal water absorption to be temporally stable
- made with materials with low magnetic susceptibility to minimise distortion artefacts
- easy to construct with cheap and readily available materials

Probably the only materials that satisfy all these criteria are LEGO® (Billund, Denmark) bricks; there are two types of plastics used for their manufacturing, transparent polycarbonate and opaque acrylonitrile butadiene styrene (ABS) polymers. These polymers are diamagnetic like water (Osswald & Menges, 2010; Wapler et al., 2014) and have high impact strengths and minimal water absorption properties (0.15% and 0.20-0.40% over 24 hours respectively) (Shackelford, Han, Kim, & Kwon, 2015). LEGO bricks are manufactured by injection moulding using high precision moulds, giving the pieces an excellent final geometrical tolerance of just 20µm (Quercioli, Tiribilli, Mannoni, & Acciai, 1998).

For the phantoms, 2×2 bricks made of transparent polycarbonate were chosen (LEGO design no. 3003; Figure 4.1 a). To seal the bottom side of the bricks, 2×2 flat tiles made of opaque ABS were chosen (LEGO design no. 3068). Using a heated metallic wire, a shallow hole was made at a specific location on the surface of each tile to act as an orientation marker for scanning (Figure 4.1 a). The central cylindrical compartment

of the transparent bricks is slightly shorter than the sides of the bricks; to ensure full contact of each tile with the central cylindrical compartment of the transparent bricks, four layers of transparent tape of size about $6 \times 6 \text{ mm}^2$ were attached to the centre of each tile (Figure 4.1 a).

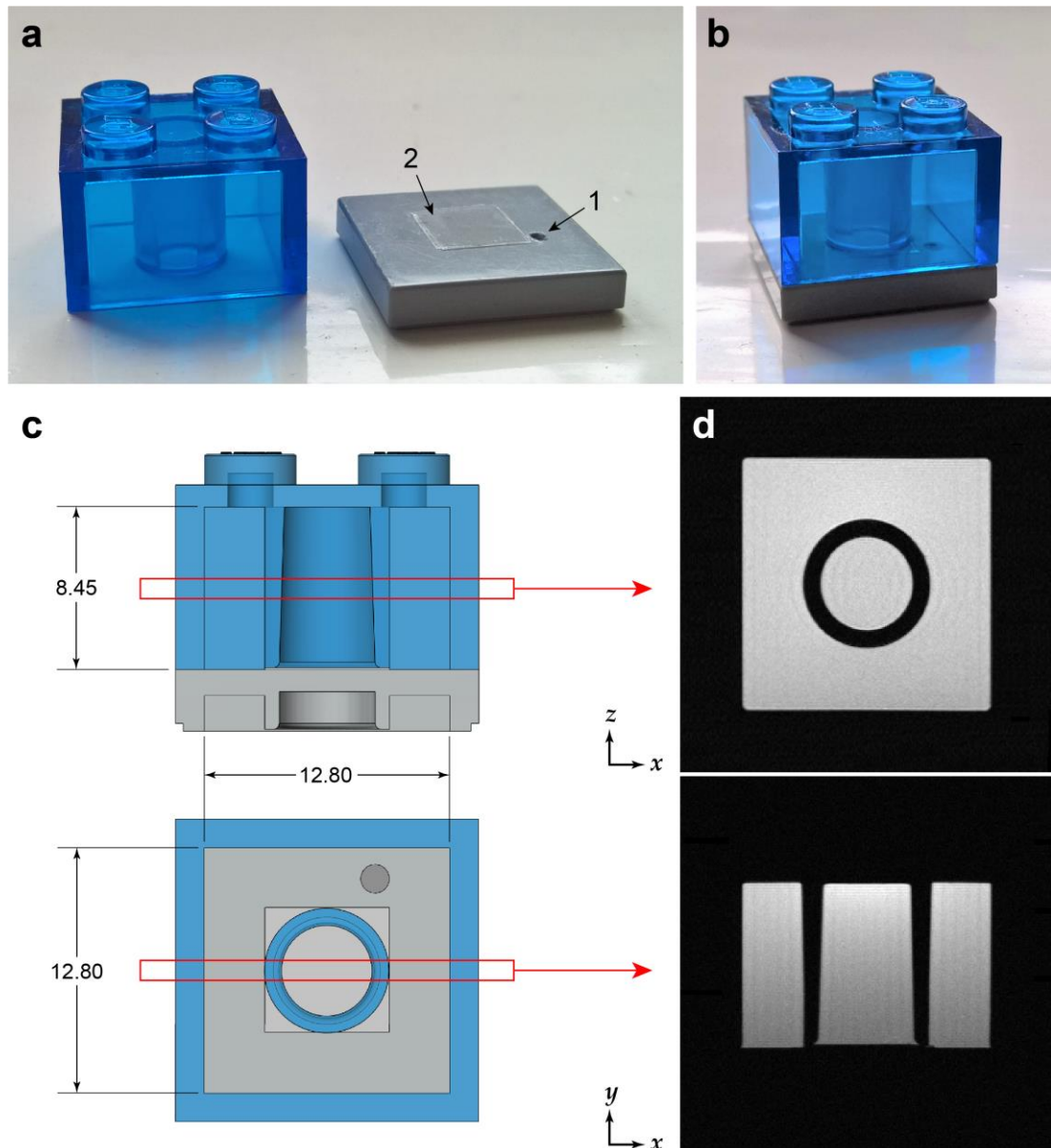


Figure 4.1 A simple structural phantom for assessing geometric accuracy in preclinical MRI scanners.

Images show (a) the two different LEGO bricks that make up the phantom, (b) the completed phantom, (c) cross sections indicating its internal dimensions (in mm), and (d) example MRI images acquired with a rat brain surface RF coil (top, coronal; bottom, axial/sagittal). The arrows in (a) indicate (1) the orientation marker and (2) transparent tape that ensures full contact between the cylindrical compartment of the transparent plastic and the opaque tile.

The transparent brick of each phantom was filled by submerging it into distilled water doped with Gd 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (Gd-DOTA; Dotarem®, Guerbet, Roissy CdG Cedex, France) at a concentration of 3 mmol/L (6 mL/L). The use of a contrast agent is necessary to shorten both relaxation times in order to allow use of a QA pulse sequence with short TR and decrease the overall QA duration. The concentration of 3 mmol/L was chosen according to a published table (Noebauer-Huhmann et al., 2008) to create a solution with a T_2 relaxation time similar to that of the healthy striatum in the rat brain at 7 T (Fagan et al., 2008). Gd-based contrast agents are often used in MRI phantoms, including a structural preclinical phantom developed recently by others (Teh, Maguire, & Schneider, 2016).

The flat tile was firstly placed over the filled transparent plastic and acetone was applied at the joints to temporarily hold them together. Then, a thin layer of epoxy resin was applied to permanently join them and seal the phantom.

4.2.2 Quality assurance criteria

The constructed phantoms enable assessment of geometric accuracy in preclinical MRI scanners via measurement of two criteria:

1. The internal dimensions across all three orthogonal directions, approximated by the distance between opposing edges horizontally or vertically in the images (Figure 4.1 c). Measurements in y direction (vertical in axial/sagittal image in Figure 4.1 d) are obtained along the sides of the phantom to avoid confounding by the tape attached to the centre of the base tile. The overall size of the internal phantom area is similar to that of the rat MCA brain territory (10-15 mm in any direction).
2. The volume of the central cylindrical frustum-shaped compartment, measured by identifying the relevant region in all slices and summing all of its voxels (the central region in images in Figure 4.1 d). This volume is similar to that of a medium-sized cortical and subcortical infarct, or a large subcortical

infarct in the rat MCAO model (100-200 mm³).

The internal dimensions of all constructed phantoms were measured before filling using callipers to verify inter-phantom precision. To measure the true volume of the cylindrical compartment of the phantom, three phantoms were constructed as described, but without filling them with liquid solution or attaching them permanently. 3-dimensional (3D) CT images of the empty phantoms were obtained using a nanoScan[®] PET/CT scanner (Mediso Ltd., Budapest, Hungary) with an isotropic resolution of 34.5 μm (35 kVp beam energy, 1 mA tube current, 450 ms exposure time). Using ImageJ (1.50b; Rasband, W.S., National Institutes of Health, Bethesda, Maryland, USA, <http://imagej.nih.gov/ij/>) the CT data were first smoothed with a Gaussian filter with an SD of 1 pixel (81.2 μm full width at half maximum (FWHM)) and then binarised using an appropriate intensity threshold to produce masks of the phantom with exactly the calliper-measured dimensions. The volume of the cylindrical compartment in all three phantoms was measured from the thresholded CT data and was found to be very precise; the mean value was considered as the ground truth (mean \pm SD = 153.1 \pm 0.4 mm³). CT scanning was performed with the assistance of Dr Adriana Tavares (University of Edinburgh).

4.2.3 MRI scanner and equipment

All experiments described in this chapter were carried out on an Agilent 7 T horizontal bore animal scanner (Agilent Technologies[®], Santa Clara, California, USA) with a magnet bore with 305 mm inner diameter (ID), located in the Edinburgh Preclinical Imaging (EPI) facility at the University of Edinburgh. An actively shielded gradient insert with a maximum gradient strength of 400 mT/m and ID of 120 mm was used. All MRI scanning described in this chapter was performed using either a 72 mm ID volume coil for RF transmission and reception, or this volume coil for transmission and a rat head two-channel phased array surface coil for reception (RAPID Biomedical GmbH, Rimpar, Germany). The software used for data acquisition was VnmrJ[®] (3.2, Agilent). All equipment used is shown in Figure 4.2.

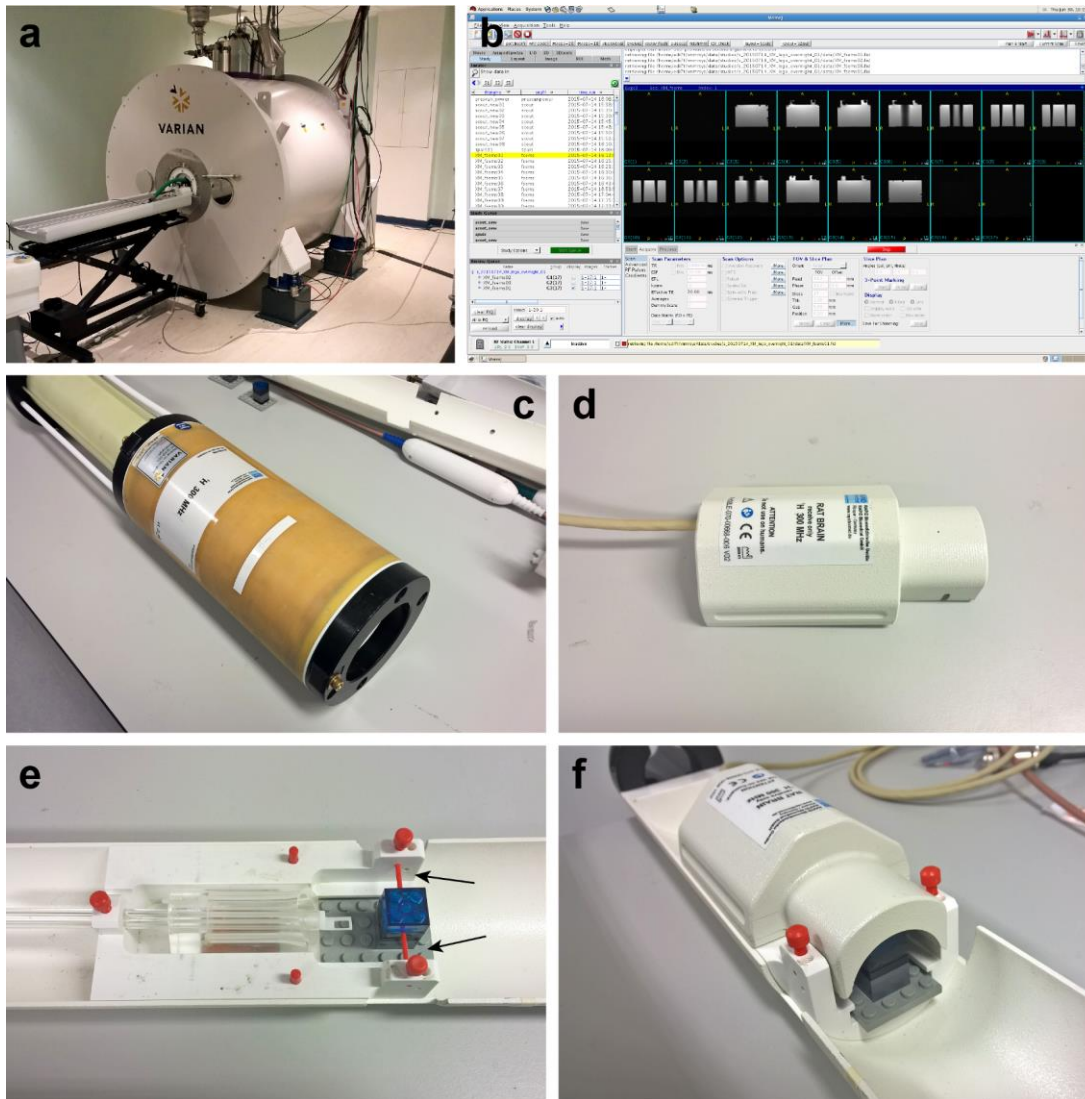


Figure 4.2 The preclinical MRI scanner and equipment used for phantom imaging.

Images show (a) the 7 T Agilent scanner, (b) the VnmrJ software for data acquisition, (c) the volume coil for signal transmission/reception, (d) the rat head surface coil for signal reception, (e) the cradle for rat imaging with arrows indicating the ear bar holders, and (f) the surface coil placed on the cradle. The volume coil is placed in the centre of the scanner and then the cradle with the surface coil is inserted in the volume coil, so that the phantom is positioned at the isocentre of the scanner. If the volume coil is used for both signal transmission and reception, the cradle is inserted in the volume coil without the surface coil.

For scanning, the phantom is placed in the rat cradle with the opaque tile at the bottom and is kept in place using stereotaxic ear bars, similar to the procedure followed for rat brain imaging (Figure 4.2 e). For phantom alignment and positioning close to the surface coil, additional LEGO bricks are placed beneath the phantom, while masking tape is also used for further immobilisation. If the surface coil is used

for signal reception, it is placed on top of the phantom and the assembled cradle is positioned at the scanner's isocentre within the volume coil.

4.2.4 T₂ relaxometry

To measure the T₂ relaxation time of the phantoms' solution, the volume coil was used for signal transmission and reception and a MSME sequence was used for capturing T₂ relaxation (*mems* protocol in VnmrJ; parameters in Table 4.1). T₂ relaxation follows an exponential decay curve over time (Figure 2.2 e); this is described by the following equation:

$$S = S_0 e^{-TE/T_2} + S_{offset} \quad (4.1)$$

where S is the signal corresponding to the relaxing M_{xy} , S_0 is the maximum signal corresponding to M_0 and S_{offset} is an offset factor taking into account any potential instrumental imperfections. The MSME sequence generates an image for each TE, the signal in which is proportional to M_{xy} over time. MATLAB® (2015a; The MathWorks® Inc., Natick, Massachusetts, USA) was used to take measurements of the mean

Table 4.1 Parameters of the phantom's T₂ relaxometry pulse sequence.

Parameter	Value
Type of sequence	MSME
TR (ms)	1500
TE* (ms)	10-200
NSA	1
FOV (mm×mm)	19.2×19.2
Matrix size (pixels×pixels)	64×64
Slice thickness† (mm)	2
Number of slices	5
Scan duration (min:sec)	1:36

FOV indicates field of view; MSME, multi-slice multi-echo; NSA, number of sample averages; TE, echo time; TR, repetition time.

*Twenty echo times with an echo spacing of 10 ms were used.

†No inter-slice gap was used.

intensity, S , within a circular region of interest (ROI) with 3 mm diameter at the centre of the phantom in all 20 images of the middle slice. Intensity measurements were then plotted against TE and equation 4.1 was fitted to the points to estimate the T_2 relaxation time using MATLAB's Curve Fitting Tool.

4.2.5 Quality assurance scanning protocol

Reviewed literature in Chapter 3 revealed that rats are mostly used in experimental studies of stroke, with conventional SE and FSE being the most prevalent pulse sequences for 2-dimensional (2D) T2WI for the assessment of infarct size. Based on these observations, a protocol was devised for the assessment of geometric accuracy in preclinical scanners. FSE was chosen over SE for its multifold improvement in speed, and sequence parameters were chosen based on parameters for rat T2WI, as shown in Table 3.1; parameters were adjusted to reach a compromise between image quality and speed of acquisition for imaging with either of the volume or the surface coil (*fsems* protocol in VnmrJ; parameters in Table 4.2).

For full assessment of geometric accuracy in a preclinical scanner, the phantom is scanned using this sequence and the preferred RF coil/s six times, once in each primary imaging plane (axial, coronal and sagittal) and all repeated with flipped frequency and phase encoding directions in the imaging plane to record any influence of spatial encoding. Acquired images are analysed to measure the two QA criteria and compare them with the ground truth values; the internal dimensions of the phantom are measured in all six scans and the volume of the phantom's central compartment in the axial scan alone, which represents the coronal plane used in rodent brain imaging (the x - y plane in Figure 2.3; in phantom scanning the human convention is often used, hence this plane is the axial plane).

Table 4.2 Parameters of the standardised QA pulse sequence.

Parameter	Value
Type of sequence	FSE
TR (ms)	1600
TE _{eff} (ms)	20
NSA	2
ETL	4
rBW (kHz)	100
FOV (mm×mm)	19.2×19.2
Matrix size (pixels×pixels)	256×256
Slice thickness* (mm)	1
Number of slices	17
Scan duration (min:sec)	3:28

A FSE sequence uses 180° refocussing RF pulses after the initial 90° excitation with a number equal to ETL, to produce a series of echoes and acquire multiple lines of *k*-space within each TR. ETL indicates echo train length; FOV, field of view; FSE, fast spin echo; NSA, number of sample averages; rBW, receiver bandwidth; TE_{eff}, effective echo time (TE corresponding to the echo when the central *k*-space lines determining the image contrast are acquired); TR, repetition time.

*No inter-slice gap was used.

4.3 Automated Image Analysis Tool

2D MRI data have a limited resolution, which could impact manual measurement of the two QA criteria and cause errors in comparisons between scanners where subtle differences in geometric accuracy are present; it is therefore important to acquire reliable and reproducible measurements in order to detect true differences between scanners. For the analysis of phantom data acquired in this chapter and the multicentre study presented in Chapter 6, a simple and fast automated tool was developed in MATLAB.

4.3.1 Analysis pipeline

An analysis pipeline was devised, utilising simple methods for the detection of the region representing the phantom in scans and the measurement of its internal dimensions or the volume of its central cylindrical compartment (Figure 4.3). Most analysis steps were used empirically and were adjusted until acceptable results were obtained. The main task of the analysis pipeline is to create accurate binary masks (black and white images) of the phantom; for this, three different methods have been implemented and compared, as described in the next section.

Initially, the intensity of the scan is normalised to the range 0 to 1. If the phantom is misaligned in the imaging plane, plain binarisation is performed to measure the alignment angle based on the slope of the phantom sides and identify the imaging plane based on the horizontal and vertical length of the phantom. The slices are then aligned and rotated respectively, to ensure equivalent analysis. The boundary of the phantom is determined with a segmentation process utilising a range of morphological steps (blue boxes in Figure 4.3). The process used for measuring the dimensions of the phantom involves scaling of the images using bicubic interpolation to perform subpixel analysis, if required; the precision of measurement improves proportionally to the scaling factor, but the tool's computational efficiency decreases.

The segmentation process for measuring the phantom's dimensions returns a binary mask of the phantom's boundary alone. The distance between opposing edges across four different locations in each direction is computed (horizontal and vertical), by identifying the pixel coordinates of opposing edges. For volume estimation, a filled version of the boundary mask is returned and the central compartment is identified to count included voxels. Both processes are guided by the location of the centroid (centre of mass of the binary object) of the whole phantom.

The analysis pipeline described above is repeated in all slices of the 2D scan that show the phantom's central cylindrical compartment, as determined automatically during the initial binarisation. The final dimension in each direction is the mode value of all

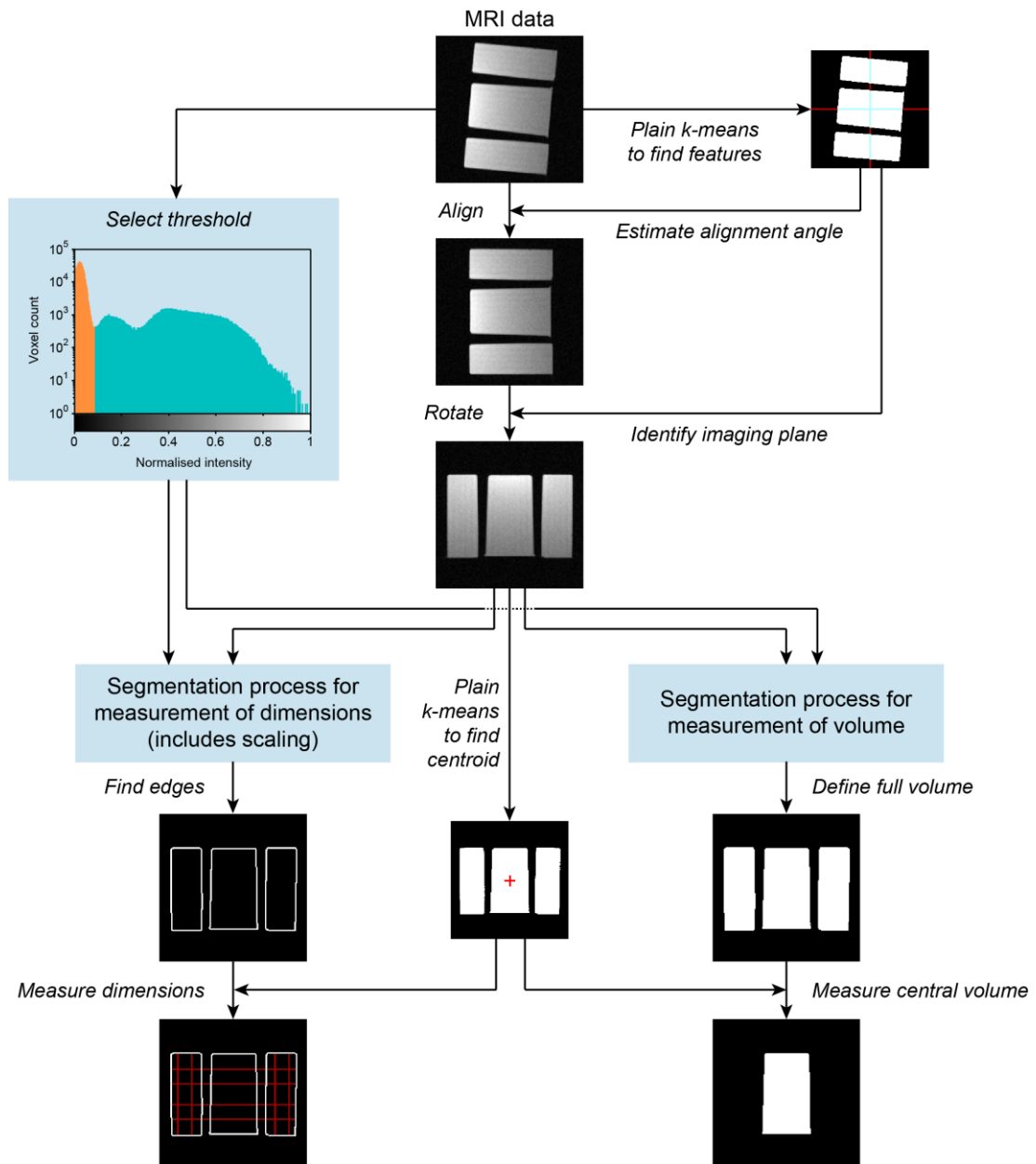


Figure 4.3 Automated pipeline for the analysis of phantom data. This is a simplified diagram showing the major processing steps used to analyse the phantom data automatically and measure its internal dimensions and the volume of its cylindrical compartment (an axial/sagittal scan is shown for illustration). Three different methods were applied and compared in the blue boxes, each utilising a series of morphological processing steps; threshold selection in the top blue box forms the basis of one of these methods. Only axial/sagittal scans are rotated to ensure equivalent analysis.

measurements (often 20-30 per direction); the mode is used to discard any abnormal measurements attributed to the phantom’s orientation marker or other artefacts in the images. The final volume is estimated by the sum of all detected voxels.

4.3.2 Segmentation methods

Various methods exist for segmenting objects in greyscale images. Since MRI images have background noise that is described by a known (Rician) distribution (Gudbjartsson & Patz, 1995) and phantom images are usually characterised by a high SNR throughout, three basic methods for segmentation were considered; these were implemented in MATLAB and were accompanied by standard morphological processing steps using dedicated MATLAB functions to accurately define the phantom (blue boxes in Figure 4.3). The three approaches were eventually compared to select the best method.

4.3.2.1 Simple thresholding

This is the simplest method implemented. The intensity histogram of the whole normalised scan is examined to find the first peak corresponding to background noise (Figure 4.3). The threshold, T , is then defined as

$$T = 5 I_{peak} \quad (4.2)$$

where I_{peak} is the intensity corresponding to the histogram peak; this threshold was found to best describe the distribution of background noise in scans acquired using either the volume coil alone or the volume-surface coil pair. Any voxels with an intensity higher than T are assumed to represent the phantom. Prior to and post application of the threshold, other processing steps are taken to refine the segmentation process, as described below.

Each rotated slice (Figure 4.3) is firstly smoothed using a small 2D Gaussian filter of size 3×3 and an SD of 1 pixel (FWHM = 2.35 pixels) to slightly reduce noise without distorting the phantom's features significantly. Then, the estimated threshold is applied to create a binary mask of the phantom. Large unwanted objects around the periphery of the image as well as other spurious pixels are removed to retain only the phantom mask. To measure the dimensions of the phantom, the mask is first dilated and eroded using a horizontal (3 pixels long) and a vertical (5 pixels long) linear

structuring element to flatten its boundary, and then the internal pixels of the mask are discarded to preserve only the boundary pixels. To measure the volume of the cylindrical compartment of the phantom, this region is first identified and retained based on the position of the mask's centroid and then its boundary is flattened as described.

4.3.2.2 *k*-means clustering

k-means is an iterative partitioning method that assigns a sample of objects into *k* mutually exclusive clusters (Lloyd, 1982), according to the following procedure:

1. Choose *k* initial centroids.
2. Calculate the distance of each object from each centroid and assign it to the cluster with the nearest centroid.
3. Compute the mean in each cluster to obtain *k* new centroids.
4. Iterate over steps 2-3 until the sum of distances within each cluster cannot be reduced further.

The result is the creation of clusters the objects in which are as close as possible around their centroid. In image segmentation, each pixel is an object, centroids correspond to intensities and partitioning is based on calculating the intensity difference between each pixel and each centroid. Here, MATLAB's **kmeans** function with default parameters was used to perform *k*-means clustering and group the pixels in the images to clusters representing the phantom and the background. The function accepts as inputs the image and *k* and employs a heuristic approach to find the initial centroids. It returns a vector containing the index of the cluster to which each pixel belongs to, but the indices are not sorted according to intensity.

k-means clustering with two and three clusters was tested; larger values of *k* make segmentation computationally expensive and the likelihood to produce errors (empty clusters) increases. Code was written to sort the vector with the cluster indices according to the intensity of corresponding pixels; for *k* = 2 the cluster with the higher

intensity represents the phantom and for $k = 3$ the two clusters with the higher intensity represent the phantom. The same pre- and post-processing morphological operations as in simple thresholding are used for k -means clustering as well. k -means with $k = 3$ but no additional processing steps was also used for binarisation in the main analysis pipeline, as indicated in Figure 4.3.

4.3.2.3 Canny edge detection

Edge detection is an image analysis method that identifies locations in an image where there are sharp changes in intensity (the intensity differs significantly between adjacent pixels). Most edge detection algorithms depict edges by simply estimating the first or second order derivative of the image to produce its gradient version. However, Canny (1986) devised a multistage approach that detects as many local maxima as possible while being less prone to errors due to noise:

1. Apply a Gaussian filter to smooth the image and remove noise.
2. Estimate the intensity gradient of the image by calculating local intensity differences (edges) in four directions.
3. Apply thinning to remove spurious results.
4. Apply a high and a low threshold to categorise remaining edges to one strong and two weak groups.
5. Preserve the strong edges and the weak edges from the middle group that are connected to the former.

In MATLAB, **edge** function is used, which accepts a two-element vector with the thresholds and the SD of the Gaussian filter, and returns a binary mask with the detected edges. The values of the input parameters are arbitrarily chosen based on the size and quality of the input image and the salience of features that need to be detected. If the thresholds are not given as a function input, **edge** automatically chooses and outputs two rough values, which can then be manually tuned to improve edge detection accuracy.

A default value of 2 is used by **edge** function for the SD of the Gaussian filter; this leads to false negatives if the image is small or its noise level is low, and to false positives in the contrary situation. To automatically estimate an appropriate value, a method that takes into account the signal characteristics and the size of the phantom images was determined. The histogram of the normalised scan is first examined to estimate a threshold using equation 4.2 and compute the mean intensity of all voxels with an intensity higher than the threshold, $I_{phantom}$, and the SD of the remaining voxels, SD_{noise} . The following formula was conceived to compute the SD of the Gaussian filter:

$$SD_{Canny} = \frac{\text{height of image}}{f_{SD}\sqrt{I_{phantom}/SD_{noise}}} \quad (4.3)$$

The height of the image is determined in pixels (256 for the QA protocol). The square root in the denominator is a measure of SNR; the effect of it is subtler than the size of the image, hence the square root is estimated. The factor f_{SD} is used to refine the formula to produce appropriate estimates of SD_{Canny} (normally within 1 to 3).

To reduce detection errors in the developed Canny edge detection pipeline for measuring the volume of the cylindrical compartment of the phantom, an iterated procedure is used to modify SD_{Canny} until the segmented region in each slice satisfies a criterion of shape. Initially, each slice is analysed with the parameters defined above to create a binary mask of the detected edges. Large unwanted objects around the periphery of the image are then removed. The remaining edges should be mostly of the phantom; these are filled to obtain the full region of the phantom, the boundary of which is flattened by dilation and erosion (as described in Section 4.3.2.1). The cylindrical compartment is then identified and retained based on the position of the phantom's centroid. Following segmentation, a shape factor, SF , is estimated as follows:

$$SF = \frac{4\pi A}{P^2} \quad (4.4)$$

where A is the area of the compartment given by the number of included pixels, P is its perimeter estimated using standard MATLAB functions and $\pi = 3.14159$. SF is called ‘isoperimetric quotient’ and increases as the circularity of the shape increases (Kremer & Weisstein, 2016); for example, it equals 0.79 for a square and 1 for a circle (maximum possible value). As the cylindrical compartment of the phantom appears as a rectangle in axial scans, a value of $SF = 0.4$ was deemed the minimum appropriate; while SF is less than 0.4, SD_{Canny} is increased by 0.1 and the procedure is repeated until convergence.

To measure the internal dimensions of the phantom, SD_{Canny} in equation 4.3 is modified to take into account the new size of the images following scaling:

$$\overline{SD}_{Canny} = SD_{Canny} \frac{\text{scaling factor}}{2} \quad (4.5)$$

This procedure is not iterated as it is less prone to errors than the volume measurement procedure; Canny edge detection with the thresholds and SD_{Canny} is applied and the produced binary mask is post-processed to remove spurious pixels and preserve only those at the boundary of the phantom.

To determine thresholds that result in detection of the boundary of the phantom while the majority of unwanted edges are excluded, the values produced by MATLAB were multiplied by a factor f_T ranging from 1 to 3 with a step of 0.5 while the value of 30 was set for f_{SD} . f_T was then kept stable at the determined value and f_{SD} was altered between 10 and 50 with a step of 10 to choose the optimal value.

4.3.3 Graphical user interface

Figure 4.4 shows a screenshot of the **PhantomAnalysisGUI** graphical user interface (GUI) developed in MATLAB for assessing the phantom scans interactively and comparing the segmentation methods described earlier. Canny edge detection is set as the default and related parameters are automatically estimated as soon as a scan is loaded, including voxel size and the slices to be analysed. A scaling factor of 5 is the

default for performing subpixel measurements of the phantom's dimensions. All these parameters are shown in the **Inputs** panel of the GUI and can be adjusted by the user on demand.

By clicking the buttons in the **Analysis** panel, the tool measures and outputs on the GUI the dimensions and the volume of the central compartment, and in MATLAB's Command Window the measurements for each slice. The detected boundary of the phantom is shown in the **Output Data** figure; the user chooses between overlaid edges or masks for display from a drop-down list. Example pictures showing the analysis procedure and the final masks for all analysed slices are saved in the image's folder by checking the **Save images?** box in the **Analysis** panel.

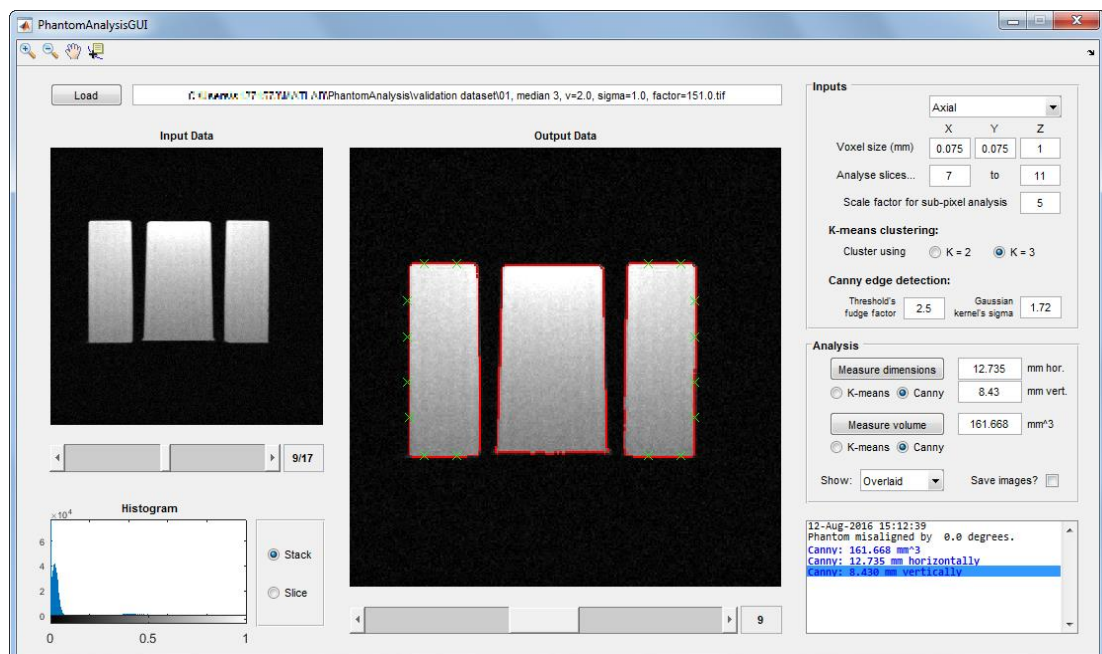


Figure 4.4 A graphical user interface for the analysis of phantom data. The tool shows the input slices and their histogram on the left, the analysed images on the middle and the panels used for controlling and carrying out the analysis on the right. The red lines over the analysed image represent the detected boundary of the phantom and the small green cross marks indicate opposing points on the boundary where the dimensions of the phantom are measured.

The tool with all associated code, instructions for its use and the validation dataset described below are freely available for download from GitHub (<https://github.com/Edinburgh-Imaging/PreclinicalMRIPhantomAnalysis>).

4.3.4 Validation dataset

To examine the accuracy of segmentation, 60 phantom scans were created by simulating realistic variation due to protocol differences and system errors at different sites. Actual phantom scans in all imaging planes were initially acquired using the QA protocol described earlier; six scans were taken with the volume coil alone and six scans with the volume-surface coil pair. Using custom-made MATLAB functions, each of these 12 scans was randomly scaled within $\pm 5\%$ of the original FOV ($19.2 \times 19.2 \text{ mm}^2$), translated in-plane within $\pm 10\%$ of the FOV in either direction, and its SNR was altered by either adding Rician noise or smoothing with median filters of size 1-3 pixels. Noise in magnetic resonance images is governed by the Rice distribution (Gudbjartsson & Patz, 1995), the probability density function of which is given by

$$p(x) = \frac{x}{\sigma^2} e^{-\frac{x^2 + \nu^2}{2\sigma^2}} I_0\left(\frac{x\nu}{\sigma^2}\right) \quad (4.6)$$

where x is the measured pixel intensity, ν is called the non-centrality parameter (corresponds to the intensity in the absence of noise), σ the scale parameter (corresponds to the SD of the distribution) and I_0 the modified Bessel function of the first kind with zero order. $p(x)$ depends on the values of ν and σ ; the ratio ν/σ is a measure of SNR. Figure 4.5 shows an actual MRI noise distribution measured in the unmodified axial scan acquired using the surface coil, and $p(x)$ for $\sigma = 1$ and various values of ν , generated using a random noise generator developed in MATLAB by Bar-Guy (2005). The shape of the Rice distribution with $\nu = 2$ and $\sigma = 1$ matched better the actual noise distribution and was chosen for simulations.

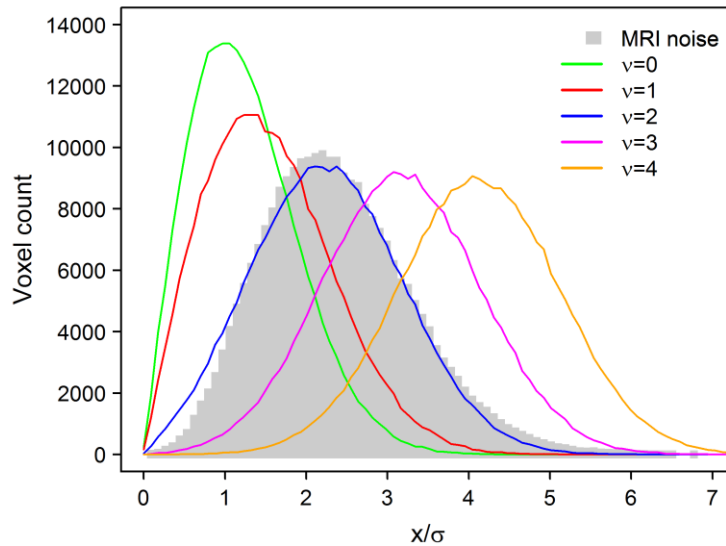


Figure 4.5 The Rician noise in MRI images. The background noise in the phantom images is characterised by a distribution similar to the Rice probability density function with $\nu = 2$ and $\sigma = 1$; this σ was used for generating all the distributions shown in colour. The histograms for all distributions are based on data with 252416 voxels and were plotted using 100 bins.

A noise image was generated for each slice of the phantom scan with the same size using Bar-Guy's generator. All noise images for each scan were then multiplied with a randomly estimated factor between 0 and 200 to simulate noise of differing magnitude (200 was chosen for 16-bit phantom images with a maximum of 65536 levels of intensity). The mean intensity of each noise image was subtracted from the corresponding phantom slice and the noise image was added to the new phantom slice, in order to preserve the overall mean intensity.

Five scans were created using each original scan, for a total of 60 scans. Simulated scans were analysed using the developed GUI to measure the dimensions of the phantom in all 60 scans and the volume of the phantom's cylindrical compartment in the 10 scans in the axial plane. The default scaling factor of 5 was used for subpixel measurements of dimensions.

4.3.5 Statistical analysis

The simulated scans were analysed manually twice by myself and once by Ross Lennen (University of Edinburgh) to examine intra-rater reliability and inter-rater reproducibility; Ross Lennen was blinded to the results of my analysis. Each rater had at least 3 years of image analysis experience at the time of analysis. The second analysis by myself was performed one month after the first analysis to avoid any learning effects. ImageJ was used with the straight line tool for measuring dimensions (0.01 mm precision) and the freehand selections tool for measuring the volume.

The accuracy of the developed automated tool for estimating both QA criteria was assessed by comparison against the measurements of my first manual analysis, which were considered as the ground truth. The agreement between intra- and inter-rater measurements, or ground truth measurements and each of the three segmentation methods was examined using Bland-Altman analysis (Bland & Altman, 1986). This method requires plotting of the difference between corresponding measurements against their mean; the compared methods agree if the mean of the differences (which represents accuracy) lies close to the ideal mean, zero, as judged based on its 95% CI. Bland-Altman analysis is superior to simple correlation analysis for comparing the performance of a new method against a gold standard, since correlation measures the strength of their relation and not the degree of absolute agreement as the former method does (Bland & Altman, 1986).

To evaluate the spatial overlap between manually defined volumes or manually versus automatically segmented volumes, the Dice coefficient (DC) was estimated (Dice, 1945). This metric is very popular for the validation of image segmentation algorithms, as it takes into account the similarity between the compared regions in terms of both shape and location in the image, instead of just their size. DC is given by

$$DC = \frac{2|V_a \cap V_b|}{|V_a| + |V_b|} \quad (4.7)$$

where V_a and V_b are the two compared volumes; the intersection $|V_a \cap V_b|$ is the number of common voxels and the sum in the denominator is the total number of voxels. A value of 0 indicates no overlap and a value of 1 indicates perfect similarity between volumes; a good correlation between volumes is often considered to be one with $DC > 0.7$. Due to the small number of images used for volume estimation (10), DC values were summarised using boxplots showing median values and ranges.

Custom-written code in MATLAB was used to perform comparisons of binary masks for estimating DC. The time required for estimating both QA criteria in each of the 10 simulated axial scans was measured using a MATLAB function running the automated routines over all images and a stopwatch for the manual analysis; a standard Intel® Core 2 Quad Q9650 processor (Intel Corp., Santa Clara, California, USA) running at 3 GHz with 4 GB of random-access memory on a Windows® 7 Enterprise machine (Microsoft Corp., Redmond, Washington, USA) was used.

In SPSS® (22.0, IBM Corp., Armonk, New York, USA), the difference in analysis time was examined using repeated measures analysis of variance (ANOVA) with a Greenhouse–Geisser correction to account for the violated sphericity (Mauchly's test; Table C.2 in Appendix C) and Bonferroni's post hoc tests to correct for multiple comparisons (repeated-measure factor is the analysis method). Distributions of measured times were normally distributed, as determined by the Shapiro–Wilk test for normality (Table C.1). A p -value less than 0.05 was considered statistically significant.

4.4 Results

4.4.1 Phantom's T_2 relaxation time

Figure 4.6 shows the mean intensity within a circle of diameter 3 mm at the centre of the phantom, plotted against TE in images acquired using the volume coil and the pulse sequence in Table 4.1. By fitting equation 4.1 to the points, the T_2 relaxation time of the phantom's liquid solution was estimated to be 53.56 ms, with an excellent goodness of fit ($\bar{R}^2 = 0.994$).

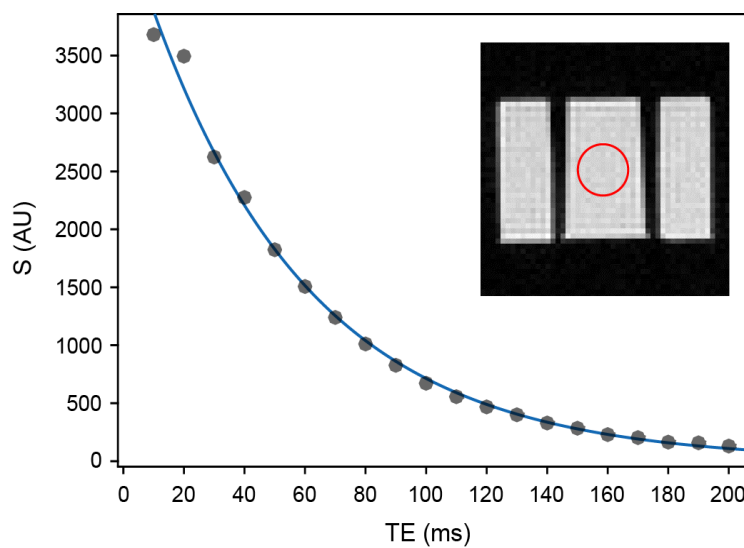


Figure 4.6 Measurement of phantom's T_2 relaxation time.

The mean intensity was measured within a circle of diameter 3 mm in all images acquired with varying TEs (red circle in insert). The T_2 exponential decay curve (blue line) was fitted on intensity versus TE points to measure the T_2 relaxation time: $S = 4685e^{-TE/53.56} - 10.25$. AU indicates arbitrary units; S, signal intensity; TE, echo time.

4.4.2 Intra- and inter-rater agreement

As shown by Bland-Altman plots in Figure 4.7, the intra-rater reliability in the manual measurement of the phantom's dimensions and the volume of its cylindrical compartment in the 60 simulated scans is excellent, with the mean difference in all plots lying very close to zero.

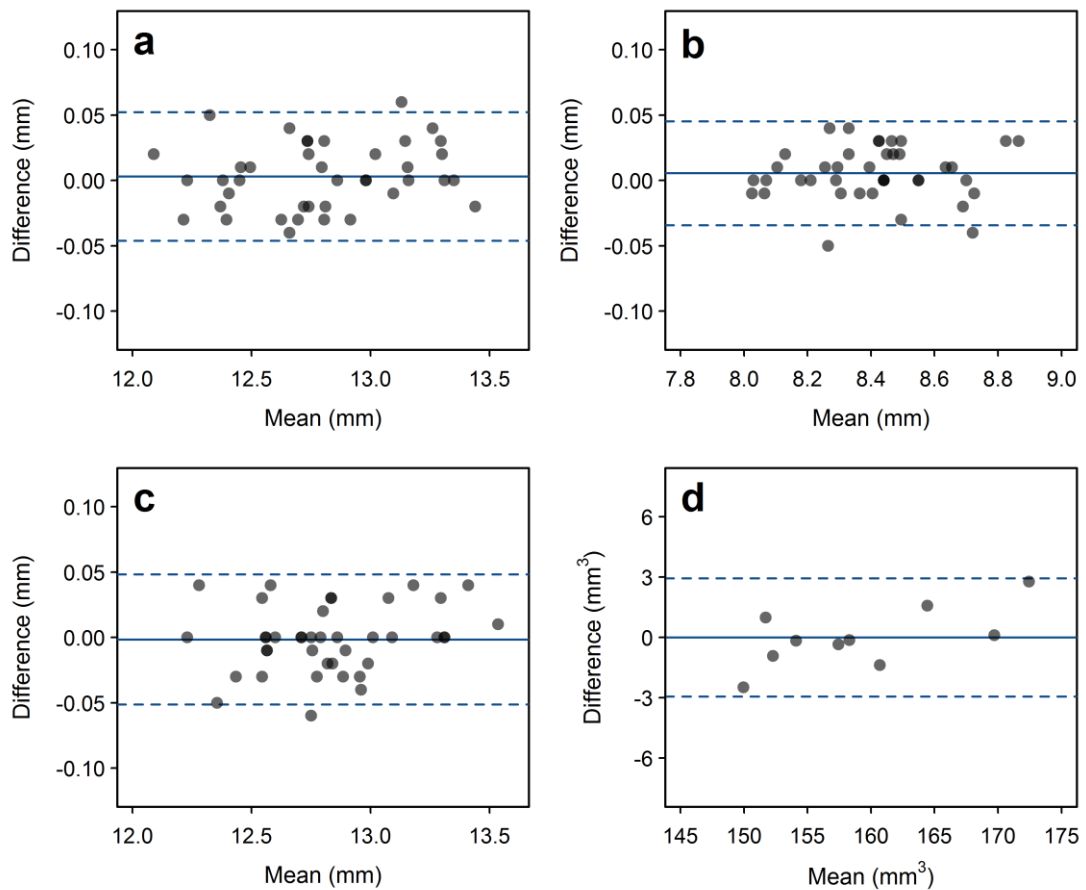


Figure 4.7 Bland-Altman plots assessing intra-rater agreement.

The plots show the measurements in (a) x direction, (b) y direction and (c) z direction, and (d) the volume measurements. For estimating the differences, the measurements of the first analysis were subtracted from the corresponding measurements of the second analysis. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

Inter-rater reproducibility is also excellent (Figure 4.8). The mean differences in measurements of the phantom's dimensions and volume are small and similar to those in corresponding comparisons in the intra-rater reliability assessment. However, there is a small increase in the dispersion of differences.

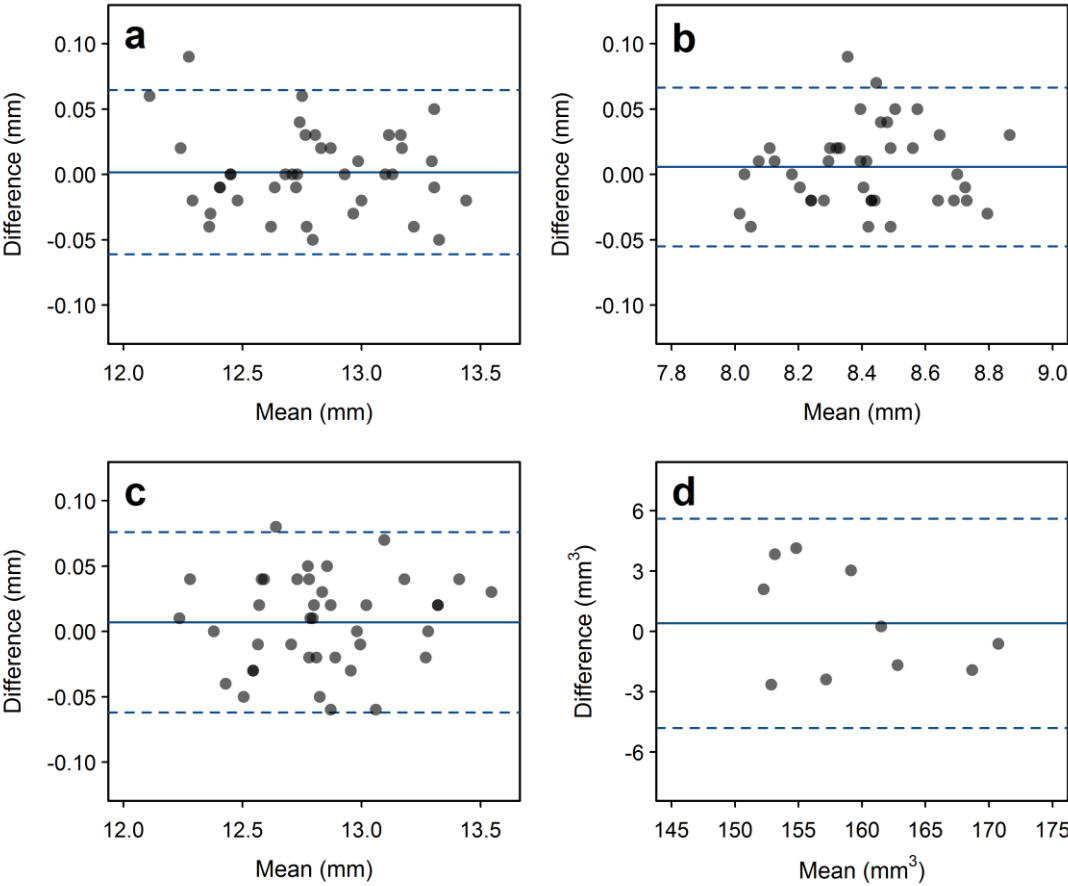


Figure 4.8 Bland-Altman plots assessing inter-rater agreement. The plots show the measurements in (a) x direction, (b) y direction and (c) z direction, and (d) the volume measurements. For estimating the differences, the measurements of the first analysis by the first rater were subtracted from the corresponding measurements by the second rater. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

The volume of the phantom's cylindrical compartment has a near perfect similarity between analyses, as assessed by DC (intra-rater: median = 0.993, interquartile range (IQR) = 0.991–0.993; inter-rater: median = 0.984, IQR = 0.981–0.988). The first set of measurements by myself are, therefore, reliable and reproducible and are used for assessing the performance of the automated methods.

4.4.3 Performance of automated analysis tool

4.4.3.1 Selection of Canny edge detection parameters

Before the three developed methods for automated analysis of phantom data are compared, the performance of Canny edge detection was optimised by selecting appropriate factors for automatic selection of the two thresholds, f_T , and the SD of the Gaussian filter, f_{SD} (Section 4.3.2.3). Figures 4.9 and 4.10 show Bland-Altman plots assessing the agreement between this method and manual analysis for different values of f_T . It is evident that the default thresholds estimated by the MATLAB's **edge** function (for $f_T = 1$) sometimes lead to detection of unwanted edges outside the phantom and overestimation of its dimensions (Figure 4.9 a-b). The accuracy of the method is almost identical for f_T values between 1.5 and 3 (Figure 4.9 d-o). The iterated procedure for estimating the volume of the phantom's cylindrical compartment was not influenced noticeably by the choice of thresholds, and differences with corresponding manual measurements had a mean close to zero (Figure 4.10). $f_T = 2.5$ was chosen for subsequent analyses, due to the high overall accuracy of all measurements obtained using this value.

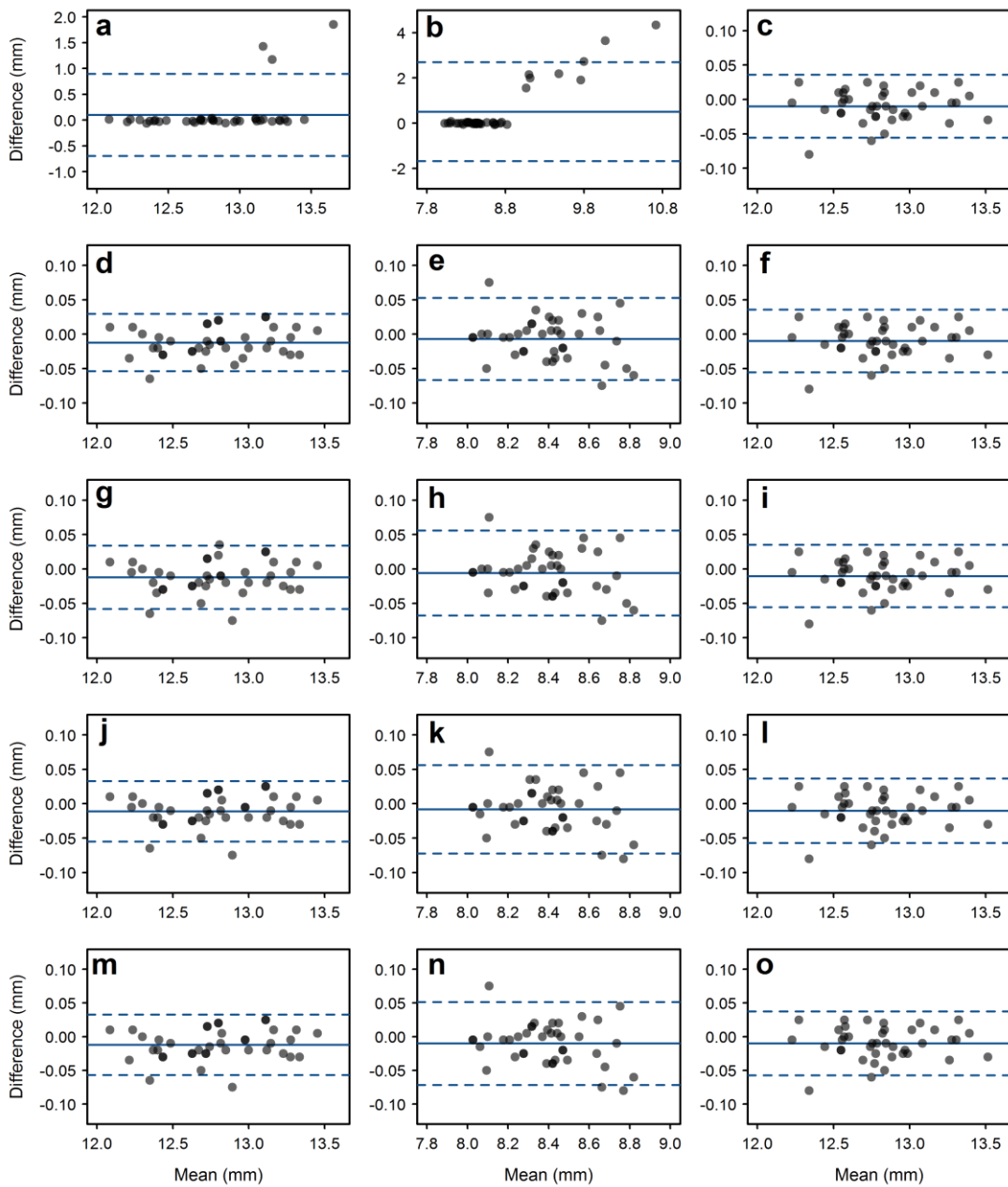


Figure 4.9 Bland-Altman plots assessing the performance of Canny edge detection for measuring the phantom's dimensions for various thresholds. Five values of factor f_T were tested: (a-c) 1, (d-f) 1.5, (g-i) 2, (j-l) 2.5, and (m-o) 3. Plots in the left column (a, d, g, j, m) show measurements in x direction, in the central column (b, e, h, k, n) in y direction, and in the right column (c, f, i, l, o) in z direction. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

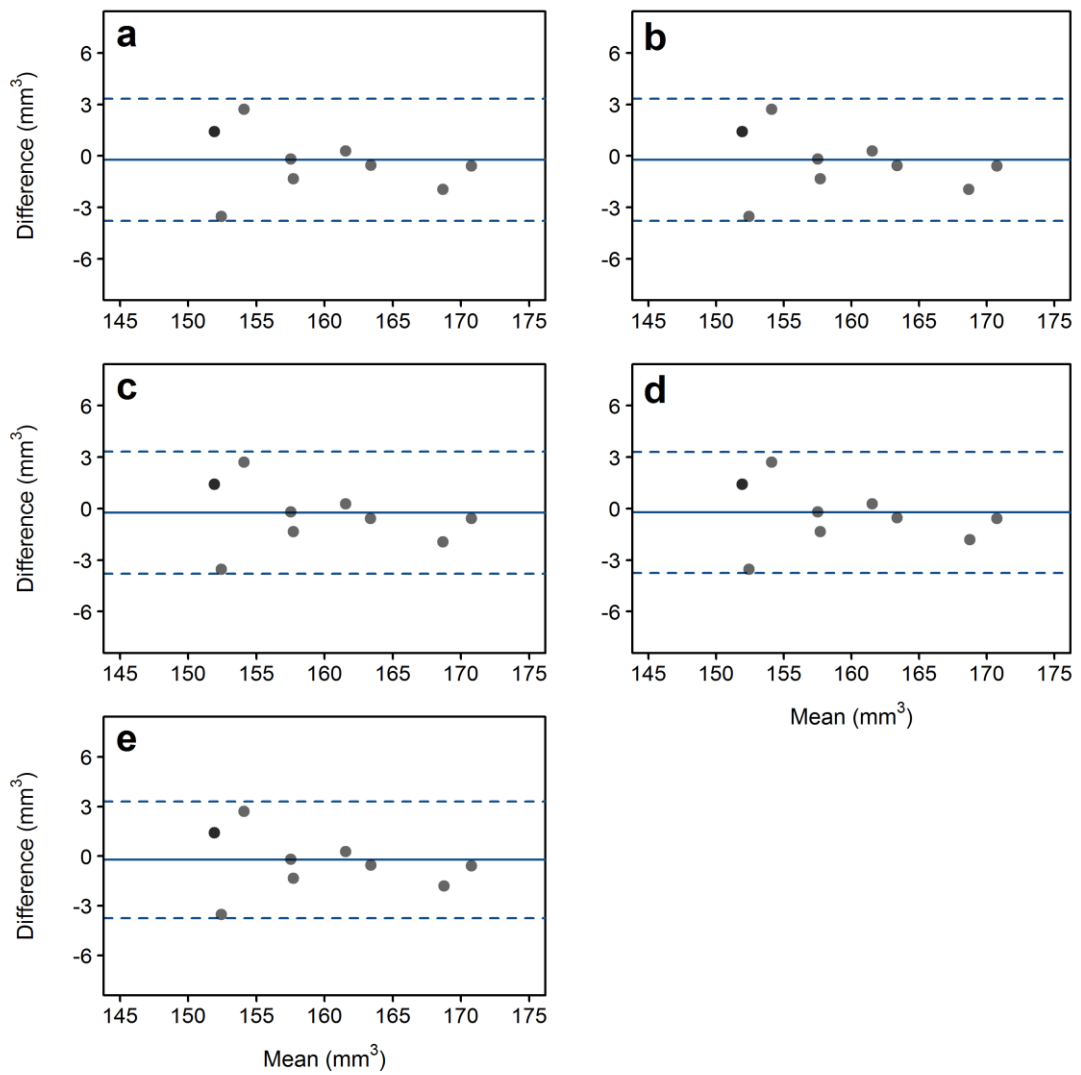


Figure 4.10 Bland-Altman plots assessing the performance of Canny edge detection for measuring the volume of the phantom’s cylindrical compartment for various thresholds. Five values of factor f_T were tested: (a) 1, (b) 1.5, (c) 2, (d) 2.5, and (e) 3. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

Setting f_T to 2.5, the factor for automatic adjustment of the SD of the Gaussian filter (f_{SD} in equation 4.3) was varied between 10 and 50; results are shown in figures 4.11 and 4.12. In contrast with f_T , f_{SD} affects mostly the measurement of volume, with small values (large Gaussian filters) causing selection of fewer voxels than manual analysis and large values (small Gaussian filters) causing selection of more voxels (Figure 4.12 a, e). $f_{SD} = 30$ was the optimal and was chosen for this method.

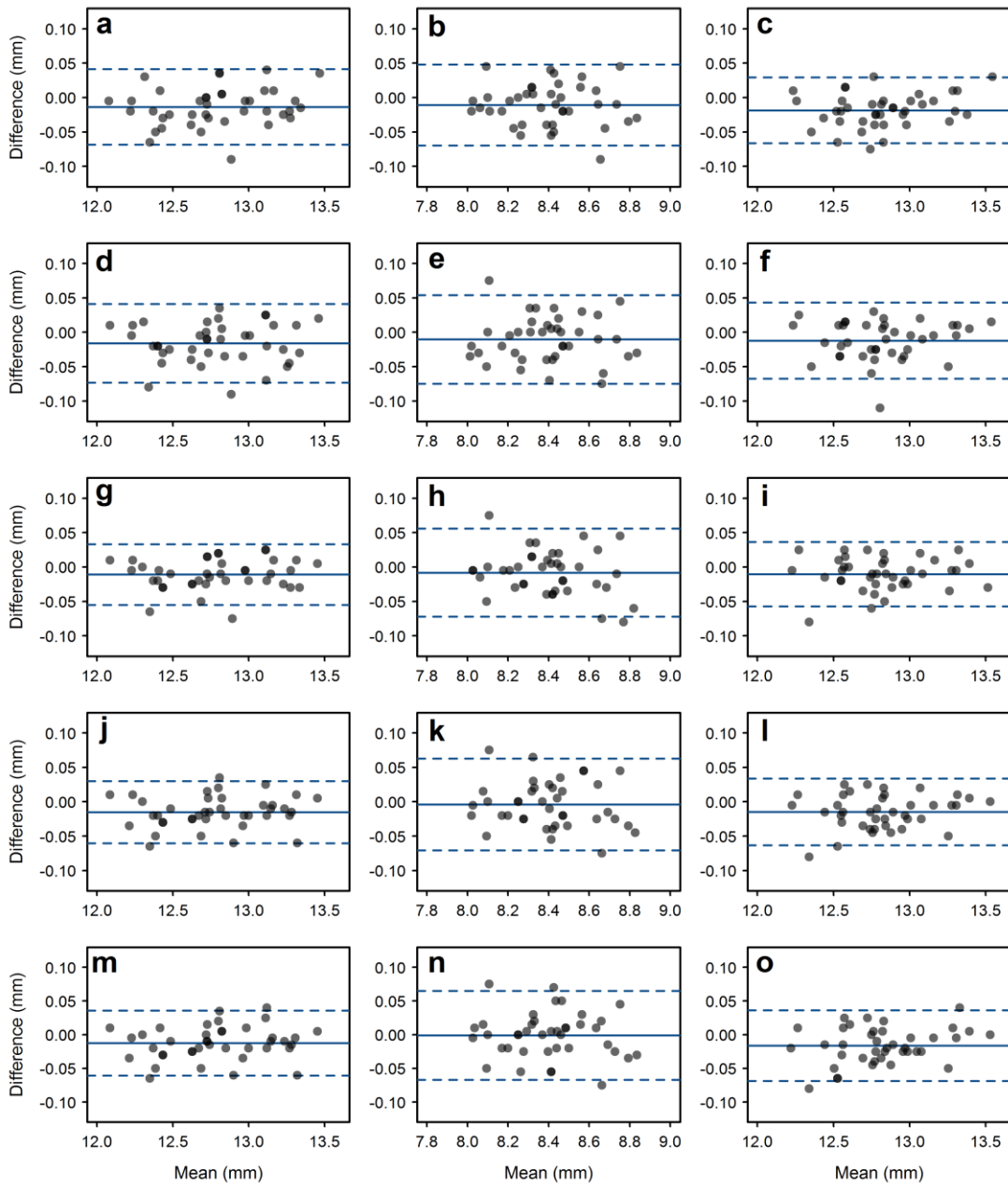


Figure 4.11 Bland-Altman plots assessing the performance of Canny edge detection for measuring the phantom's dimensions for various Gaussian filters. Five values of factor f_{SD} were tested: (a-c) 10, (d-f) 20, (g-i) 30, (j-l) 40, and (m-o) 50. Plots in the left column (a, d, g, j, m) show measurements in x direction, in the central column (b, e, h, k, n) in y direction, and in the right column (c, f, i, l, o) in z direction. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

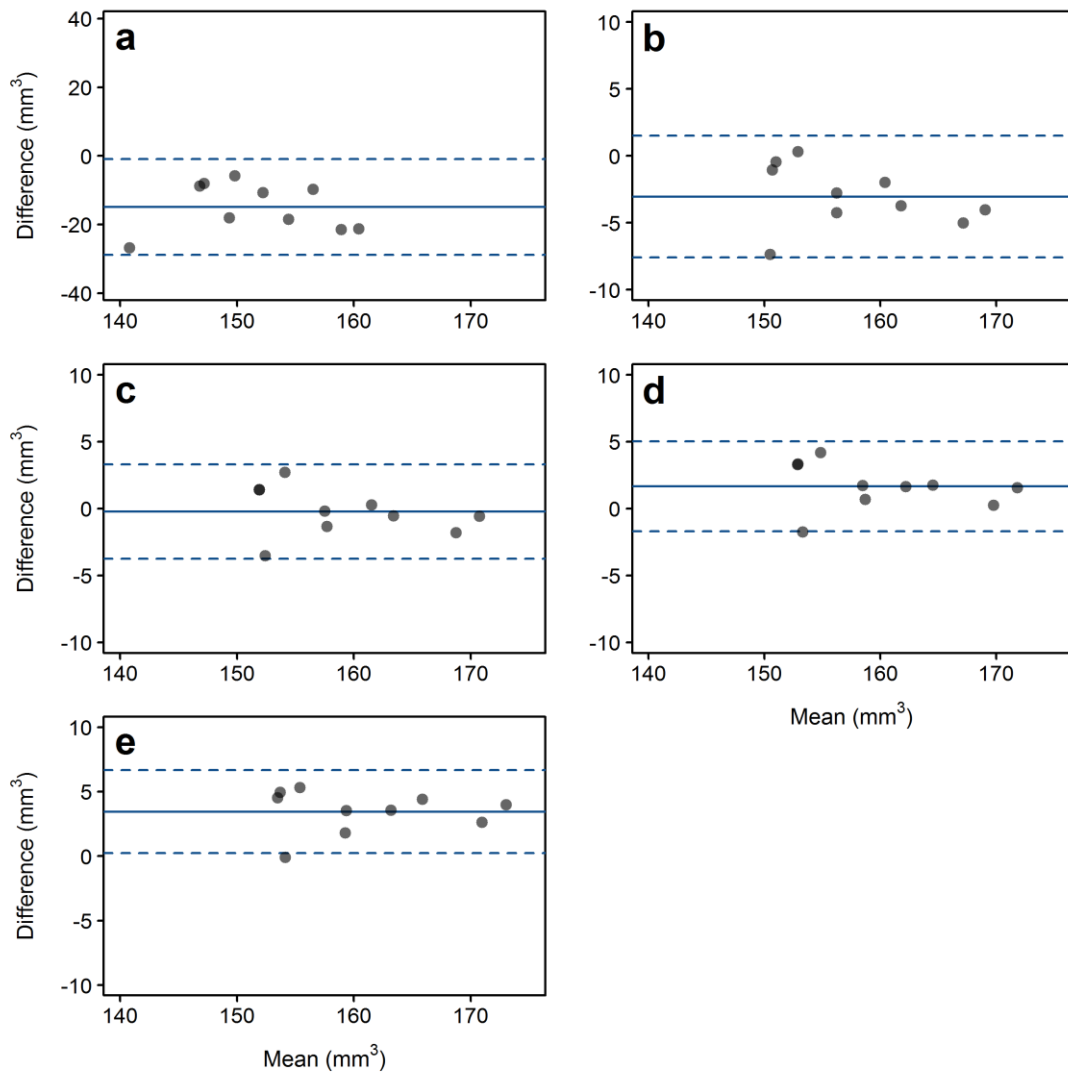


Figure 4.12 Bland-Altman plots assessing the performance of Canny edge detection for measuring the volume of the phantom’s cylindrical compartment for various Gaussian filters. Five values of factor f_{SD} were tested: (a) 10, (b) 20, (c) 30, (d) 40, and (e) 50. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

In agreement with previous observations, DC for the automatically segmented volumes remained stable for different values of f_T , but varied significantly for different values of f_{SD} (Figure 4.13). The highest median DC is 0.982 (IQR = 0.975–0.983) for the chosen values $f_T = 2.5$ and $f_{SD} = 30$, denoting almost perfect similarity with manually defined volumes.

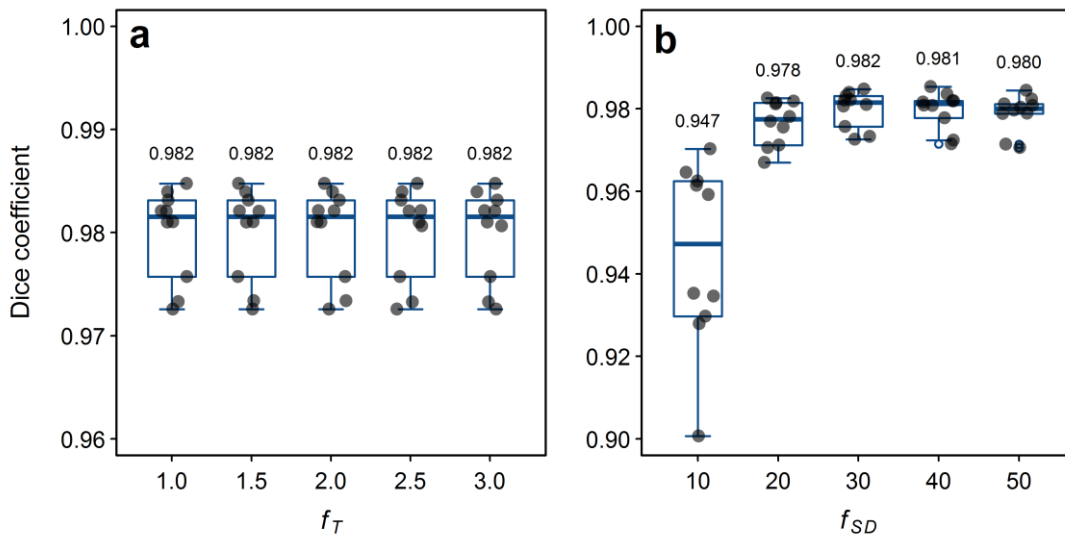


Figure 4.13 Dice coefficient for various Canny thresholds and Gaussian filters. Five values for each of factors (a) f_T and (b) f_{SD} were tested, as indicated, while $f_{SD} = 30$ and $f_T = 2.5$ respectively. The thick blue line of each boxplot indicates the median Dice coefficient (value given above).

4.4.3.2 Comparison of segmentation methods

Figures 4.14 and 4.15 compare the performance of all segmentation methods developed for the analysis of phantom data. Automated analysis by simple thresholding is the worst performing method, with the variability in measurements for both QA criteria being large, despite a small mean difference between automated and manual measurements of phantom's dimensions (Figure 4.14 a-c; Figure 4.15 a). k -means with $k = 2$ provides accurate estimates of phantom's dimensions, but the volume is underestimated by -9.4 mm^3 on average (Figure 4.14 d-f; Figure 4.15 b). In contrast, k -means with $k = 3$ usually overestimates the size of the phantom in all directions, but provides fairly accurate estimates of volume (0.6 mm^3 mean volume difference; Figure 4.14 g-i; Figure 4.15 c). Overall, Canny's method is the best, with measurements for both QA criteria being accurate and the dispersion of their differences with manual measurements being small (-0.2 mm^3 mean volume difference; Figure 4.14 j-l; Figure 4.15 d).

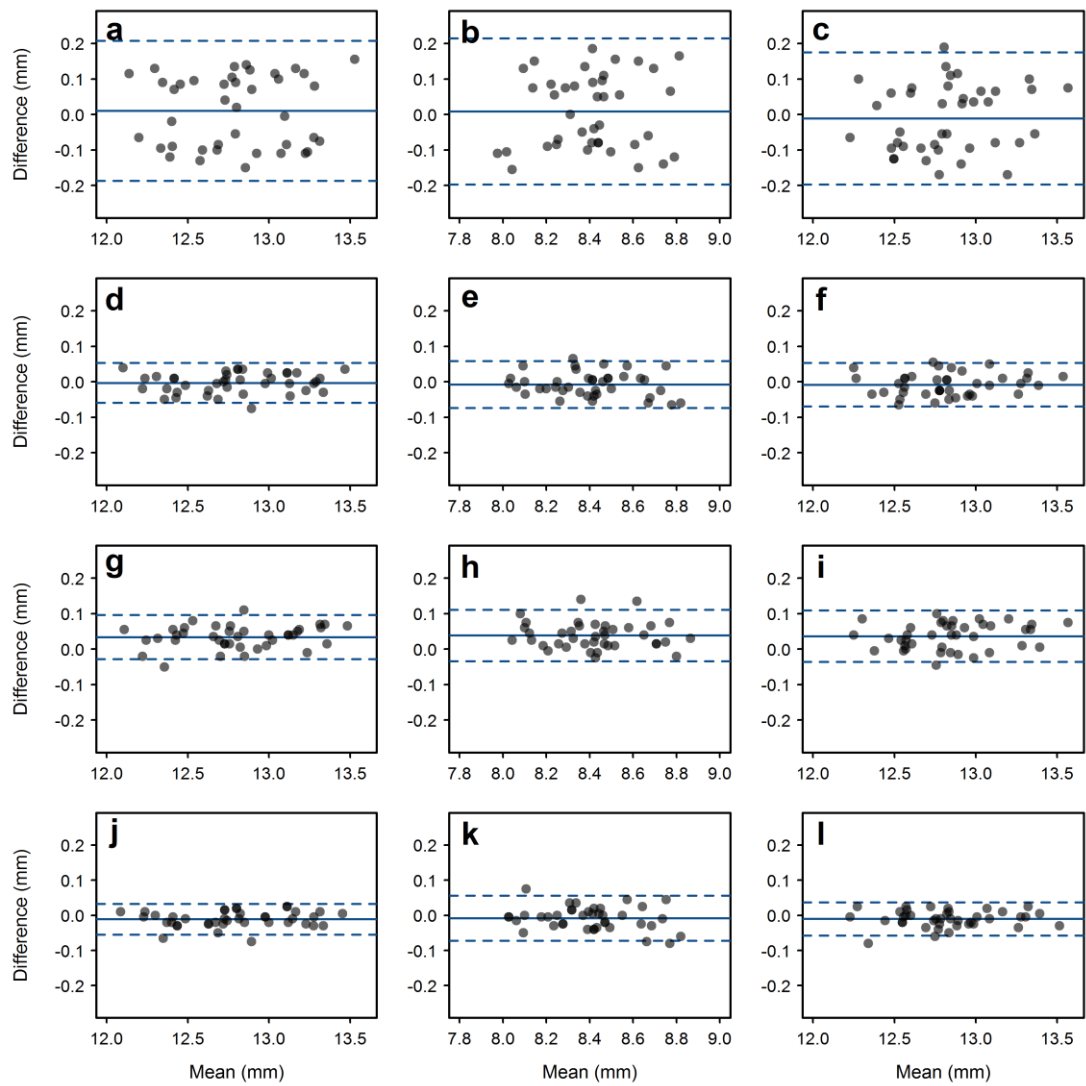


Figure 4.14 Bland-Altman plots comparing different methods for measuring the phantom's dimensions. Four methods were compared: (a-c) simple thresholding, (d-f) k -means clustering with $k = 2$, (g-i) k -means clustering with $k = 3$, and (j-l) Canny edge detection. Plots in the left column (a, d, g, j) show measurements in x direction, in the central column (b, e, h, k) in y direction, and in the right column (c, f, i, l) in z direction. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

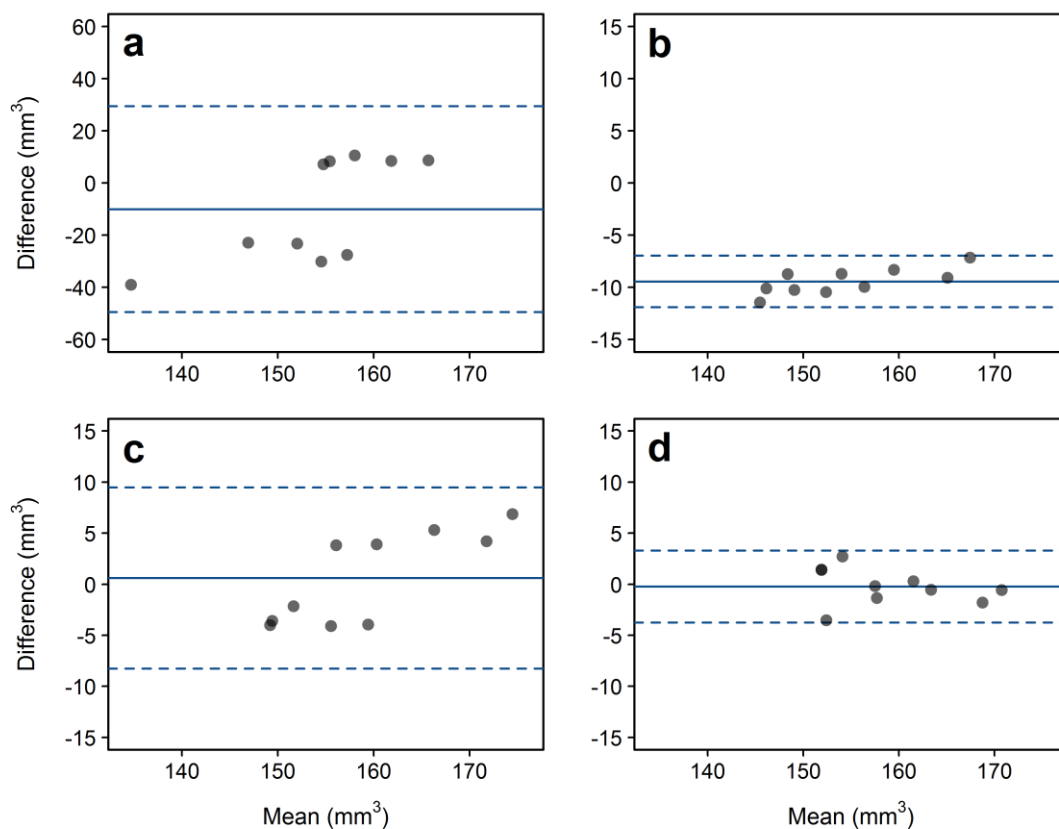


Figure 4.15 Bland-Altman plots comparing different methods for measuring the volume of the phantom's cylindrical compartment. Four methods were compared: (a) simple thresholding, (b) k -means clustering with $k = 2$, (c) k -means clustering with $k = 3$, and (d) Canny edge detection. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

Repeated measures ANOVA revealed that the analysis time differed significantly between all methods ($F(1.711, 15.399) = 3092.230, p < 0.001$), as well as in all pairwise comparisons performed with post hoc tests using the Bonferroni correction ($p < 0.001$; Figure 4.16 b). All automated methods are faster than manual analysis. Expectedly, despite being the most computationally efficient method, simple thresholding produces more variable DC estimates than the rest of the methods (Figure 4.16 a). The median DC is very high for k -means with $k = 3$ and Canny edge detection, but Canny's method is significantly faster; it requires just 5.3 seconds on average for each axial scan to measure both QA criteria (56.4 times faster than manual analysis), in contrast with 37.6 seconds spent by k -means with $k = 3$. Overall, the performance of the

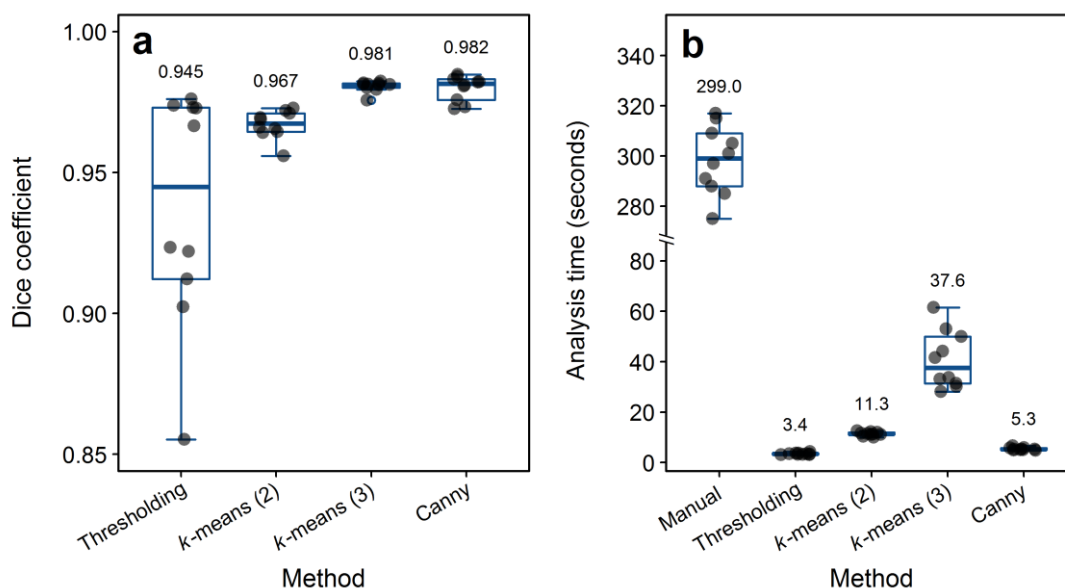


Figure 4.16 Dice coefficient and analysis time for different phantom data analysis methods.

The plots show (a) the Dice coefficient for methods tested against manual analysis and (b) the time required for full analysis of each axial scan by all methods. The thick blue line of each boxplot indicates the median estimate (value given above). Analysis time is significantly different between all methods in analysis of variance pairwise comparisons ($p < 0.001$).

automated Canny method is similar to that of the second manual rater, as evaluated against the ground truth estimates by the first rater, but has the additional benefit of being perfectly reproducible.

4.5 Discussion

This chapter describes the procedure followed for the construction of several identical sMRI phantoms and the development of an automated image analysis tool for assessing geometric accuracy in preclinical scanners. The phantom was scanned using both volume and surface RF coils with a simple QA protocol, in order to validate the accuracy of the automated tool. The produced images have high resolution and great quality, suggesting that the phantom materials and QA protocol are suitable for the aforementioned purpose.

LEGO bricks were used successfully for the development of clinical MRI phantoms previously (Caramanos et al., 2010). In the context of multicentre studies in rat models of stroke, the chosen bricks are ideal for a number of reasons: (1) the sizes of the whole phantom and its cylindrical compartment are similar to those of the part of the rat brain often imaged in stroke and a large subcortical infarct respectively; (2) their cuboid shape allows reproducible positioning in the scanners and reliable measurement of dimensions, in contrast with spherical or cylindrical phantoms; (3) they are manufactured by injection moulding with superior precision (20 μm) than most alternative approaches, such as 3D printing (Quercioli et al., 1998); (4) they are made of plastics with great dimensional stability and MRI compatibility (Osswald & Menges, 2010; Shackelford et al., 2015; Wapler et al., 2014); and (5) they are very cheap and widely available. It should be noted, however, that the small size of the phantom precludes monitoring of geometric accuracy in studies assessing morphological outcomes larger than those in models of neurological diseases in rodents. Furthermore, while the phantom has similarities with a rat brain in terms of overall size and T_2 relaxation, the volume of its cylindrical compartment does not mimic the shape of an infarct or the contrast in intensities between injured and healthy tissue. Even so, the phantom constitutes a useful method for examining the influence of scanning parameters and identifying sources of geometric variability in small animal MRI. Methods for creating small phantoms with exact lesion characteristics are currently not available.

The need for routine QA assurance in preclinical MRI has only recently been recognised; others proposed phantom designs to facilitate cancer studies (Lee, Fullerton, Baiu, Lescrenier, & Goins, 2011) or perform 3D geometric distortion correction (O'Callaghan et al., 2014; Teh et al., 2016; Yoshimaru, Totenhagen, Alexander, & Trouard, 2014). These phantoms could not be used in the studies described in this thesis, as their size hinders use with various types of imaging coils at different centres, most are constructed with lower precision or require more complex procedures for construction and post-processing of acquired data, and are

not as cost-effective.

It is widely accepted that manual analysis is time-consuming, laborious and depends on many factors, ranging from the experience of the rater to the display and room lighting settings used; these may compromise the intra-rater reliability and inter-rater reproducibility. QA of clinical MRI scanners with standardised phantoms has been the norm for decades, but only recently the assessment of data has shifted from manual to automated (Davids et al., 2014; Panych et al., 2015). Following this path, an automated tool was also developed for the analysis of phantom data acquired in this thesis to improve the reliability and reproducibility of QA and enable valid comparisons between different scanners. A few common methods were implemented and compared, and Canny edge detection was found to be superior; this is used in the multicentre study presented in Chapter 6. The superiority of this method can be attributed to the following two reasons: Firstly, the presence of global intensity inhomogeneities (bias field) governs the detection of edges in simple thresholding and k -means clustering, which rely on assumptions of intensity homogeneity. In the images acquired using the surface RF coil, segmentation using these two methods could cause the inclusion of more pixels in the part of the image that lies closer to the coil, as compared to the rest of the image. Secondly, the distribution of the signal coming from the phantom's liquid solution may lie very close to that of background noise, or even overlap with it. This problem can become apparent in the low SNR images taken using the volume coil, which can affect the selection of an appropriate threshold or cause clustering errors. That being said, Canny's method depends on more input parameters than the other two methods (two intensity thresholds and the SD of the Gaussian filter); methods for estimating these automatically have been devised, and were found to perform very well in all scenarios. The robustness of the developed analysis pipeline is showcased by the often identical results for various values for these parameters above or below the optimal.

The FSE pulse sequence was used for acquiring images with volume and surface RF coils; as indicated by the systematic review presented in the previous chapter, these

are widely used for MRI in experimental stroke for the assessment of infarct size (Milidonis, Marshall, Macleod, & Sena, 2015). Despite that sequence parameters were chosen based on those used in T2WI in rat models of stroke (Table 3.1), some were adjusted to produce images with good overall quality in a short period of time for the purpose of QA, and may not be appropriate for use in experimental in vivo imaging. For example, a short TR was used to increase the speed of acquisition and a small FOV was used to improve the resolution and therefore the precision of measurements; this TR would not be sufficient for nuclei in the brain to fully relax for the acquisition of the next line of data, and a small FOV could introduce significant aliasing artefacts in, at least, rat brain imaging.

Fast, accurate and reliable methods for the assessment of geometric accuracy in preclinical MRI scanners in the context of multicentre studies in rat models of stroke were developed. These are used in Chapter 6 to carry out a multicentre assessment of geometric accuracy in scanners routinely used in stroke.

Chapter 5

Development of a Semi-Automated Infarct Segmentation Tool

Reliable assessment of the effect of MRI variability on infarct volume quantification requires an accurate and reproducible measurement method. In this chapter, a simple threshold-based semi-automated method for the analysis of rodent brain 2D T2WI data and the measurement of infarct volume is proposed. Its accuracy and reproducibility is assessed by comparison against measurements of gold standard manual analysis and a common volumetric technique, stereological point counting.

Parts of this chapter have been published or presented as:

Milidonis, X., McCabe, C., Macrae, M. I., Sena, E. S., Macleod, M. R., & Marshall, I. (2016). Segmentation of infarct on structural MRI data in rodent models of stroke using a semi-automated tool. *European Stroke Journal*, 1(1), 3-612. doi:10.1177/2396987316642909. Proceedings of the 2nd European Stroke Organisation Conference (p. 404). Barcelona, Spain.

Milidonis, X., McCabe, C., Macrae, M. I., Sena, E. S., Macleod, M. R., & Marshall, I. (2016). Comparison of infarct quantification methods in experimental focal cerebral ischaemia. *Cerebrovascular Diseases*, 41(1), 1-2. doi:10.1159/000446380. Proceedings of the 25th European Stroke Conference (p. 207). Venice, Italy. [awarded the European Stroke Research Foundation Travel Grant Award].

Milidonis, X., McCabe, C., Macrae, M. I., Sena, E. S., Macleod, M. R., & Marshall, I. (2016). Segmentation of infarct on structural MRI data in rodent models of stroke using a semi-automated tool. *Presented at the 8th Annual Scientific Meeting of the Scottish Imaging Network: A Platform for Scientific Excellence (SINAPSE)*. Stirling, UK.

5.1 Introduction

Rodents are the most widely used animals in experimental stroke and infarct volume is the primary outcome measured (Howells et al., 2010; Macrae, 2011). Infarct volume exhibits high inter-animal heterogeneity in the majority of rodent MCAO models (Liu et al., 2009), which, as shown earlier, may be magnified by an excess of sources of errors related to the outcome assessment method and particularly MRI. Chapter 4 suggested an MRI QA protocol to reduce the variability and improve the accuracy of measured morphological outcomes, but still the most significant contribution to heterogeneity often comes from post-processing methods used for the analysis of acquired images. The consistency in measurements within groups of animals can be improved by increasing the number of animals used or excluding outliers (Dirnagl, 2006), but using a more accurate and reproducible method of measurement may have a similar effect. A decreased measurement variability can help increase the statistical power of the study, improve the generalisability of findings and reduce the waste of animals in studies where large cohorts are needed, such as multicentre studies. In addition, automating the method as much as possible can further reduce the burden to the assessors and the overall duration of the study.

As revealed by the systematic review presented in Chapter 3, manual planimetric analysis is often employed for estimating infarct size on 2D T2WI data. Semi-automated methods including basic thresholding are also used, but these are currently applied on quantitative T₂ maps or intensity-homogeneous images acquired using volume RF coils. Reasons behind this may vary, as described below. Firstly, qualitative MRI data often suffer from severe bias field effects that preclude the use of automated software that rely on the assumption of global intensity homogeneity, particularly when surface RF coils are used. Secondly, infarcts usually have complex shapes and indistinct boundaries, resulting in ROIs with inhomogeneous contrast that are difficult to define automatically. Thirdly, infarcts in the subacute and chronic phases of stroke are enlarged due to the presence of vasogenic oedema, which results in a shift of the midline between the brain hemispheres and other abnormalities in the

brain, such as cerebral microbleeds and enlargement of ventricles in some MCAO models that may confound unsupervised analysis techniques. Therefore, manual planimetric analysis of 2D T2WI data is still considered as the gold standard approach; it requires the rater to manually outline the boundary of the infarct area in each slice, making the procedure extremely labour intensive, time-consuming and subjective, as it depends on the experience and personal judgement of the rater, amongst other.

Expectedly, there is an absence of dedicated tools for automated or semi-automated analysis of such data in rodent models of stroke. As a result, data acquired in preclinical centres are often analysed manually by corresponding researchers. Numerous studies describe dedicated infarct segmentation tools for use on clinical brain MRI data (Bauer, Gratz, Gralla, Reyes, & Wiest, 2014; Filippi et al., 2014), but a few have created basic computerised methods for estimating infarct size from histologically stained rodent brain slices (Regan, Detwiler, Huang, Lynch, & Regan, 2007; Swanson et al., 1990). Herein, an attempt is made to define an analysis pipeline for the analysis of qualitative T2WI data for the measurement of infarct size, aiming to improve the repeatability, reproducibility and speed of analysis. Most steps of this pipeline are performed by a dedicated MATLAB tool, allowing automated segmentation of the infarct and subsequent manual editing of the selected regions in each slice to further improve segmentation accuracy. The tool was validated against gold standard full manual analysis of a standard set of rat brain T2WI data acquired as part of an MCAO experiment. The tool was also compared with stereological point counting, a method previously used for the analysis of multi-slice data in clinical studies with reported high accuracy and reproducibility (Acer, Sahin, Usanmaz, Tatoğlu, & Irmak, 2008; Saunders et al., 1995; Thrippleton et al., 2014). Intra- and inter-rater agreement in the measurement of infarct volume was evaluated for all three methods.

5.2 Methods

5.2.1 Analysis pipeline

A clear definition of the infarct as observed in qualitative 2D T2WI data in standard MCAO models is not available, and manual analysis is performed on subjective terms across centres. Here, a pipeline for the semi-automated analysis of such data was devised based on the following four observations:

1. The infarct can be defined based on the intensity of healthy tissue in the corresponding contralateral hemisphere, as proposed by articles systematically evaluated in Chapter 3 (Jiang et al., 2006; Modo, Beech, Meade, Williams, & Price, 2009; Veldhuis et al., 2003).
2. The infarct extends across a single brain hemisphere (ipsilateral hemisphere) within the territory supplied by the corresponding MCA.
3. The infarct initiates at the occluded origin of the artery and extends outwards to encompass surrounding regions, therefore no healthy tissue is enclosed by infarcted tissue.
4. More than one infarcted regions may be visible in a single MRI slice through the brain, but all regions are connected across slices into a single blob (Bauer et al., 2014).

The semi-automated analysis pipeline is shown in Figure 5.1; most analysis steps were used empirically and were optimised for use on coronal scans (indicated in Figure 2.3), which is the imaging plane primarily used in rodent models of stroke according to the systematic review presented in Chapter 3. The steps within the blue box in Figure 5.1 are performed using a custom-made dedicated MATLAB tool.

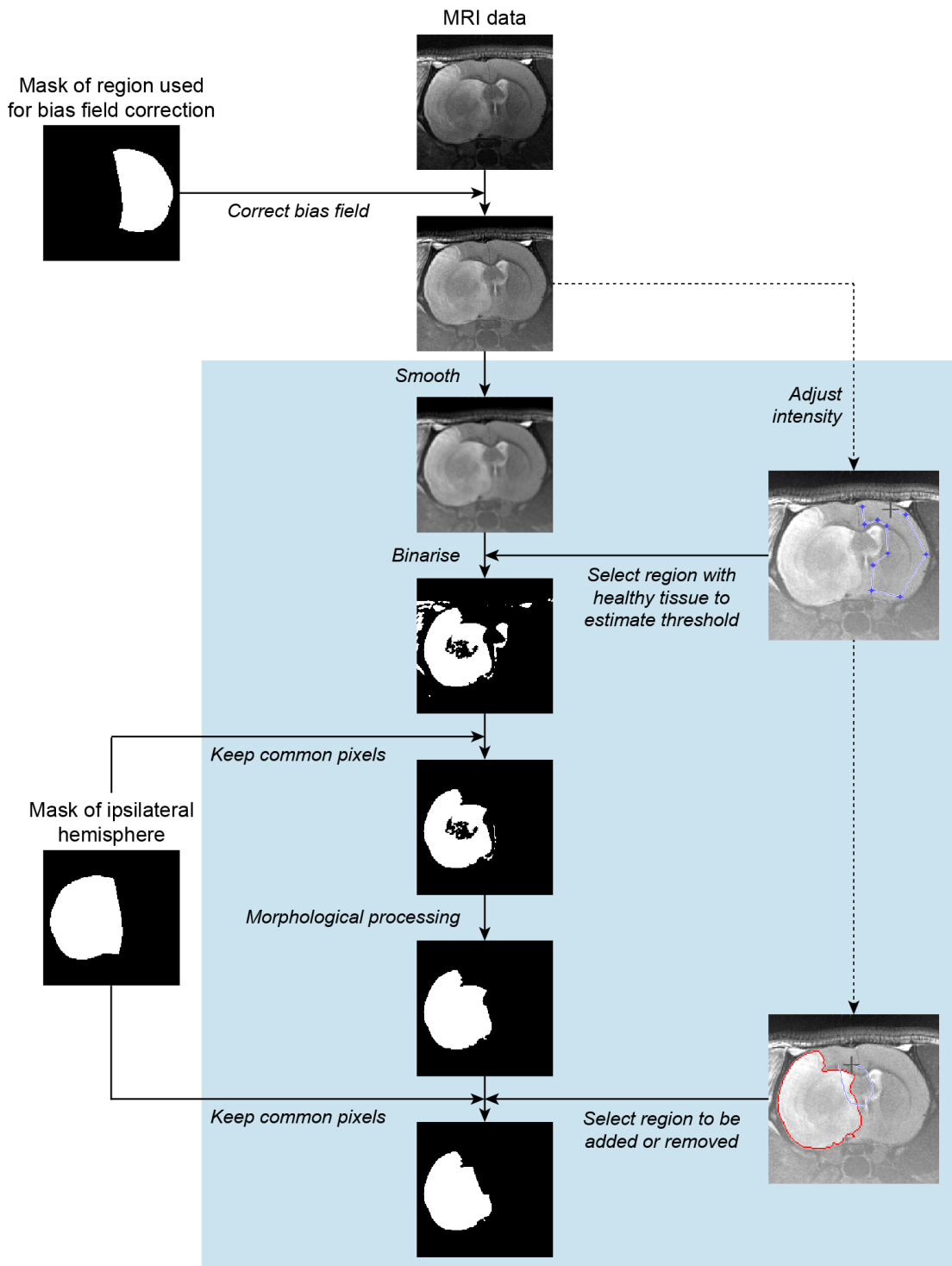


Figure 5.1 Semi-automated pipeline for the measurement of infarct size. This is a simplified diagram showing the major processing steps used to analyse the rodent brain data and measure infarct volume in the ipsilateral hemisphere; analysis is performed in coronal T2WI scans. The analysis steps within the blue box are incorporated in a custom-made MATLAB tool, with those on the right indicating manual input.

The first and most critical step is the correction of bias field in the images; this can be very challenging in stroke images as the intensity is expectedly not homogeneous across the brain (the lesion is brighter than healthy tissue in T2WI). Bias field correction was performed using the N4 algorithm (Tustison & Gee, 2010) in 3D Slicer (4.4.0; <http://www.slicer.org>). The corrected images are loaded in the developed MATLAB tool along with masks of the corresponding ipsilateral hemisphere. Their intensity is first normalised to the range 0 to 1 and the normalised images are passed in two different branches of the analysis pipeline. In one branch (right column in Figure 5.1) their intensity is automatically adjusted by scaling it to the intensity range in the middle slice; these images are used in steps requiring manual input. In the other branch (middle column in Figure 5.1) the tool smooths the images using a Gaussian filter to remove noise. Then, MATLAB's polygon tool is used to manually select a large region in the healthy contralateral hemisphere without ventricles or MRI artefacts. The mean and SD of the intensity in this region, I_{contra} and SD_{contra} respectively, are used to calculate an intensity threshold using the formula

$$T_I = I_{contra} + f_{T_I} SD_{contra} \quad (5.1)$$

T_I is applied on the smoothed slices to obtain a binary mask of the infarct; voxels with an intensity higher than T_I are classified as infarct. T_I depends on a factor f_{T_I} , as used previously by Veldhuis et al. (2003), Jiang et al. (2006) and Modo et al. (2009). The produced binary masks are intersected with corresponding masks of the ipsilateral hemisphere to identify common pixels that lie within the hemisphere. Basic morphological operations in MATLAB are applied to refine the masks automatically, including filling of holes, closing of local regions using a disk-shaped structuring element and removing of small spurious regions. These processes are performed on each slice of the 2D scan where the infarct is present; slices are specified by the user. Eventually, the largest connected object across slices is detected and preserved.

Following initial selection of the infarct, the regions can be manually edited using the MATLAB tool to refine the boundary of the infarct. MATLAB's freehand tool was

incorporated to allow selection of regions to be added or removed. When adding regions, the new infarct masks are intersected again with corresponding masks of the ipsilateral hemisphere. The final infarct volume is estimated by the sum of all voxels included in the selected regions.

To optimise the performance of the analysis pipeline, a standard MRI dataset (Section 5.2.3) was analysed repeatedly by altering parameter values of all crucial analysis steps, as shown in Table 5.1. Manual editing of the infarct regions was skipped (last step in Figure 5.1), in order to optimise the automated part of the pipeline alone. Initially, the performance of bias field correction when different regions are taken into account was examined. Binary masks of the regions listed in Table 5.1 were fed to the N4 algorithm, except when the whole image was taken into account (no masks required). After the optimal mask for bias field correction was determined, the size of the Gaussian smoothing filter, the radius of the disk-shaped structuring element for morphological closing and, finally, the factor f_{T_I} determining the intensity threshold in equation 5.1 were altered in order; the optimal value for each processing step was set for optimisation of the rest.

Table 5.1 Optimising the performance of the infarct segmentation tool.

Bias field correction masks	Size of Gaussian filter (pixels)	Radius of structuring element for closing (pixels)	Intensity threshold factor f_{T_I}
No masks	No smoothing	No closing	1.0
Ipsilateral hemisphere	3×3, SD = 1	1	1.2
Contralateral hemisphere	3×3, SD = 2	2	1.4
Brain	5×5, SD = 1	3	1.6
	5×5, SD = 2	4	1.8
	7×7, SD = 1	5	2.0
	7×7, SD = 2		2.2
			2.4
			2.6
		2.8	
		3.0	

Analysis of a standard MRI dataset was repeated by using different parameters for each of the four processing steps to select the optimal, starting from the step on the left while using the parameters in bold for the other steps. SD indicates standard deviation in pixel values (an SD = 1 corresponds to a full width at half maximum of 2.35 pixels).

5.2.2 Graphical user interface

Figure 5.2 shows a screenshot of the **RodentAnalysisGUI** developed in MATLAB for assessing coronal rodent brain scans interactively and measuring infarct volume. The GUI incorporates all processing steps within the blue box in Figure 5.1. Using the buttons in the **Inputs** panel on the right of the GUI, a brain scan and corresponding masks of the ipsilateral hemisphere are loaded. The voxel size is read from the raw data, but this and the number of slices where the infarct is present can be adjusted in the **Inputs** panel by the user. The intensity threshold can also be adjusted in this panel by setting factors that alter the contribution of either term in equation 5.1.

The user loads or roughly draws a region in the middle slice in the healthy contralateral hemisphere that is used to estimate the intensity threshold (as shown in the middle panel in Figure 5.2 a). By clicking the **Measure volume** button in the **Analysis** panel, the tool measures and outputs on the GUI the infarct volume, and in MATLAB's Command Window the measurements for each slice. By clicking the **Edit** button, the panel in the middle of the GUI changes into an edit mode, which can be used to manually refine the boundary of the infarct (Figure 5.2 c). The identified infarct is shown in the **Output Data** figure; the user chooses between overlaid edges or masks for display from a drop-down list. The final infarct masks for all analysed slices are saved in the image's folder by checking the **Save masks?** box in the **Analysis** panel. The GUI also allows manual editing of pre-saved masks by clicking the **Load to edit** button in the **Analysis** panel.

The tool with all associated code, instructions for its use and an example scan from the validation dataset described below are freely available for download from GitHub (https://github.com/Edinburgh-Imaging/Infarct_Segmentation_RodentMRI-).

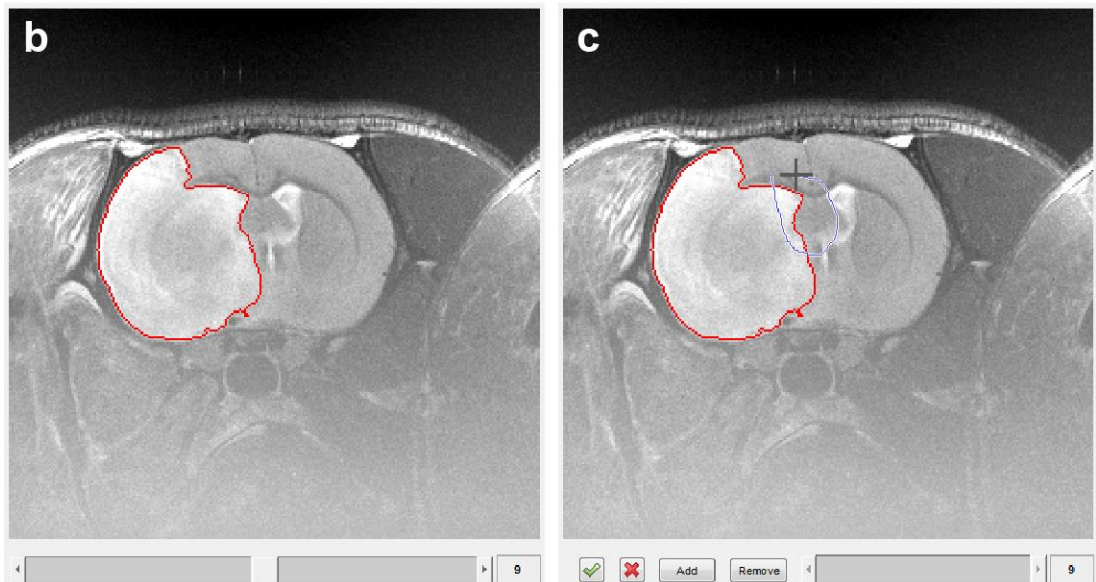
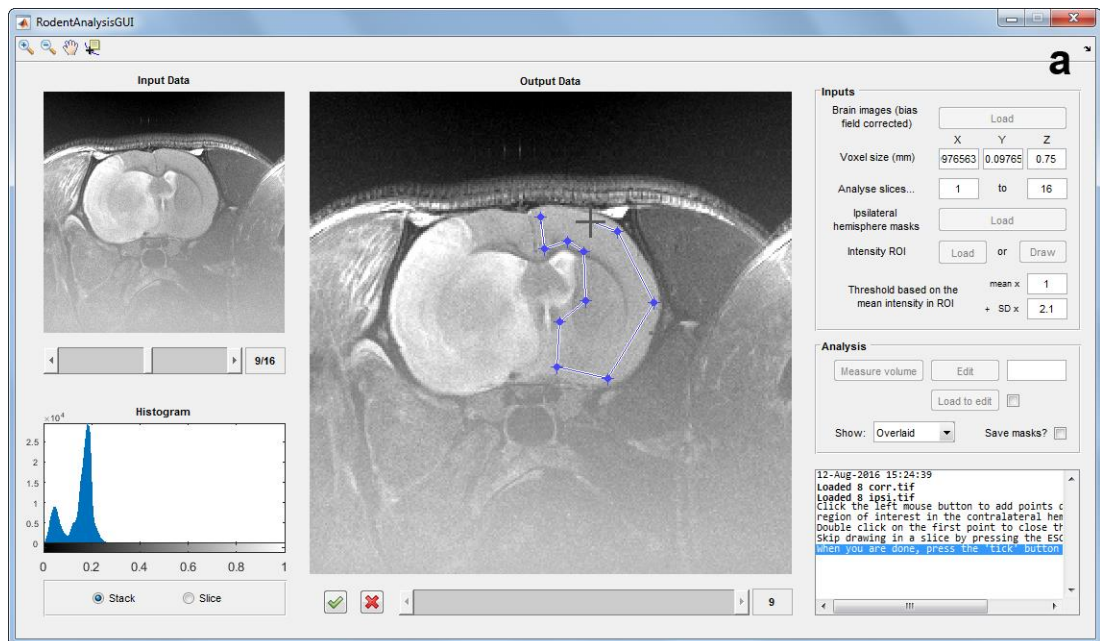


Figure 5.2 A graphical user interface for the measurement of infarct size. On the left of (a) the input slices and their histogram are shown, while on the right the panels used for controlling and carrying out the analysis are shown. On the middle of the tool the user can (a) select a region in the healthy contralateral hemisphere for estimating the threshold, (b) view the segmentation results and (c) manually edit the segmented regions. The red lines over the image in (b, c) represent the detected boundary of the infarct.

5.2.3 Validation dataset

To examine the accuracy of segmentation, coronal brain T₂-weighted scans of 21 different Sprague-Dawley rats were analysed using the procedure described above. This dataset was acquired as part of a study assessing the influence of gender on the evolution of ischaemic injury and penumbra (Baskerville, Macrae, Holmes, & McCabe, 2016). The rats were subjected to permanent MCAO using the intraluminal filament suture model described in Section 2.1.3. A 7 T scanner was used to perform T2WI at 24 hours after occlusion of the artery to visualise the extent of the infarct.

Scans were analysed manually using the freehand selection tool in ImageJ to trace the infarct, the ipsilateral and contralateral hemispheres and create corresponding masks. Custom-written code in MATLAB was used to create an image sequence containing the masks of each hemisphere, or combine the masks for each slice to create a sequence with masks of the whole brain. These were used to test the performance of bias field correction. Masks of the ipsilateral hemisphere were used along with the corrected T2WI data to measure infarct volume with the developed MATLAB tool and optimise its performance, as described in Section 5.2.1.

5.2.4 Comparison with stereological point counting

While the accuracy of the developed methodology can be evaluated by comparison against gold standard manual analysis, the often poor reproducibility and speed of the latter may constitute inappropriate benchmarks. Therefore, an additional method for volume measurement with reported excellent reproducibility and efficiency in a range of clinical applications was used, and specifically stereological point counting (Acer et al., 2008; Saunders et al., 1995; Thrippleton et al., 2014). This is a method similar to manual planimetric analysis, but reduces analysis time by subsampling the data in all or fewer equidistant slices.

Point counting was performed on the uncorrected raw brain images of the Sprague-Dawley rats using the dedicated **Stereology** module in Analyze software (11.0; Mayo

Clinic, Rochester, Minnesota, USA). A randomly positioned and oriented grid of points is superimposed over the slices by setting the required slice gap and spacing between adjacent points. Points that lie over the ROI are selected using the cursor (Figure 5.3). In this study, a grid with a unique position and orientation for each animal was used, superimposed on all slices. Further strict criteria to reduce sampling bias were set: points that were exactly at the top-right boundary of the ROI were included but points at the bottom-left boundary were excluded. Similar to manual analysis, the volume of the ROI is calculated using the following formula:

$$V = td^2 \sum_{i=1}^n p_i \quad (5.2)$$

where t is the slice thickness, d is the distance between adjacent points (grid spacing), p_i is the number of points selected in each slice, i , and n is the total number of slices used (Garcia-Finana, Cruz-Orive, Mackay, Pakkenberg, & Roberts, 2003; Thrippleton et al., 2014).

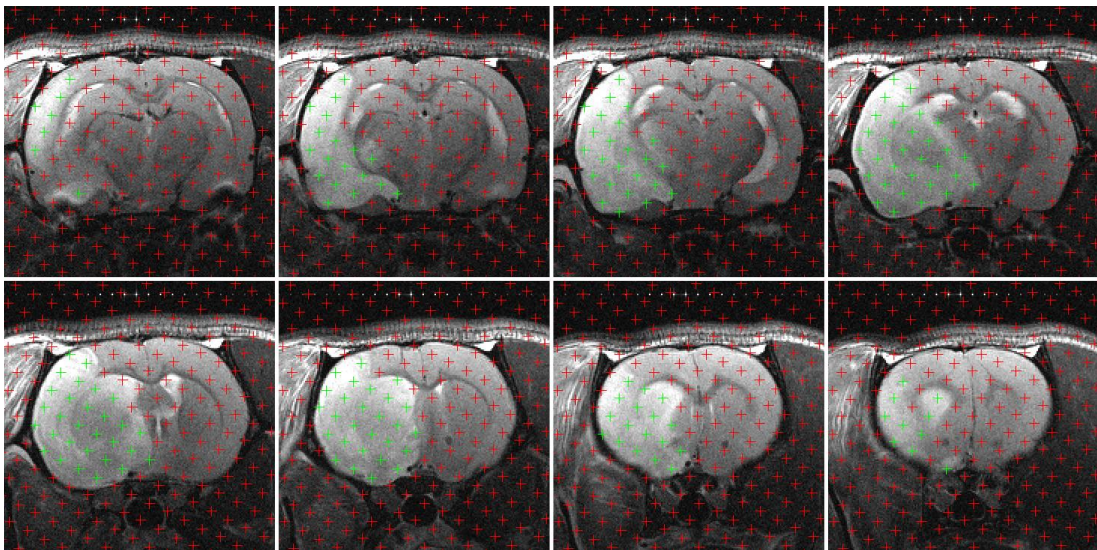


Figure 5.3 Measurement of infarct volume using stereological point counting. These eight coronal slices across the brain of a rat (caudal to rostral from top left to bottom right), show a randomly positioned and oriented grid of points (red). The user-selected green points represent the infarct. The spacing between adjacent points is 1.27 mm.

According to Keller and Roberts (2009), the grid should be dense enough to permit a reliable and precise volume estimation without compromising efficiency; at least 200 points on average must fall within the ROI to achieve a coefficient of error (CE) lower than 5.0%. To verify this, the CE was estimated in Microsoft Excel using the following formula:

$$CE = \sqrt{\alpha(q)(3(C_0 - \hat{v}) - 4C_1 + C_2) \left(\sum_{i=1}^n p_i\right)^{-2} + \hat{v} \left(\sum_{i=1}^n p_i\right)^{-2}} \quad (5.3)$$

where

$$\alpha(q) = \frac{\Gamma(2q + 2) \zeta(2q + 2) \cos(\pi q)}{(2\pi)^{2q+2} (1 - 2^{2q-1})} \quad (5.4)$$

$$C_k = \sum_{i=1}^{n-k} p_i p_{i+k}, \quad k = 0, 1, \dots, n - 1 \quad (5.5)$$

$$\hat{v} = 0.0724 \frac{P}{\sqrt{A}} \sqrt{n \sum_{i=1}^n p_i} \quad (5.6)$$

P/\sqrt{A} in equation 5.6 is a shape factor, determined by the area, A , and the perimeter, P , of the ROI over a few representative slices; it equals 3.54 for a circle and increases for more irregular shapes. Γ is the gamma function and ζ the Riemann zeta function estimated using MATLAB's **gamma** and **zeta** functions respectively; $\pi = 3.14159$. q takes values between 0 and 1, with smaller values denoting more irregular objects; for $n \geq 5$, it is estimated using the formula:

$$q = \max \left\{ 0, \frac{1}{2 \log(2)} \log \left(\frac{3(C_0 - \hat{v}) - 4C_2 + C_4}{3(C_0 - \hat{v}) - 4C_1 + C_2} \right) - \frac{1}{2} \right\} \quad (5.7)$$

The first term in the square root of equation 5.3 is the contribution of the variability between slices and the second term is the contribution of the variability due to point counting within slices (Garcia-Finana et al., 2003).

The validation dataset was analysed using point counting to calculate the volume of

the infarct, the ipsilateral and contralateral hemisphere. Based on the mean volume of each structure estimated by manual tracing in ImageJ, ideal grid sizes providing on average at least 200 points over each structure were determined. Manually traced regions in slices from five randomly selected animals were used to estimate mean values for A and P and then the shape factor in equation 5.6 for each structure. Random selection was performed using the RANDBETWEEN function in Microsoft Excel.

5.2.5 Measurement of corrected infarct volume

The volumes of the two brain hemispheres are necessary to account for the space-occupying effect of vasogenic oedema and brain swelling when calculating infarct volume, which is significant at the time point when the rat dataset was acquired (24 hours following MCAO). According to Swanson et al. (1990), the volume of the vasogenic oedema in the ipsilateral hemisphere is equal to the difference in the volumes of the two brain hemispheres; the infarct volume is then corrected using the following formula:

$$V_I = V_{I_u} + V_{CH} - V_{IH} \quad (5.8)$$

where V_{I_u} is the uncorrected infarct volume (measured directly using each method), V_{IH} the volume of the ipsilateral hemisphere and V_{CH} the volume of the contralateral hemisphere. However, this formula does not take into account the fact that the healthy contralateral hemisphere is compressed by the oedema to the same extent as the ipsilateral hemisphere is extended, whereas the infarct is not compressed. Gerriets et al. (2004) defined a compression factor to account for this:

$$CF = \frac{V_{CH} + V_{IH}}{2V_{CH}} \quad (5.9)$$

The corrected infarct volume, V_{I_c} , is then

$$V_{I_c} = V_I CF \quad (5.10)$$

As mentioned previously, the validation dataset was analysed by myself to measure V_{I_u} , V_{IH} and V_{CH} using manual tracing in ImageJ and stereological point counting in Analyze. Semi-automated segmentation in MATLAB was used to measure V_{I_u} alone; manual measurements of V_{IH} and V_{CH} were used to correct these V_{I_u} measurements for brain swelling. V_{IH} and V_{CH} were measured on slices including infarcted tissue. Analysis was guided by a commonly used brain atlas (Paxinos & Watson, 1997), while training in infarct identification was provided by an experienced neuroscientist (Dr Chris McCabe, University of Glasgow).

5.2.6 Statistical analysis

Similar to the methodology described in Section 4.3.5, the processing steps of the semi-automated segmentation method (Table 5.1) were optimised based on DC estimated using equation 4.7, which assesses the spatial overlap between manually and semi-automatically defined regions (Dice, 1945). To estimate the optimal intensity threshold factor f_{T_I} in equation 5.1, a piecewise cubic polynomial (smoothing spline curve) was fitted to the mean DC estimates using the MATLAB's Curve Fitting Tool with default parameters; the optimal f_{T_I} corresponding to the maximum DC value on the curve was identified using the **feval** function in MATLAB.

The accuracy of the optimised semi-automated tool and point counting was assessed by comparison against manual measurements in images from all 21 animals performed by myself, which were considered as the ground truth. The agreement between groups of measurements was examined using Bland-Altman analysis (Bland & Altman, 1986), as described in Section 4.3.5. DC was used to compare the performance of the optimised semi-automated method against manual analysis; stereological point counting does not produce masks that can be compared using DC. The time required for delineating the infarct in each of the 21 animals using each method was measured using a stopwatch. SPSS was used to compare measured times, V_{I_u} and V_{I_c} between methods using repeated measures ANOVA with a

Greenhouse–Geisser correction where sphericity was violated (Mauchly's test; Table C.2) and Bonferroni's post hoc tests to correct for multiple comparisons (repeated-measure factor is the analysis method). Distributions of measurements were normally distributed, as determined by the Shapiro–Wilk test for normality (Table C.1). A p -value less than 0.05 was considered statistically significant.

Finally, ten animals were randomly selected using the RANDBETWEEN function in Microsoft Excel to examine intra-rater reliability and inter-rater reproducibility of all three methods in the measurement of infarct volume. For this reason, data were analysed twice by myself and once by Ross Lennen (University of Edinburgh); the latter was blinded to the results of my analysis. For repeated point counting by either rater, the grids used in the first analysis by myself were reused, so that any potential variability in measurements can be attributed to the raters and not to grid differences (Thrippleton et al., 2014). For each rater, repeated analyses by various methods were performed with an interval of at least two weeks to avoid any learning effects. Each rater had at least 3 years of image analysis experience at the time of analysis. The agreement between groups of measurements and delineated infarct regions was examined using Bland-Altman analysis and DC; since the same grid was used for each animal for point counting, DC was estimated by identifying common points.

Due to the small number of images used in intra- and inter-rater analyses (10), DC values were summarised using boxplots showing median values and ranges. For all other analyses, DC values were summarised using the mean and SD.

5.3 Results

5.3.1 Selection of parameters for semi-automated segmentation

The semi-automated segmentation tool was optimised before comparison with manual analysis and stereological point counting, by repeated analysis of the data using the parameters in Table 5.1. Initially, different methods for correcting the bias field were examined by feeding masks to the N4 algorithm in 3D Slicer to guide the correction; each mask led to noticeably different results (Figure 5.4). If no mask is given (default method) the correction fails to detect the true bias field in the images produced primarily by the surface RF receive coil. Masks that excluded the areas surrounding the brain led to detection of more slowly varying bias fields across the FOV, with correction based on the healthy contralateral hemisphere providing the best estimate of the expected bias field in the images.

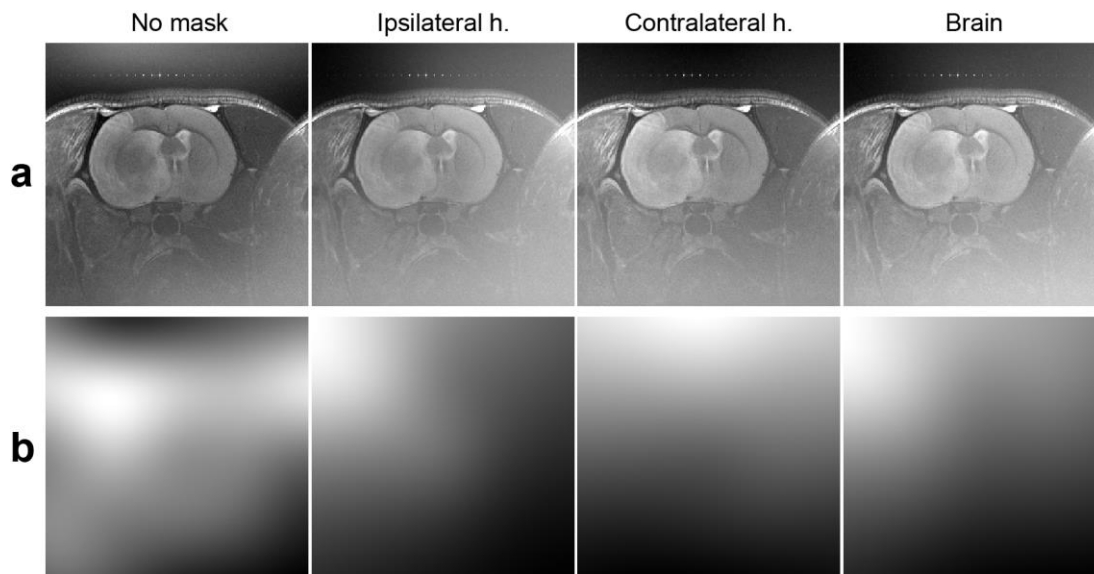


Figure 5.4 Comparison of bias field correction methods.

The images show (a) a corrected slice through the middle of the brain (uncropped matrix) and (b) corresponding bias fields estimated in 3D Slicer when various masks are used, as indicated at the top. The bias field when masks of the contralateral hemisphere are used is similar to the expected radial intensity gradient produced by the surface coil that sits at the top of the rat's head. h. indicates hemisphere

Figure 5.5 shows how significant the bias field correction is to the semi-automated analysis pipeline (Figure 5.5 a). The spatial overlap between manually defined and semi-automatically segmented infarct volumes, as determined with the DC, was poor when images were corrected based on the whole FOV (mean \pm SD = 0.085 ± 0.057), the ipsilateral hemisphere (mean \pm SD = 0.020 ± 0.045) and the whole brain (mean \pm SD = 0.135 ± 0.089). DC was highest for the contralateral hemisphere (mean \pm SD = 0.868 ± 0.051); data for all subsequent analyses using the semi-automated method were corrected based on masks of this hemisphere.

The accuracy of segmentation was not influenced in a similarly great extent by other analysis steps (Figure 5.5 b-d). Identified infarct regions were mostly identical for various sizes of the Gaussian filter used for smoothing the images and various radii of the disk-shaped structuring element used for morphologically closing the binary infarct mask. For smoothing, the minimum DC was obtained when bias field corrected images were not smoothed (mean \pm SD = 0.861 ± 0.058) and the maximum for the small Gaussian filter of size 3×3 pixels and an SD of 1 pixel (FWHM = 2.35 pixels; DC: mean \pm SD = 0.867 ± 0.052); the latter was used for subsequent analyses. For morphological closing, the minimum DC was obtained for a radius of the structuring element of 5 pixels (mean \pm SD = 0.857 ± 0.056) and the maximum for no closing (mean \pm SD = 0.867 ± 0.052); the latter was set as the default in the MATLAB tool.

The second most influential analysis step is the intensity threshold used for the initial selection of the infarct; the factor f_{T_I} determining its value (equation 5.1) was altered between 1.0 and 3.0 after all other steps were optimised. Then, a curve was fitted to the mean DC estimates to determine the optimal f_{T_I} that corresponded to the peak of the curve (maximum DC = 0.868; Figure 5.5 d); this was found to be $f_{T_I} = 2.1$ and was set as the default in the developed MATLAB GUI. The fitted curve had an excellent goodness of fit ($\bar{R}^2 = 0.999$).

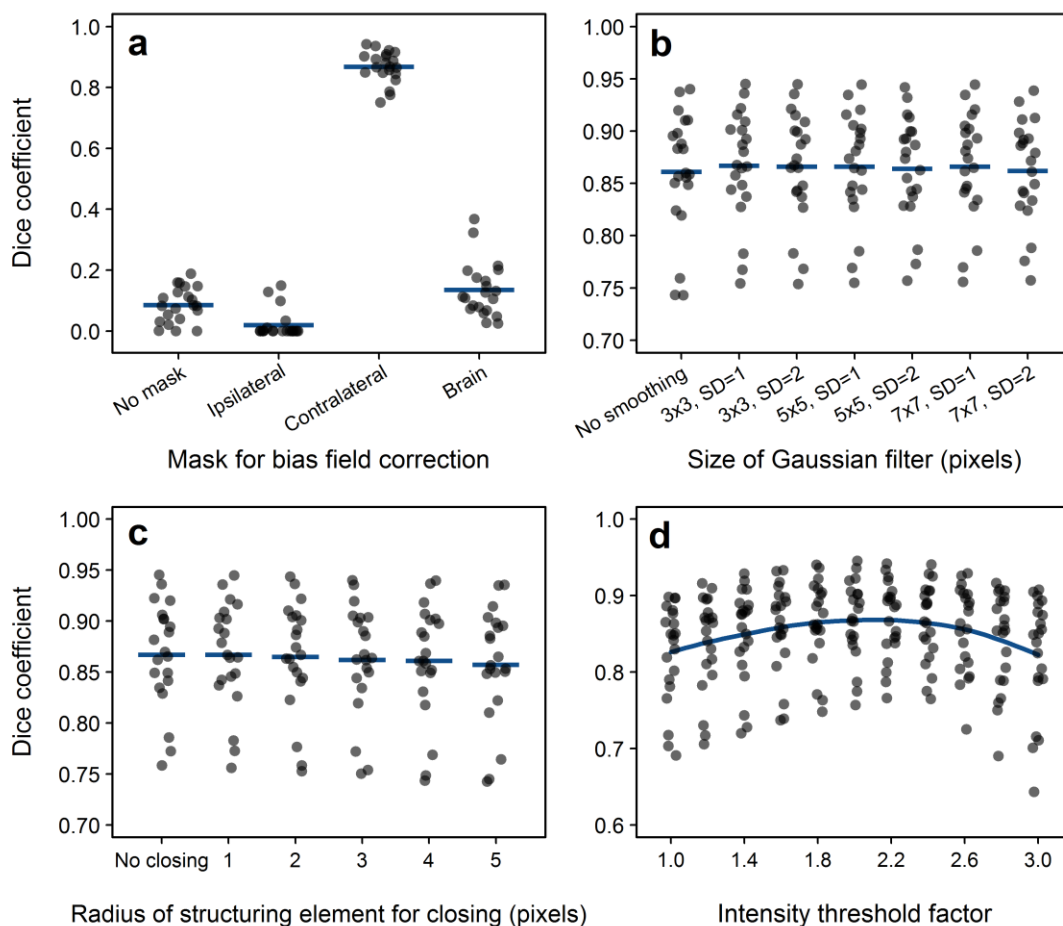


Figure 5.5 Dice coefficient for various parameters in the semi-automated analysis pipeline. Different (a) masks to correct images for bias field, (b) sizes of Gaussian smoothing filter, (c) radii of structuring element for morphological closing, and (d) intensity threshold factors were used. The four analysis steps were optimised in the order shown, starting with default parameters: size of the Gaussian filter: 3×3, SD = 1; radius of structuring element for closing: 1; intensity threshold factor: 2. The thick blue line for each group of measurements indicates the mean Dice coefficient; for (d), a curve was fitted to the data to accurately define the optimal threshold factor. SD indicates standard deviation.

Figure 5.6 shows the detected infarct without any manual editing in a coronal slice through the middle of the brain of a representative animal, for various parameters of the four analysis steps. Bias field correction using different masks results in highly heterogeneous infarcts, with the algorithm often failing to select any voxels when the ipsilateral hemisphere is used (Figure 5.6 a). Increasing the size of the Gaussian filter or the intensity threshold leads to selection of fewer voxels (Figure 5.6 b, d), whereas increasing the radius of the structuring element for morphological closing causes

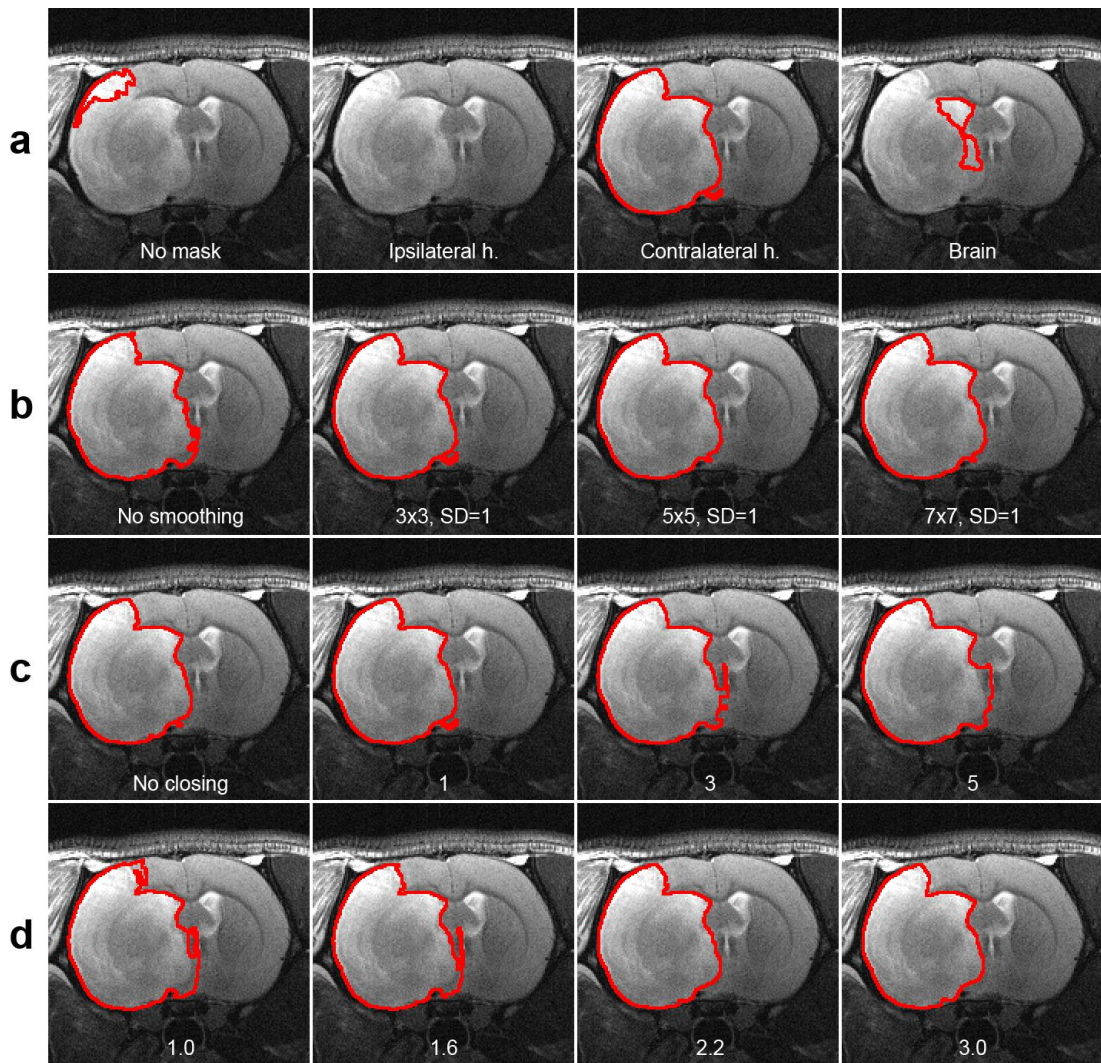


Figure 5.6 Infarct for various parameters in the semi-automated analysis pipeline. The detected boundary of the infarct before any manual editing is shown on the same coronal slice through the middle of the brain of a rat (red). Different (a) masks to correct images for bias field, (b) sizes of the Gaussian smoothing filter (in pixels), (c) radii of the structuring element for morphological closing (in pixels), and (d) intensity threshold factors were used. Bias field correction using masks of the ipsilateral hemisphere led to complete miss of the infarct in this animal. h. indicates hemisphere; SD, standard deviation.

selection of more pixels around the infarct boundary (Figure 5.6 c).

5.3.2 Selection of parameters for stereological point counting

For stereological point counting, the shape factor (P/\sqrt{A} in equation 5.6) for the infarct, the ipsilateral and contralateral brain hemispheres was determined as 4.35, 3.89 and 4.15 respectively, indicating that the infarct is more irregularly shaped than the hemispheres. Using the manually determined mean volume of each structure, the optimal grid spacing (d in equation 5.2) was determined to be 1.27 mm, 1.86 mm and 1.66 mm respectively (or 13, 19 and 17 pixels respectively for the dataset used). Grids generated using this spacing provided on average 227.14, 213.19 and 214.24 points within the whole volume of each structure respectively. The mean CE (equation 5.3) was estimated as 2.0%, 3.1% and 3.4% for these structures respectively, well below the maximum CE of 5.0% for reliable point counting.

5.3.3 Comparison of methods

Figure 5.7 assesses the agreement between the two methods and gold standard manual tracing for the measurement of infarct volume. Stereological point counting was found to be more accurate than semi-automated segmentation and with a lower dispersion of differences (mean difference is -0.2 mm^3 and 6.7 mm^3 respectively).

Nevertheless, the DC was high for semi-automated segmentation, even before manual editing of the detected infarct using the developed tool (pre-editing: mean \pm SD = 0.868 ± 0.052 ; post-editing: mean \pm SD = 0.931 ± 0.022 ; Figure 5.8 a). The time required for analysis of each scan differed significantly between methods, as revealed by a repeated measures ANOVA test ($F(2, 40) = 257.166$, $p < 0.001$; Figure 5.8 b). Pairwise comparisons performed with post hoc tests using the Bonferroni correction also indicate that semi-automated analysis including manual editing is significantly faster than manual tracing and point counting (mean \pm SD is 3.68 ± 0.88 , 10.54 ± 1.58 and 4.76 ± 0.88 minutes respectively; $p < 0.001$ for all pairwise comparisons).

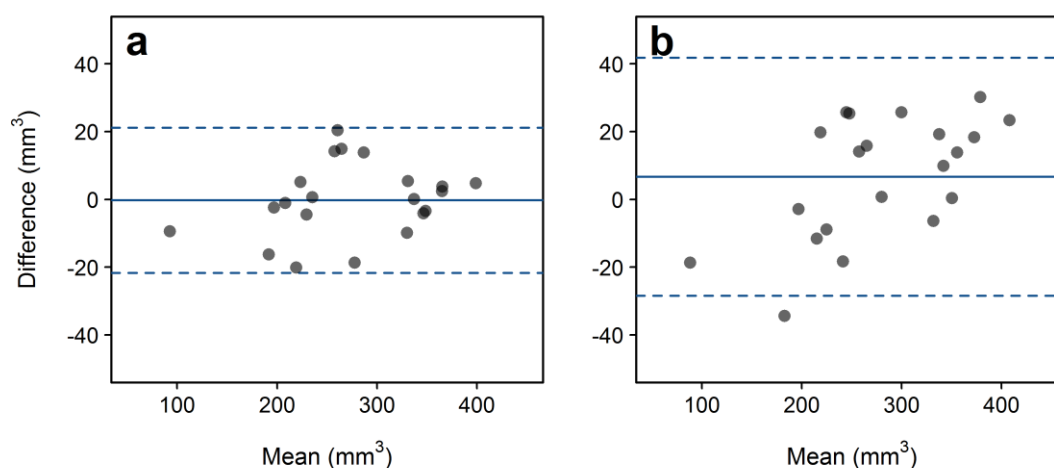


Figure 5.7 Bland-Altman plots comparing methods for measuring infarct volume. The plots assess the agreement between manual tracing and (a) stereological point counting or (b) semi-automated threshold-based segmentation. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

Absolute V_{I_u} and V_{I_c} suggest that the lower accuracy and higher dispersion of measurements of the semi-automated tool compared with point counting did not affect its correlation with either of the other two methods (Figure 5.8 c, d). Repeated measures ANOVA showed that both V_{I_u} and V_{I_c} did not differ significantly between methods ($F(1.514, 30.283) = 2.415, p = 0.118$ and $F(2, 40) = 2.175, p = 0.127$ respectively, with a Greenhouse–Geisser correction for the former; Table C.2). As shown in Figure 5.8, the distributions of infarct volumes determined by each method were similar (for manual tracing, point counting and semi-automated analysis: uncorrected: mean \pm SD is $274.8 \pm 74.0 \text{ mm}^3$, $274.6 \pm 76.7 \text{ mm}^3$ and $281.5 \pm 84.4 \text{ mm}^3$ respectively; corrected: mean \pm SD is $190.2 \pm 55.8 \text{ mm}^3$, $187.0 \pm 69.6 \text{ mm}^3$ and $197.8 \pm 67.9 \text{ mm}^3$ respectively).

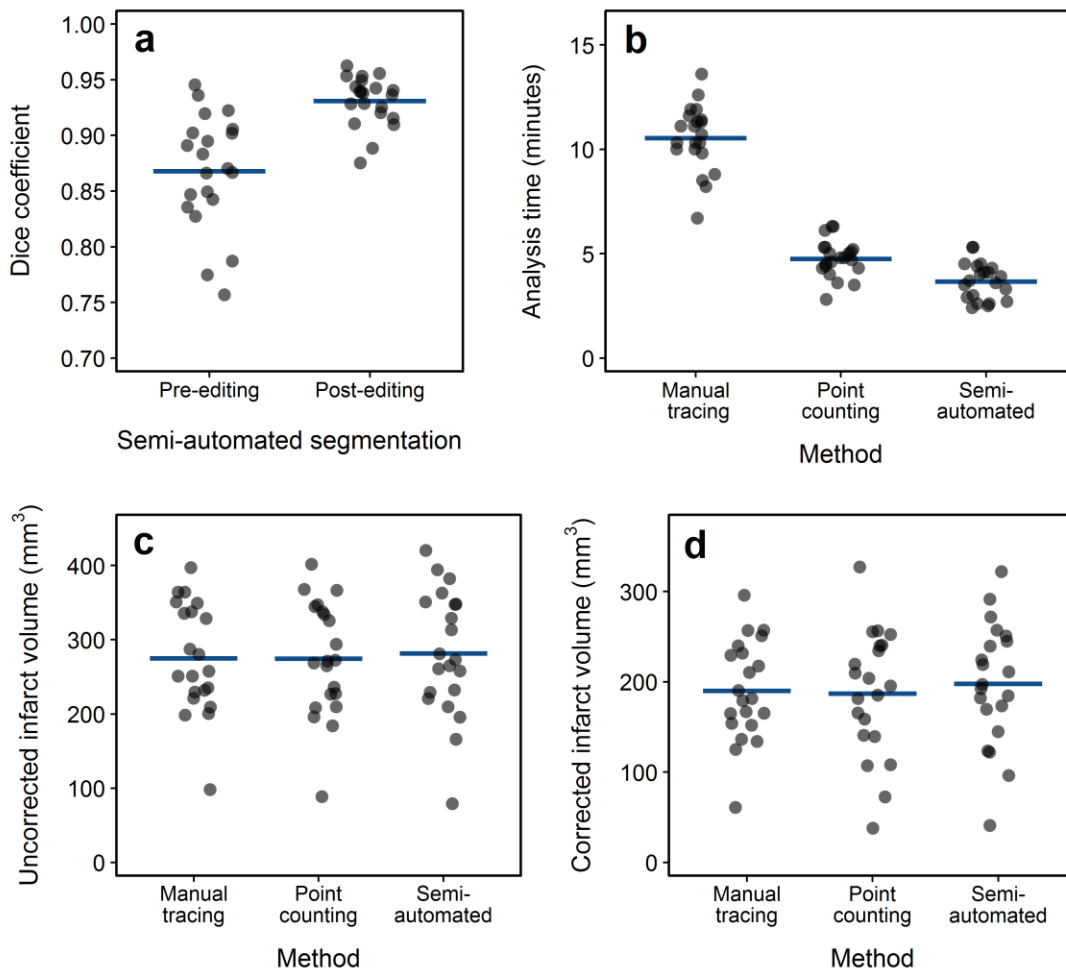


Figure 5.8 Dice coefficient, analysis time and infarct volume for all methods. The plots show (a) the Dice coefficient comparing infarct volumes defined by manual tracing and the semi-automated method (pre- and post-manual editing), (b) the time required for full analysis of each scan, (c) the uncorrected infarct volume and (d) the infarct volume corrected for brain swelling. The Dice coefficient comparing point counting and manual analysis cannot be estimated. The thick blue line for each group of measurements indicates the mean value. Analysis time is significantly different between all methods in analysis of variance pairwise comparisons ($p < 0.001$).

Figure 5.9 shows the infarct regions detected by the optimised semi-automated tool in a representative scan with a low DC and a representative scan with a high DC before manual editing. Selected regions tend to include false positives in slices at the caudal or rostral end of the MCA territory where the true infarct area is often small (Figure 5.9 a), and may also include the hyperintense ventricles where they are adjacent to the lesion (Figure 5.9 c); these slices often require manual editing. However, semi-automated segmentation can assist in the identification of potentially

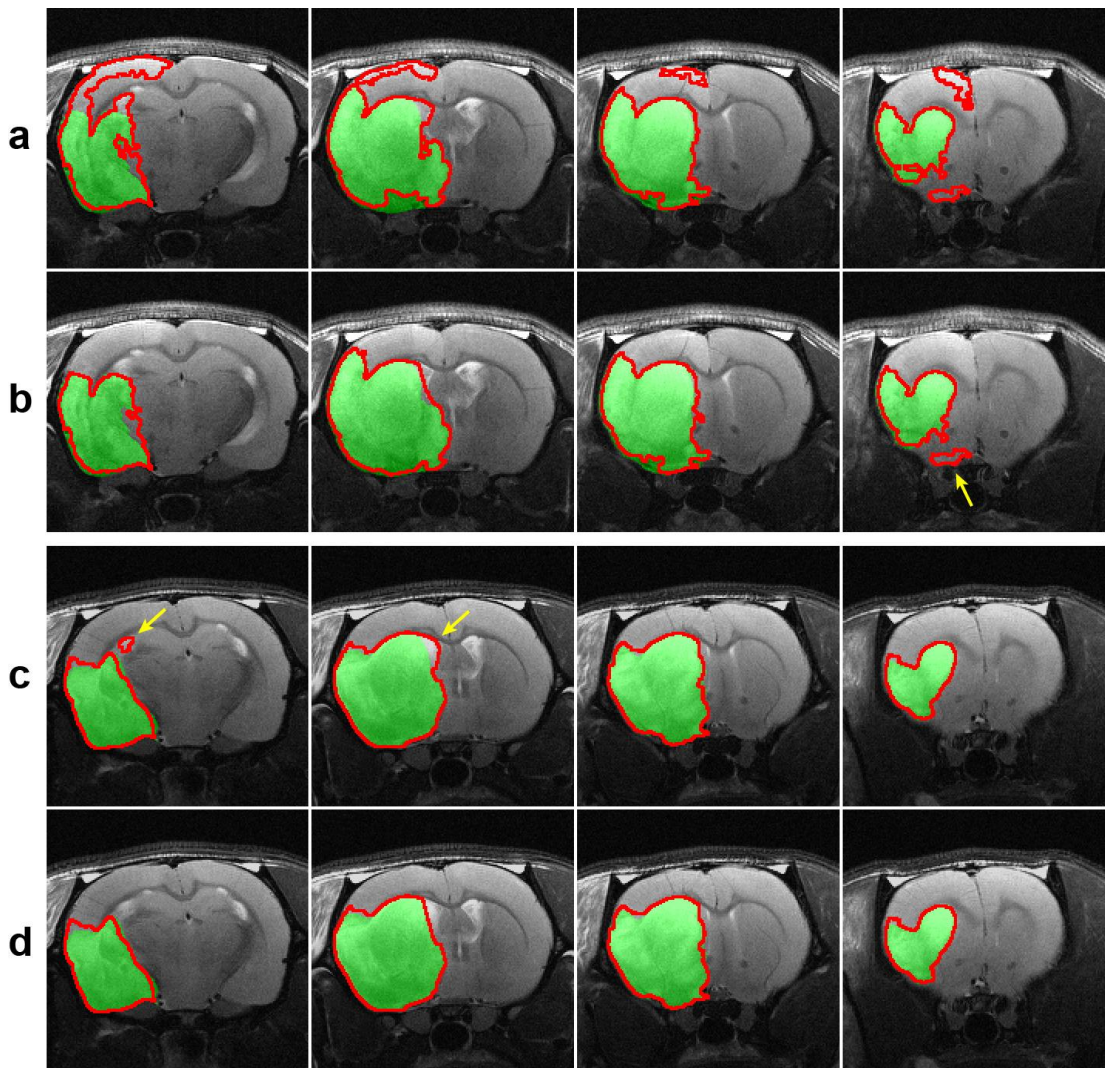


Figure 5.9 Infarct in two rats identified by semi-automated segmentation. These four coronal slices (caudal to rostral from left to right) show the infarct boundary for (a, b) an animal for which segmentation before manual editing had a low Dice coefficient and (c, d) an animal for which segmentation before manual editing had a high Dice coefficient. The ground truth manually traced infarct is shown in green (identical in a, b and in c, d), while the segmented infarct boundary is shown in red (a, c) before manual editing and (b, d) after manual editing. Semi-automated segmentation can assist in the identification of infarct areas that manual analysis fails to identify (arrow in b), but sometimes includes areas that are not infarct (for example, the ventricles indicated by the arrows in c). Semi-automated segmentation had a Dice coefficient of 0.775 and 0.936 before manual editing (a and c respectively), and 0.915 and 0.949 after manual editing (b and d respectively).

infarcted tissue (Figure 5.9 b).

5.3.4 Intra- and inter-rater agreement

Ten of the 21 scans were randomly selected to perform repeated analyses using all three methods to assess their intra-rater reliability and inter-rater reproducibility. Bland-Altman plots in Figure 5.10 show that the dispersion of differences between corresponding measurements in inter-rater analyses is higher than intra-rater analyses for all three methods, but the mean difference is smaller for manual tracing and point counting in inter-rater analyses. Manual tracing is the most reproducible method between different raters (mean difference is 0.5 mm^3 ; Figure 5.10 b), while the other two methods were equally reproducible (mean difference is 3.6 mm^3 for point counting and 2.8 mm^3 for semi-automated analysis; Figure 5.10 d, f). Semi-automated segmentation has the best reliability among all methods (mean difference is -0.6 mm^3 ; Figure 5.10 e).

Infarct regions defined by repeated analyses by the same or different raters have a high similarity overall, as indicated by DC in Figure 5.11. Expectedly, DC is higher for intra-rater analysis than for inter-rater analysis using either method. The same rater produced the most homologous infarcts using the semi-automated method (median = 0.988, IQR = 0.979–0.990) and the least homologous infarcts using manual analysis (median = 0.970, IQR = 0.959–0.975), despite the low variability in infarct volume differences for the latter (Figure 5.10 a). Median DC estimates for intra- and inter-rater analyses using stereological point counting are the most comparable (median = 0.972, IQR = 0.967–0.978 and median = 0.966, IQR = 0.949–0.969 respectively).

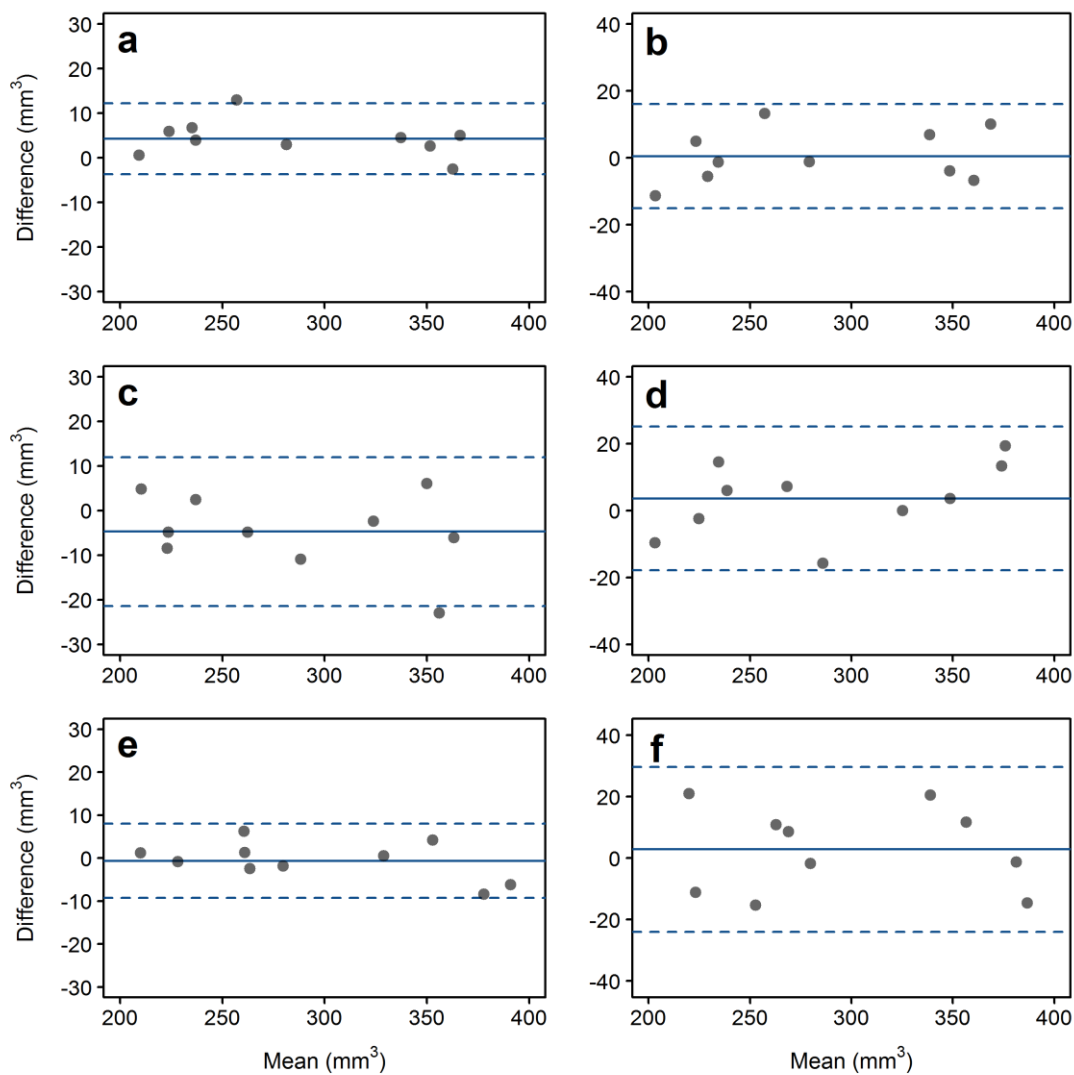


Figure 5.10 Bland-Altman plots assessing intra- and inter-rater agreement. Repeated analysis (a, c, e) by the same rater or (b, d, f) by two different raters is assessed for (a, b) gold standard manual tracing, (c, d) stereological point counting and (e, f) semi-automated threshold-based segmentation for measuring infarct volume. For estimating the differences, the measurements of the first analysis by the first rater were subtracted from the corresponding measurements of the second analysis or the second rater. The solid blue line in each plot indicates the mean difference and the dashed blue lines its 95% CI.

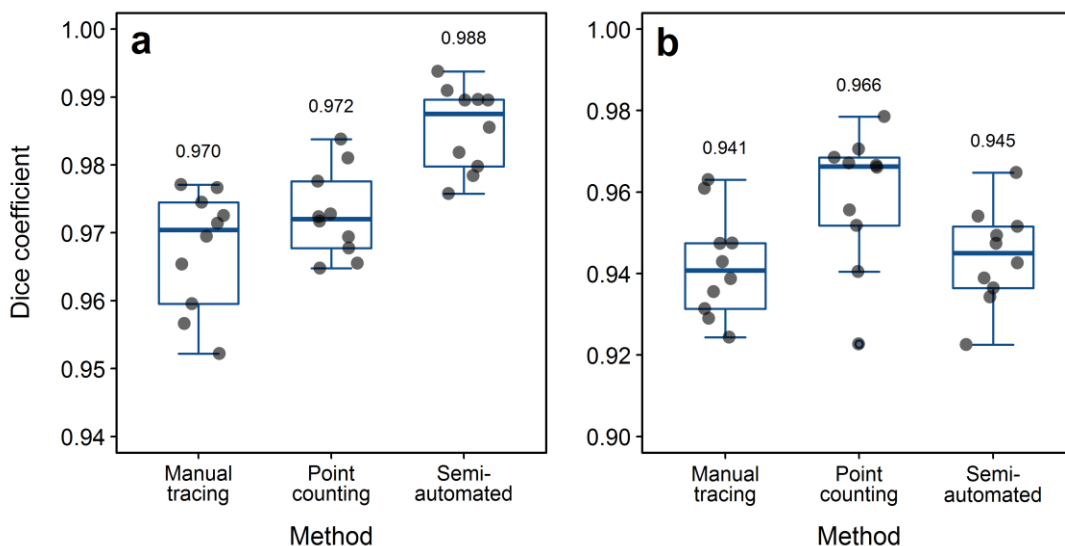


Figure 5.11 Dice coefficient comparing repeated analyses. The plots show the Dice coefficient for infarct volumes defined in (a) intra-rater and (b) inter-rater analyses using each method. The thick blue line of each boxplot indicates the median Dice coefficient (value given above).

5.4 Discussion

This chapter describes the procedure followed for the development of a semi-automated method for measuring infarct volume on 2D T₂-weighted images in rodent models of stroke. Its accuracy, reproducibility and speed was evaluated by analysis of a dataset of rat images acquired as part of a recent study assessing the influence of gender on the evolution of ischaemic injury and penumbra (Baskerville et al., 2016); measured infarct volumes were compared against those determined using gold standard manual planimetric analysis of the same data. Infarct volumes between the two methods did not differ significantly, but the developed method was found to be significantly faster and produced volumes in repeated analyses that had a greater similarity. The semi-automated method was further compared with point counting, a stereological technique often employed for the measurement of volumes in clinical studies (Acer et al., 2008; Saunders et al., 1995; Thrippleton et al., 2014). While stereological point counting is an accurate method when using a dense grid of points, semi-automated analysis is faster and more repeatable.

The pipeline for semi-automated analysis was devised based on observations made in previous studies, brain anatomy and stroke pathophysiology. Initially, widely used academic tools are employed to trace the brain hemispheres and correct the raw data for bias field. Then, a dedicated MATLAB GUI is used to apply an intensity threshold to binarise the images and define the boundary of the infarct using simple morphological operations. The infarct was defined as the areas in the ipsilateral hemisphere where the intensity is higher than the mean plus 2.1 SDs of the intensity in healthy tissue in the contralateral hemisphere sampled in the middle coronal slice. No study has ever applied a thresholding technique in qualitative T2WI acquired using surface RF coils before, which is currently a commonly used imaging protocol. A similar infarct definition was used in previous studies in rat models of stroke in intensity normalised images acquired using either quantitative T2WI or volume RF coils (Jiang et al., 2006; Veldhuis et al., 2003), but no study provided further details regarding the image analysis procedure. While the tool developed here performs an initial identification of the infarct with good accuracy (DC is larger than the normally accepted minimum of 0.7), optional manual editing of the identified infarct regions can be performed using the same tool to further improve measurement accuracy. The tool fills a gap in current preclinical stroke practice and can be used to assist manual segmentation of the infarct by detecting areas where necrotic tissue is present but is too dim to be traced by the human eye in common intensity heterogeneous images. Its performance is not affected by the severity of stroke, the extent and consequences of vasogenic oedema, the shape of the infarct or the alignment and positioning of the brain in the image that may confound unsupervised analysis techniques. It has a high intra-rater reliability, suggesting that its use in a multicentre study employing centralised data analysis could reduce the variability and improve the consistency of measurements, allowing the execution of more reliable comparisons between cohorts.

Despite these advantages, the proposed method requires use of various software for pre-processing the images, which could increase analysis burden and time in the long run, though not in a larger extent than full manual analysis. The N4 algorithm in 3D

Slicer software was employed to correct the images for bias field and homogenise the intensity across the brain (Tustison & Gee, 2010). The algorithm models the bias field by performing curve fitting in the log space and was successfully applied in a recent study to correct sMRI data of rats and mice (Oguz, Zhang, Rumple, & Sonka, 2014). A few MATLAB tools for performing bias field correction are available (Kroon D-J, 2009; Salvado, 2006), but are not optimised for use on stroke images and perform poorly for the data used in this study. Furthermore, the rat images were pre-processed to manually draw the brain hemispheres in ImageJ; masks of the contralateral hemisphere are used to perform bias field correction, whereas masks of the ipsilateral hemisphere are used to guide intensity thresholding to detect the infarct. Recently, various methods for automatic rodent brain segmentation in MRI data were proposed (Chou, Wu, Bai Bingren, Qiu, & Chuang, 2011; Oguz et al., 2014; Zhang et al., 2009). None of these techniques can segment each hemisphere separately, which is important in stroke not only for the aforementioned reasons, but also for correcting potential infarct volume overestimations due to brain swelling, or normalising measurements to V_{IH} or V_{CH} . In addition, these algorithms were validated on high resolution 3D brain images of healthy animals. Brains of ischaemic animals are characterised by asymmetrical anatomy and intensity; this factor alone could be sufficient to cause serious performance issues to these tools. The technique by Chou et al. (2011) was also applied on a 2D MRI dataset, but processing time ranged from 32 to 39 minutes per animal; this is longer than the time spent for manual tracing in this study. Due to time constraints, new methods for automatic identification of each brain hemisphere and bias field correction were not implemented in the developed MATLAB tool. Future work could focus on the development of a brain hemisphere segmentation method that utilises computer vision algorithms to adapt to the shape and appearance of the boundary of each hemisphere, without relying explicitly on assumptions of global intensity homogeneity (Babalola, Cootes, Twining, Petrovic, & Taylor, 2008; Wells, Grimson, Kikinis, & Jolesz, 1996). Then, a bias field correction step guided by the voxels in the segmented contralateral hemisphere should follow. In contrast with the lesioned ipsilateral hemisphere, the contralateral hemisphere

leads to a relatively precise estimate of the bias field, suggesting that the T_2 relaxation time and thus the signal across it is relatively homogeneous (Figure 5.12). A tool with this capacity will eliminate the need for manual pre- and post-processing of data for intensity normalisation and calculation of V_{Ic} respectively, improving its efficiency and bringing it a step closer to routine preclinical utility.

This is the first study employing stereological point counting for the measurement of the volume of the infarct and each brain hemisphere in a rodent model of stroke; shape factors for each ROI, as well as optimal grid sizes ensuring the presence of at least 200 points on average within each ROI for low measurement error ($CE < 5.0\%$) were determined for the intraluminal filament suture model in Sprague-Dawley rats scanned at 24 hours after stroke. While the size of each brain hemisphere in rats or mice is similar for most stroke models at this time point, infarct size could differ significantly and point counting parameters should be adjusted for its measurement in individual studies based on the expected infarct size.

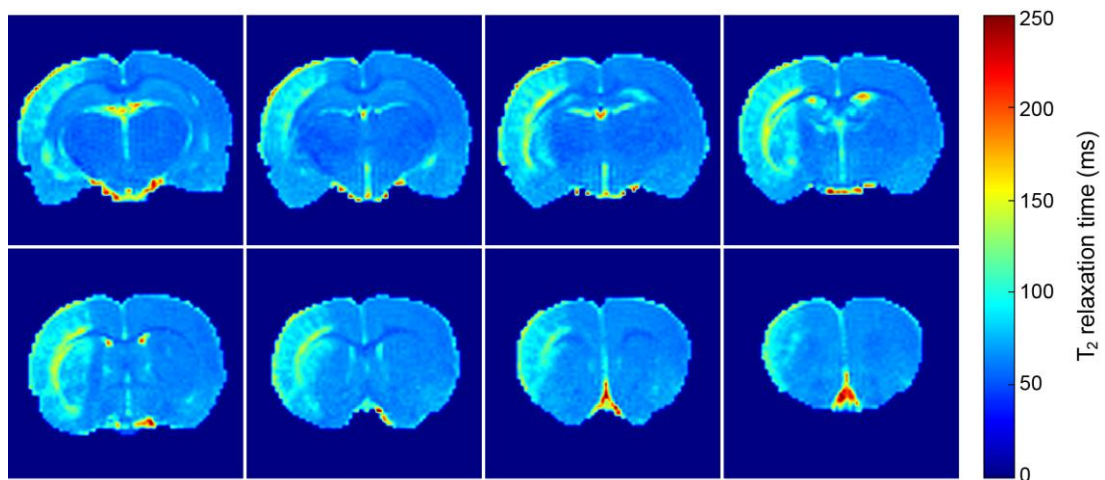


Figure 5.12 T_2 relaxation map in the rat brain following stroke.

These eight coronal slices (caudal to rostral from top left to bottom right) are of rat that was subjected to the same stroke model as the rats in this study and was scanned using quantitative T_2 -weighted imaging. The T_2 relaxation time is relatively uniform within the healthy contralateral hemisphere across slices (right side of brain), apart from the cerebrospinal fluid that has a slower relaxation rate. Provided by and used with permission from Dr Tracy Farr (University of Nottingham).

A well optimised point counting procedure has high accuracy and reproducibility in repeated analyses, as shown here, but it also has a particular limitation that might discourage its use in most modern experiments: it does not provide masks of the measured ROIs that are necessary for various post-processing analyses, such as assessment of its shape and appearance across animals, or template formation using registration methods. On the other hand, manually traced ROIs in ImageJ can easily be transformed into corresponding masks, while the developed MATLAB tool can save the masks automatically on demand.

A fast, accurate and reliable method for assisted manual analysis of 2D T2WI data in rodent models of stroke and measurement of infarct volume was developed. The improved repeatability of this method over gold standard manual analysis holds promise for the implementation of a standardised data analysis protocol in future large multicentre studies. The tool will have a central role in the multicentre simulation study presented in the following chapter, aiming to compare infarct volume and infarct-based study effects from a large number of simulated MRI scans across different preclinical scanners to examine the influence of MRI variability.

Chapter 6

Between- and Within-Scanner Variability

Using the structural MRI phantoms and the image analysis tool described in Chapter 4 a multicentre QA study is carried out, assessing geometric accuracy in six different preclinical MRI scanners across Europe as used for the assessment of infarct size in rat models of stroke. Longitudinal imaging over a period of six months is performed to examine the daily, weekly and monthly within-scanner variability in geometric accuracy. Then, T₂-weighted scans simulating the scaling errors in each scanner and the voxel size of the imaging protocol used at each centre are generated. The scans are analysed using the semi-automated tool presented in Chapter 5 to determine the impact of MRI variability on the measurement of infarct volume and related statistical group comparisons.

Parts of this chapter have been published or presented as:

Milidonis, X., Lennen, R. J., Jansen, M. A., Mueller, S., Boehm-Sturm, P., Holmes, W. M., . . . Marshall, I. (2016). Multicenter Evaluation of Geometric Accuracy of MRI Protocols Used in Experimental Stroke. *PLOS ONE*, 11(9), e0162545. doi:10.1371/journal.pone.0162545

Milidonis, X., Lennen, R. J., Sena, E. S., Macleod, M. R., & Marshall, I. (2015). Accuracy of geometric measurements in high-resolution MRI scanners: between- and within-system evaluation. *Presented at the 21st Annual Scientific Meeting of the British Chapter of the International Society for Magnetic Resonance in Medicine (ISMRM)*. London, UK.

6.1 Introduction

Currently, studies in animal models of stroke are performed at a single centre, use sample sizes that are too small to allow valid statistical comparisons between animal groups and have poor generalisability (Dirnagl, 2006; Landis et al., 2012). Multicentre studies can be the way forward, as data from multiple cohorts at participating centres are pooled to improve statistical power, which can be useful for the investigation of subgroups or detection of subtle effects (Bath et al., 2009; Dirnagl et al., 2013). Of great interest in animal models of stroke is sMRI for longitudinal *in vivo* measurement of various morphological outcomes and most importantly infarct size, the primary outcome used for the assessment of treatment effects (Macrae, 2011). However, combining sMRI data in multicentre studies can be problematic, as differences in the performance of scanners can introduce systematic variation in measurements and obscure the effects of interest. Furthermore, alterations in the performance of individual scanners over time, either due to normal drift or hardware and software upgrades, can affect the comparability of longitudinal measurements even within the same cohort. Of all possible confounders in the assessment of morphological outcomes, the geometric accuracy of the MRI scanner is probably the most important; this must be monitored and improved at regular intervals across centres in order to minimise errors and ensure correspondence between measurements.

Before undertaking large multicentre stroke studies involving sMRI, it is important to investigate whether the variability between and within scanners is small enough to allow pooling of data; if not, strategies to address this variability must be determined. Until now, the accuracy and reproducibility of sMRI in different preclinical scanners has not been evaluated, despite an extensive assessment of clinical scanners in multicentre studies over the years. Those studies found that geometric accuracy is influenced by magnetic field inhomogeneity, gradient nonlinearity and positioning within the scanner (Jovicich et al., 2006), hardware and imaging protocols (Chen et al., 2014; Kruggel, Turner, Muftuler, & The Alzheimer's Disease Neuroimaging Initiative, 2010), and system upgrade or normal drift (Takao,

Hayashi, & Ohtomo, 2011; Takao et al., 2013). Meanwhile, others compared scanners using QA approaches involving standardised phantoms that limit the variance associated with inter-individual differences; they found that many scanners failed to pass QA criteria essential to geometric accuracy, such as spatial resolution and slice thickness, with results being dependent on the type of sequence used (Ewers et al., 2006; Ihalainen et al., 2011). Recently, possibly the first multicentre preclinical stroke trial was performed and its findings hint that treatment effects based on MRI lesion volume may vary significantly between centres, but no examination of the role of MRI-related factors was carried out (Maysami et al., 2015).

This chapter presents the first multicentre study assessing the impact of MRI heterogeneity on geometric accuracy of different preclinical MRI systems, as used routinely in experimental stroke. The aim of the study was to observe the variability in stroke imaging for the assessment of infarct size and not perform an absolute comparison of identical systems and imaging protocols; study procedures mimicked those normally followed in such in vivo experiments. However, since lesion progression in stroke is a constantly evolving process (Karki et al., 2010; Wegener et al., 2006), a QA approach was chosen instead of an in vivo experiment, utilising the phantoms and image analysis tool described in Chapter 4. The phantoms were scanned using a standardised sequence and other centre-specific sequences to determine the potential benefit of standardised over heterogeneous imaging. Moreover, within-scanner stability was evaluated at a single centre over a period of six months, covering time points often used in longitudinal stroke studies to monitor lesion evolution. Further examination of both between- and within-scanner imaging data allowed identification of scanning parameters that mostly contributed to variability in geometric accuracy. Finally, synthetic scans were generated by transforming a standard in vivo dataset; this simulated simultaneous scanning of the same animals at the six participating centres and allowed between-scanner comparison under various pre-specified scenarios. This led to evaluation of the impact of MRI variability on infarct size and related study effects across centres.

6.2 Methods

6.2.1 Between-scanner variability

In this multicentre study, the scanner used in experiments presented in Chapter 4 and five additional preclinical MRI scanners were evaluated; all were located in six different centres across Europe (centres or scanners labelled 'A', 'B', 'C', 'D', 'E' and 'F'; Table 6.1). For comparison, the location of each centre 'B'–'F' is concealed throughout this chapter. All scanners had a magnetic field strength of 7 T and all except 'A' are routinely used for the assessment of infarct size in rat models of stroke by Multi-PART collaborators.

Each of the constructed identical phantoms described in Chapter 4 was sent to each participating centre. To ensure correspondence in imaging across centres, a scanning form with detailed instructions for phantom placement, slice positioning, scanning and preparation of data, as well as exemplar images, was pre-prepared and sent along with the phantom (see Appendix D); forms were filled in by participating centres and were sent back along with the acquired data. Each centre was asked to use the standard coil set-up for rat brain imaging; all used a volume coil for RF transmission and a surface coil for RF reception, the details of which are given in Table 6.1. Centre 'E' used a mouse head surface coil due to unavailability of the rat head surface coil at the time of imaging. Following the standard rodent neuroimaging procedure and as shown previously (Figure 4.2), the phantom was attached to the surface coil and the latter was placed on the animal cradle. The whole assembly was then positioned in the scanner so that the phantom was at the isocentre of the magnet. Before the start of the study, all six scanners were undergoing only annual preventive maintenance by external service engineers and none were calibrated for geometric accuracy using this or any other phantom by study participants. Furthermore, none of the systems employed a method for automatic gradient nonlinearity distortion correction.

Table 6.1 Details of scanners and imaging coils.

Details	A	B	C	D	E	F
Scanner manufacturer and model	Agilent Technologies	Bruker BioSpec 70/30	Bruker BioSpec 70/30	Magnex Scientific magnet, Bruker gradient coils	Bruker BioSpec 70/30	Bruker BioSpec 70/20
Field strength (T)	7	7	7	7	7	7
Magnet bore diameter (mm)	305	300	300	160	300	200
Maximum gradient strength (mT/m)	400	400	600	750	200	440
Scanner software and version	VnmrJ 3.2	ParaVision 5.0	ParaVision 5.1	ParaVision 5.0	ParaVision 5.1	ParaVision 6.0
ID of volume RF transmit coil (mm)	72	72	72	72	72	86
Type of surface RF receive coil	Rat head 2-channel phased array	Rat head 4-channel phased array	Rat head 4-channel phased array	Rat head	Mouse head	Rat head 2-channel phased array

ID indicates inner diameter; RF, radiofrequency. Scanner 'A' is that used in Chapter 4.

Each centre scanned the phantom using the optimised standardised QA sequence, as described in Chapter 4 (labelled 'a'; Table 6.2) and the centre's favourite in vivo sequence for the assessment of infarct size in the rat model of stroke (labelled 'b', 'c', 'd', 'e' and 'f'; Table 6.2), with the exception of scanner 'A' which was assessed using sequence 'a' alone, since no stroke experiments were performed in this centre prior to this study. All stroke sequences 'b'–'f' were basic structural T₂-weighted sequences, as shown in Table 6.2. Each sequence was used to acquire six scans in various imaging planes, as described by the QA protocol in Chapter 4 (Section 4.2.5). Phantoms were imaged before any other experiment on the day of scanning to ensure similar operating temperatures of the systems.

Table 6.2 Parameters of standardised and stroke pulse sequences.

Parameters	a	b	c	d	e	f
Type of sequence	FSE	RARE	MSME	RARE	RARE	RARE
TR (ms)	1600	5000	3375	2742	3000	3500
TE _{eff} * (ms)	20	47	11-176	33	24	33
NSA	2	2	1	4	1	4
ETL	4	8	N/A	8	4	8
rBW (kHz)	100.0	50.0	59.5	47.0	50.0	32.9
FOV (mm×mm)	19.2×19.2	25.0×25.0	19.2×19.2	40.0×40.0	25.0×25.0	25.6×25.6
Matrix size (pixels×pixels)	256×256	256×256	256×256	256×256	256×256	256×256
Slice thickness† (mm)	1	0.75	1	1	0.6	0.5
Number of slices	17	17	17	25	12	31
Voxel size (×10 ⁻³ mm ³)	5.6	7.2	5.6	24.4	5.7	5.0
Scan duration (min:sec)	3:28	5:20	10:48	5:28	3:12	7:28

Sequence 'a' is the standardised quality assurance sequence used in Chapter 4. In vivo T₂-weighted sequences 'b'–'f' are used at corresponding centres 'B'–'F' (Table 6.1) for assessing infarct size in rat models of stroke. ETL indicates echo train length; FOV, field of view; FSE, fast spin echo; MSME, multi-slice multi-echo; NSA, number of sample averages; N/A, not applicable; RARE, rapid acquisition with relaxation enhancement (the name of FSE sequence in Bruker systems); rBW, receiver bandwidth; TE_{eff}, effective echo time; TR, repetition time.

*Applies to all sequences except 'c', for which 16 TEs with an echo spacing of 11 ms were used.

†None of the sequences used an inter-slice gap.

6.2.2 Within-scanner variability

6.2.2.1 Gradient calibration

Based on the internal dimensions of the phantom, as measured in images taken using sequence 'a' in the between-scanner variability study, the three gradient coils of scanner 'A' were calibrated by altering the strengths of the three gradient coils until scaling errors in each orthogonal direction were nulled. This procedure is performed using the Digital eddy current compensation tool (DecTool) in the scanner's software VnmrJ 3.2, the GUI of which is shown in Figure 6.1 (Agilent Technologies, 2011).

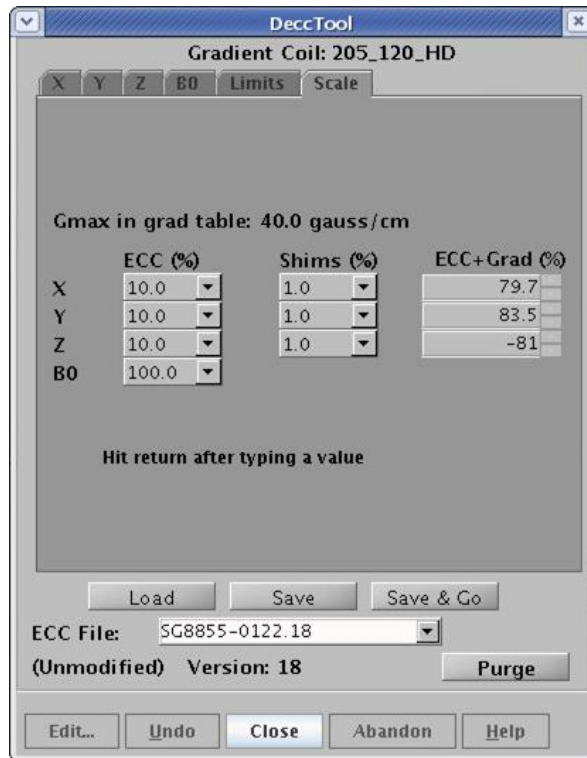


Figure 6.1 The digital eddy current compensation tool in VnmrJ software. The strength of each gradient coil depends on the **ECC+Grad (%)** scaling factor in the **Scale** panel, estimated by VnmrJ by providing the true and measured dimension of the phantom in the corresponding direction.

Basic gradient calibration of the other scanners can be performed using similar procedures, but assessment of within-scanner variability was restricted to Edinburgh's own scanner 'A' for logistical reasons.

6.2.2.2 Imaging

Starting immediately after calibration of all three gradient coils, scanning was performed at daily, weekly and monthly intervals, with five time points per interval; this corresponds to imaging over a total time period of about six months. The QA protocol involving centre's 'A' equipment (Table 6.1) and the standardised sequence 'a' (Table 6.2) was used to acquire six scans per time point, as described earlier. The phantom was kept in the scanner room during the whole imaging experiment to ensure temperature stability. Each scanning session including phantom positioning lasted between 1–1.5 hours.

6.2.3 Post hoc imaging

Post hoc imaging was performed with the calibrated scanner 'A' to examine the influence of possible confounders, as described below.

6.2.3.1 Comparison of imaging coils

The loop coil in a mouse head surface coil has a smaller diameter than that of a rat head surface coil, which makes it sensitive over a smaller ROI. As centre 'E' used a mouse head surface coil for RF reception instead of a rat head surface coil, imaging was performed to determine its impact on geometric accuracy that could possibly affect between-scanner comparisons of QA criteria. A rat head and a mouse head two-channel phased array surface coil that are used routinely in centre 'A' for rodent brain imaging, were used in conjunction with a volume RF transmit coil (centre 'A' in Table 6.1) to scan the phantom consecutively. Sequence 'a' was used to acquire the six scans for each coil, as described before.

6.2.3.2 Influence of magnetic susceptibility

Geometric distortion in images can be affected significantly by magnetic susceptibility artefacts that produce local field inhomogeneities; these depend not only on the difference in magnetic susceptibility between adjacent imaged structures, but can be amplified by increasing TE or decreasing receiver BW (rBW) (McRobbie et al., 2006; Schenck, 1996). In vivo stroke sequences 'b'-'f' had a longer TE and a narrower rBW than the standardised sequence 'a' (Table 6.2), therefore susceptibility effects in those sequences may be prevailing. To further examine these effects using scanner 'A', consecutive phantom scanning using three different sequences was performed: sequence 'a' and sequence 'b' as described in Table 6.2, and a modified version of 'b' (here labelled 'b_{modified}'); the latter sequence theoretically amplifies magnetic susceptibility effects by using a large effective TE (93 ms) and a narrow rBW (40.3 kHz), the narrowest allowed by the scanner's software given this value of TE_{eff}. Other parameters of 'b_{modified}' matched the parameters of 'b' in Table 6.2.

6.2.4 Simulations

6.2.4.1 Reference data

The scans used previously for the validation of the semi-automated infarct segmentation tool were as the reference data for in vivo simulation. In short, these are coronal brain T₂-weighted scans of 21 different Sprague-Dawley rats (males: $n = 10$, 300–350 g; females: $n = 11$, 250–280 g), acquired 24 hours after MCAO using scanner ‘B’ and imaging protocol ‘b’ described in Tables 6.1 and 6.2 respectively. They were recently used to assess the influence of gender on the evolution of ischaemic injury and penumbra (Baskerville et al., 2016).

6.2.4.2 Scaling using 3D affine geometric transformations

The freehand selection tool in ImageJ was used to manually trace the contralateral hemisphere in the reference scans and custom-made MATLAB functions were used to create corresponding binary masks. The scans and the masks were fed to the N4 algorithm in 3D Slicer (Tustison & Gee, 2010) to correct the bias field in the scans.

For data simulation, it was assumed that at small distances from the isocentre of each preclinical MRI scanner (covering the brain of a rodent), the predominant type of geometric distortion is linear scaling in each orthogonal direction due to improper calibration of gradient coils’ strengths, and that any global nonlinearities due to magnetic field inhomogeneity or local nonlinearities due to magnetic susceptibility artefacts in the brain are minute and similar between scanners (as suggested by phantom imaging); any nonlinearities due to either reason may already have been present in the reference scans. A function was written in MATLAB to linearly scale the bias field corrected scans in each direction by applying a 3D affine geometric transformation, represented by the following transformation matrix:

$$A_s = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.1)$$

where s_x , s_y and s_z are the scaling factors in x , y and z direction, respectively, in multiples of the image size (for example, for $s_x = 2$ the image size is doubled in x direction). All zero elements denote that no translation, rotation and shearing is performed, whereas the element of 1 at the bottom right corner is necessary for normalisation of the matrix in homogeneous coordinates. MATLAB's **affine3d** function is used to create the transformation matrix and **imwarp** function is used to apply the transformation to an image. Bicubic interpolation was chosen for generating the new images.

The centre of the produced multi-slice scan is the same as that of the reference scan, which simulates identical positioning of each animal in the scanners. However, depending on the scaling factors, the outermost pixels or slices in the produced scan that are located outside of the boundaries of the reference scan may be interpolated only partially. This is crucial for the subsequent semi-automated threshold-based analysis of the data, as higher or lower pixel intensities within the infarct at these locations could lead to segmentation failure. To avoid this, the developed MATLAB function estimates the median intensity in the middle slice of the reference scan and passes it to **imwarp** function to 'fill' the outermost pixels or slices in the generated scan; this value was found to be similar to the mean intensity in the healthy contralateral hemisphere in these reference scans and the transformation preserves the intensity homogeneity within the brain across slices.

Figure 6.2 shows an example scan scaled with $s_x = 1$, $s_y = 1$ and $s_z = 1.1$ (simulates a scaling error of 10.0% in z direction). This scaling corresponds to 17.6 new slices based on the 16 slices in the reference scan; instead, **imwarp** generates 18 slices, the first and last of which are interpolated partially using the reference data and partially by filling with a specified intensity. The default filling intensity is zero, producing abnormally low pixel intensities in the first and last slices, and slightly higher pixel intensities in the second and penultimate slices compared to the rest of the slices (Figure 6.2 b); this is due to the bicubic interpolation method used in the transformation. Filling with an intensity similar to that in the healthy contralateral

brain hemisphere ensures a homogeneous intensity in the brain across slices, the profile of which is similar to that in the original scan (Figure 6.2 a, c). The intensity of any infarcted areas in the outermost pixels or slices is reduced accordingly, simulating representative contributions of PVE for each combination of scaling factors used.

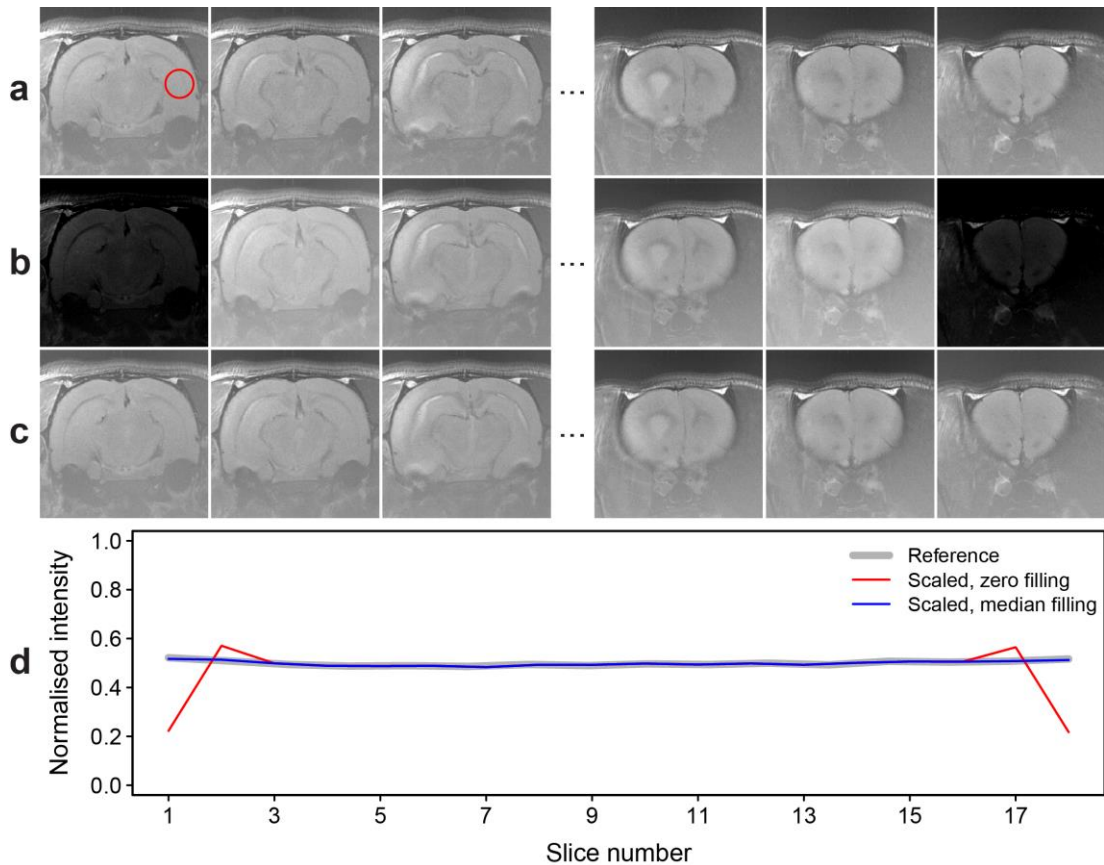


Figure 6.2 Scaling scans across the slice direction.

The three first and three last slices (caudal to rostral from left to right) in (a) an example bias field corrected reference scan are shown, as well as the scan scaled by filling outermost slices with (b) the default zero intensity and (c) the median intensity measured in the middle slice. The intensity in all images was scaled to the same range for comparison. Plot (d) shows the mean intensity across slices for all three scans, measured within a circle of diameter 3 mm in the cerebral cortex of the healthy contralateral hemisphere at a similar location in corresponding slices (for example, the red circle in the first slice in (a)). The slice number for the reference scan was scaled to the range 1–18 for direct comparison with the scaled scans.

Following transformation, the new scans are cropped by the developed MATLAB function to the required in-plane matrix size, which is 256×256 for all simulated imaging protocols (Table 6.2). If the new scans have a smaller matrix size, new rows and/or columns of pixels are added accordingly in each side to keep the brain centred in the FOV; the intensity chosen for these pixels is the minimum intensity in the reference scan. SNR differences were not modelled as the SNR in scans acquired using scanner 'B' and protocol 'b' is lower than that expected in scans from other centres (as suggested by phantom imaging), and the value in the uncorrected for bias field reference scans was found to be mean \pm SD = 26.4 \pm 3.8, based on 5 randomly selected scans (3 mm diameter circular ROI in infarct at the subcortex area; 19.2×2 mm² rectangular ROI in artefact-free background); SNR higher than 20 offers little advantage to image analysis (McRobbie et al., 2006). For the MSME sequence 'c' (Table 6.2), it was not possible to simulate quantitative T₂ maps and was assumed that the extent of the infarct in such images is identical to that on qualitative bias field corrected images; scans for this sequence were created using the method described above.

6.2.4.3 Estimation of scaling factors

The scaling factors in the transformation matrix (equation 6.1) are estimated by a series of calculations devised to combine the contribution of:

1. the linear scaling errors in scanner 'B' estimated using protocol 'b' (the scanner and protocol used to acquire the reference scans; Tables 6.1 and 6.2 respectively), and
2. the linear scaling errors in each other centre, and/or
3. the differences in voxel size between protocol 'b' and the rest of the protocols.

The reference scans must first be corrected for the presence of linear scaling errors in scanner 'B'; the percentage correction error in each direction is estimated using the formula

$$\% \text{ correction error}_{x,y,z} = \frac{100 - (100 + \bar{e}_{x,y,z})}{100 + \bar{e}_{x,y,z}} 100\% \quad (6.2)$$

where $\bar{e}_{x,y,z}$ is the median linear scaling error across each direction in scanner 'B', as measured in phantom scans. In x or y direction, the percentage difference between the pixel size for sequence 'b', $d_{x,y,b}$, and the required pixel size for each other sequence, $d_{x,y}$, is estimated as

$$\% \text{ difference}_{x,y} = \frac{d_{x,y,b} - d_{x,y}}{d_{x,y}} 100\% \quad (6.3)$$

The same percentage difference can be computed by using the FOV in each direction instead, as the matrix size is identical for all simulated scans. A similar formula is used to estimate the percentage difference between the two corresponding slice thicknesses in z direction, t_b and t :

$$\% \text{ difference}_z = \frac{t_b - t}{t} 100\% \quad (6.4)$$

The percentage correction errors and differences estimated using equations 6.2–6.4, as well as the median scaling errors in each scanner, are used to estimate three normalised scaling factors for each direction in each scanner as follows:

$$\text{scaling factor}_i = \frac{100 + \% \text{ error or difference}_i}{100} \quad (6.5)$$

where i indicates each of the three contributing components listed above (1 to 3). The three scaling factors are multiplied to estimate a pooled estimate for each orthogonal direction in each scanner:

$$s_{x,y,z} = \prod_{i=1}^3 \text{scaling factor}_i \quad (6.6)$$

s_x , s_y and s_z are used by the developed MATLAB function to perform scaling in all three directions using the transformation matrix defined previously (equation 6.1). The new pixel size, FOV and slice thickness can be estimated by dividing the relevant

parameter in sequence 'b' with the corresponding scaling factor. An example calculation of the pooled scaling factors in each direction for a simulated dataset is shown in Table 6.3. These factors were used to transform the reference dataset and simulate 21 scans for system 'A' and the standardised sequence 'a' (Tables 6.1 and 6.2 respectively), based on corresponding median linear scaling errors and voxel size differences between this sequence and sequence 'b'.

The scaled FOV and slice thickness based on the factors accounting for pixel and slice thickness differences are indeed those for sequence 'a' (19.2×19.2 mm² and 1 mm respectively); these are used for calculating infarct volume in this example dataset. The final pooled FOV and slice thickness represent the expected ones in the simulated scans, which are smaller than those specified by sequence 'a' due to the contribution of the positive linear scaling errors in this system. This hints that the brain and the infarct across all directions in the simulated scans will be larger than those in scans with the same voxel size but without the effect of linear scaling errors.

Table 6.3 Estimating pooled scaling factors for data simulation.

Direction in scanner	Contributing component	% error or difference	Scaling factor	New FOV or slice thickness (mm)
x	Scanner 'B'/sequence 'b' correction	-0.70	0.99	25.18
	Linear scaling error	4.47	1.04	23.93
	Pixel size difference	30.21	1.30	19.20
	Pooled estimate	35.08	1.35	18.51
y	Scanner 'B'/sequence 'b' correction	-1.45	0.99	25.37
	Linear scaling error	4.82	1.05	23.85
	Pixel size difference	30.21	1.30	19.20
	Pooled estimate	34.51	1.34	18.59
z	Scanner 'B'/sequence 'b' correction	-0.93	0.99	0.76
	Linear scaling error	4.47	1.04	0.72
	Slice thickness difference	-25.00	0.75	1.00
	Pooled estimate	-22.37	0.78	0.97

Correction errors are estimated using the errors in scanner 'B' for sequence 'b' (Figure 6.4 b). In this example, the linear scaling errors are the median errors in system 'A' for sequence 'a', as shown later in Figure 6.4 a. Pixel sizes and slice thicknesses to estimate percentage differences are listed in Table 6.2. The pooled scaling factors s_x , s_y and s_z are shown in bold; these were used to back calculate the corresponding pooled percentage error. The pooled new field of view (FOV) and slice thickness are the expected in the data due to scaling errors.

The procedure described above was repeated 18 times to simulate datasets of 21 scans and compare infarct volume measurements between all six preclinical centres (Table 6.1) for the following three cases:

1. Scanners are not calibrated and sequences 'a'–'f' are used at corresponding centres (both the linear scaling errors measured using each sequence and the differences in voxel size are simulated).
2. Scanners are not calibrated and the standardised sequence 'a' is used (only the linear scaling errors measured using sequence 'a' are simulated).
3. Scanners are calibrated and sequences 'a'–'f' are used at corresponding centres (only the differences in voxel size are simulated).

The factors shown in Table 6.3 for correcting the images for scaling in scanner 'B' and sequence 'b' were common in all simulations. The reference and all simulated datasets are freely available for download (<http://dx.doi.org/10.7488/ds/1496>).

6.2.5 Image analysis

6.2.5.1 Measurement of quality assurance criteria

In accordance with the stroke imaging protocol used in centre 'C', multi-echo images acquired using stroke sequence 'c' were pre-processed to create quantitative T₂ relaxation maps. For this purpose, *MRI Processor* plugin in ImageJ with default parameters was used; the plugin estimates T₂ relaxation using equation 4.1 in a voxel-wise fashion, based on the list of TE values of the MSME sequence given in Table 6.2 (Prodanov & Verstreken, 2012). Background noise in the generated maps was brighter than the phantom's liquid solution, therefore the intensity in the maps was inverted so that the contrast between the phantom and background was matched with that of images from all other centres and sequences.

According to the QA protocol and using the automated tool described in Chapter 4, the internal dimensions of the phantom were measured in scans in all imaging planes

and the volume of the phantom's cylindrical compartment was measured in axial scans. To examine the impact of SNR on volumetric accuracy in the between-scanner variability study, its value was measured in the axial scans acquired with the standardised sequence 'a' using the formula

$$SNR = 0.655 \frac{I_{phantom}}{SD_{noise}} \quad (6.7)$$

where $I_{phantom}$ is the mean intensity within a circular ROI measured at the centre of the phantom in the middle slice (3 mm diameter) and SD_{noise} the SD of the background noise measured in an artefact-free rectangular ROI in the same slice (19.2×2 mm²). The 0.655 factor is used to take into account the Rician distribution of the noise, which takes only positive values in these magnitude images which reduce the true SD.

6.2.5.2 Deformation mapping

To aid the interpretation of geometric errors in this multicentre study, the middle slices through the centre of the phantom in axial and coronal MRI scans were used to create 2D deformation maps quantifying the distance each pixel in the MRI images must move in order to recover the shape of the phantom in CT images or in baseline MRI scans; the latter was used for within-scanner variability images only. For this purpose, registration methods were used, as shown in Figure 6.3.

The binary 3D CT masks of the phantom, created as described previously in Chapter 4 (Section 4.2.2), were pre-processed to create the template images for registration. Initially, a sufficient number of the 34.5 µm thick slices around the centre of the phantom were selected to match the slice thickness in sequences 'a'-'f', either in the axial or the coronal plane; for example the middle 29 slices of the 3D CT scan were selected to create the template image for MRI sequences 'a', 'c' and 'd' that used 1 mm thick slices. The intensity of the slices was averaged in the slice direction to obtain a single greyscale image for each plane, which was then resized to match the in-plane resolution of the standardised sequence 'a' using bicubic interpolation (matrix size

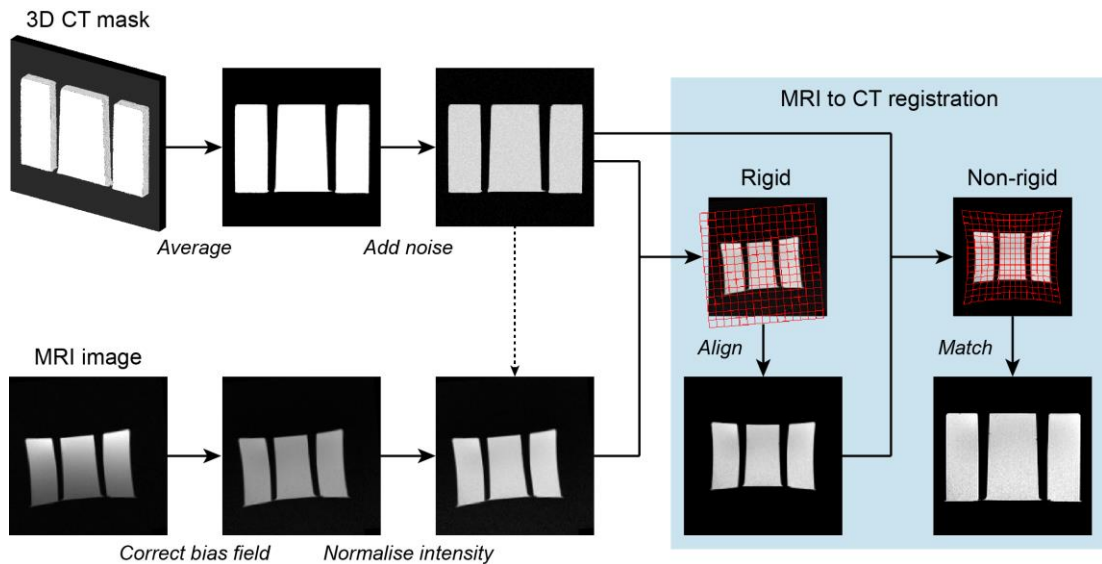


Figure 6.3 Generating 2D geometric deformation maps.

Binary 3-dimensional (3D) computed tomography (CT) masks are processed to create template images showing the true shape of the phantom, while MRI images are processed to match the intensity profile and homogeneity of the CT templates. MRI images (source images) are then registered to the CT templates (target images) using rigid and non-rigid transformations in order (blue box), which deform them until the true shape of the phantom is recovered. For the within-scanner variability study, the baseline MRI images taken post-gradient calibration were used as target images for registration. A prior transformation was applied to the original MRI image in the figure for demonstration.

256×256 pixels² and FOV 19.2×19.2 mm²). Low Rician noise was added to the averaged images as described before (Section 4.3.4) to obtain the final CT templates. The MRI slices were pre-processed for correcting for bias field using the N4 algorithm (Tustison & Gee, 2010) in 3D Slicer. Then, the intensity of MRI slices was normalised to that in the CT templates by multiplication. Images acquired using stroke sequences were resized to match the in-plane resolution of the CT templates using bicubic interpolation. For the within-scanner variability study, the baseline MRI images acquired soon after gradient calibration were used as template images instead of CT images, to observe changes in geometric distortion from baseline; these images were bias field corrected, while follow-up images were corrected and normalised, as described above. All pre-processing steps except bias field correction were carried out in ImageJ and MATLAB.

2D deformation maps were created using freely available plugins in ImageJ. A rigid

transformation (global translation and rotation) with the *StackReg* plugin was used to align MRI images over corresponding template images (Thevenaz, Ruttimann, & Unser, 1998). Then, a non-rigid b-splines transformation (local composite reshaping) with the *bUnwarpJ* plugin was used to precisely match the aligned MRI images over the templates (Arganda-Carreras et al., 2006). This tool produces a text file containing the new horizontal and vertical coordinates of each pixel of the MRI image; each value is the sum of the original coordinate and the horizontal or vertical displacement induced on that pixel. Code was written in MATLAB to read the coordinates in the text file and compute the pixel-wise Euclidean distance, $d(x, y)$, according to

$$d(x, y) = \sqrt{(\text{xdim}(x_{new} - x_{old}))^2 + (\text{ydim}(y_{new} - y_{old}))^2} \quad (6.8)$$

where x_{old} and y_{old} are the original pixel coordinates (0 to 255), x_{new} and y_{new} the new pixel coordinates, and xdim and ydim the size of the pixel in each direction (75 μm for sequence 'a'). The MATLAB code also generates colour maps with $d(x, y)$ values to visualise geometric distortion. Daily, weekly and monthly pixel-wise mean and SD maps were created to combine the deformation maps in the longitudinal study.

6.2.5.3 Measurement of infarct volume in simulated data

The freehand selection tool in ImageJ was used to manually trace the ipsilateral and contralateral brain hemispheres in all generated scans, and then the developed semi-automated threshold-based MATLAB tool was used to segment the infarct (Section 5.2.2). Analysis was performed by myself, whose repeatability in the use of the tool was found to be excellent (Figures 5.10 and 5.11). Scans within each dataset were analysed randomly, starting with the dataset with the lowest in-plane resolution for cases 1 and 3 where different sequences are simulated (Nieuwstadt et al., 2015). Random selection was performed using the `RANDBETWEEN` function in Microsoft Excel, while image analysis was guided by a commonly used brain atlas (Paxinos & Watson, 1997). For each dataset, the volume of each ROI was estimated by

multiplying the number of selected voxels with the corresponding voxel size for each simulated sequence, as given in Table 6.2.

6.2.6 Statistical analysis

Measurements of phantom dimensions and volumes were compared with corresponding ground truth values (Section 4.2.2) to estimate percentage linear scaling and volumetric errors, according to

$$\% \text{ error} = \frac{\text{measured} - \text{true}}{\text{true}} 100\% \quad (6.9)$$

From each set of six scans, 12 measurements in total were obtained for the phantom dimensions, corresponding to four measurements per orthogonal direction; errors were summarised using median values and ranges where appropriate. To identify imaging parameters that might have an impact on volumetric accuracy, the measured volumetric errors in the between-scanner variability study were compared with corresponding errors predicted by linear scaling errors alone. For this purpose, the median scaling errors across each direction, \bar{e}_x , \bar{e}_y and \bar{e}_z , were used to estimate the predicted volumetric errors, according to

$$\% \text{ predicted error}_v = \frac{(\bar{e}_x + 100)(\bar{e}_y + 100)(\bar{e}_z + 100)}{10000} - 100\% \quad (6.10)$$

The distributions of linear scaling errors across the two encoding directions in the within-scanner variability study were assessed for normality using the Shapiro–Wilk test; measurements were not normally distributed (Table C.1), therefore a nonparametric Mann–Whitney U test was used to examine their difference. Statistical comparisons between centres, sequences or time points were not performed due to the small number of measurements for each.

In the simulation study, V_{IH} and V_{CH} were used to correct infarct volume for vasogenic oedema using equation 5.10 and estimate V_{I_c} . V_{I_c} was also expressed as the percentage of V_{IH} , V_{CH} or the volume of the whole brain ($V_{IH} + V_{CH}$); these are relative measures

of infarct size that are often reported and statistically compared between groups in published articles (Table B.3). Hemispheric and whole brain volumes for oedema correction and estimation of relative infarct volume were estimated either based on all acquired slices through the MCA territory, or only the slices which included infarcted tissue to determine the impact of each approach.

To examine the overall comparability of imaging across centres for each of the three simulated cases, measurements were compared between centres using repeated measures ANOVA with a Greenhouse–Geisser correction where sphericity was violated (Mauchly's test; Table C.2) and Bonferroni's post hoc tests to correct for multiple comparisons (repeated-measure factor is the centre/scanner). The ANOVA output included η^2 (partial eta squared), which is an effect size indicating the proportion of variance in measurements explained by variation between scanners. Distributions of measurements for all simulated datasets were normally distributed, as determined by the Shapiro–Wilk test for normality (Table C.1), thus were summarised using the mean value.

Using the measurements in the two animal groups (male and female rats), statistical comparisons were performed to determine the impact of MRI variability in infarct volume-based study effects across centres. The two groups in each simulated dataset were compared using an exact two-tailed Mann–Whitney U test to account for their small size, and were summarised using the median and its 95% CI estimated using 10000 bootstrap samples. The percentage median difference in infarct volume between groups was also estimated using the formula

$$\% \text{ median difference} = \frac{\text{median}_m - \text{median}_f}{1/2 (\text{median}_m + \text{median}_f)} 100\% \quad (6.11)$$

where median_m and median_f are the median infarct volumes in male and female rats in each dataset respectively.

All statistical tests were performed in SPSS. A p -value less than 0.05 was considered statistically significant.

6.3 Results

6.3.1 Between-scanner variability

Geometric errors in scanners are shown in Figure 6.4. The majority of linear scaling errors were positive; the median error was negative only in z direction in scanner 'D' when the stroke sequence was used. Overall, geometric accuracy in stroke sequences was not worse than the accuracy in the standardised sequence, but measurements with the former were more dispersed in the majority of scanners. In systems 'B'–'F', median errors per direction were within 0.03% to 1.84% for the standardised sequence (Figure 6.4 a) and within -0.27% to 2.19% for the stroke sequences (Figure 6.4 b). Scanner 'A' was assessed using the standardised sequence 'a' alone and was found that it overestimated the phantom's dimensions by a median error of 4.47% across x and z directions and 4.82% across y direction. In all systems except 'A', a consistent difference in scaling errors between frequency and phase encoding directions was not observed; errors in scanner 'A' were larger in frequency encoding in all three directions, but the number of measurements was too small to verify this statistically.

Correspondingly, volumetric errors were positive for both standardised and stroke sequences in all scanners. With the standardised sequence, system 'A' overestimated the volume by 18.73%, while the error ranged from 5.71% to 11.67% in the rest of the scanners (values for 'B' and 'F' respectively; Figure 6.4 c). Stroke sequences had a better overall volumetric accuracy compared to the standardised sequence (median error 6.19% versus 9.02% in all scanners 'B'–'F'), but errors were dispersed between 3.96% and 12.51% (values for 'E' and 'C' respectively; Figure 6.4 c). Measured volumes were higher than corresponding predicted volumes for all systems and sequences, with percentage differences between the two estimates ranging from 2.19% to 10.10% for the standardised sequence (scanners 'B' and 'F' respectively) and 2.49% to 10.53% for stroke sequences (scanners 'E' and 'D' respectively; Figure 6.4 c).

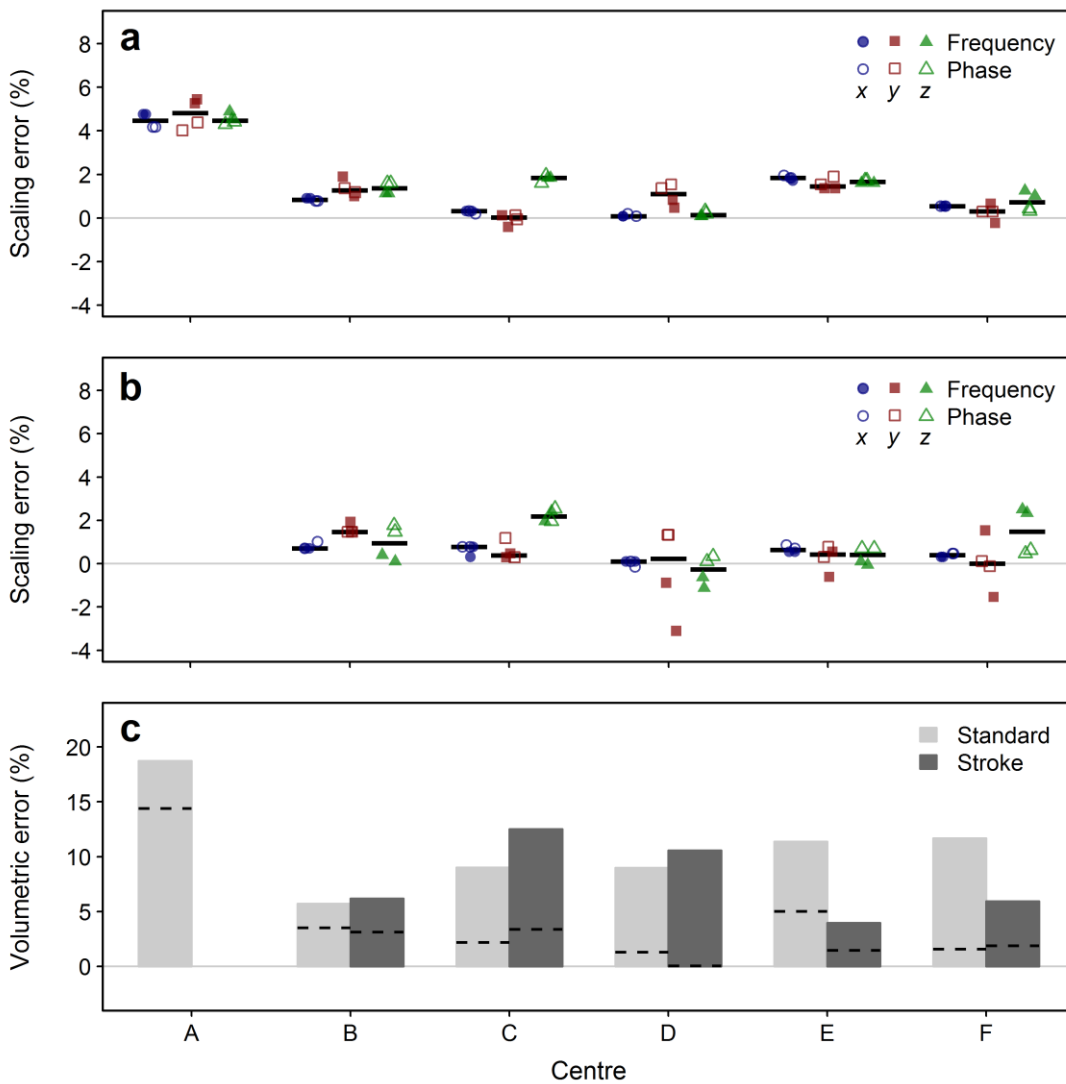


Figure 6.4 Between-scanner variability in scaling and volumetric error. Geometric accuracy in six different scanners was evaluated by measuring percentage linear scaling errors across the three orthogonal directions with (a) the standardised sequence 'a' in all scanners and (b) the stroke sequences 'b'-'f' in corresponding scanners 'B'-'F', with median values shown by the thick black lines. Plot (c) shows the percentage volumetric error with values predicted by the median scaling errors indicated by the dashed lines. Scaling errors were measured in scans in all imaging planes and volumetric errors in the axial scans alone.

The difference between measured and predicted volumetric errors suggests that factors other than simple linear scaling errors may contribute to volumetric accuracy; Figure 6.5 shows a positive association between this error difference and SNR or slice thickness. For example, when the standardised sequence is used the difference is small for scanner 'B' and high for scanner 'F', images of which had the highest and lowest SNR respectively. Considering the different stroke sequences alone, a large

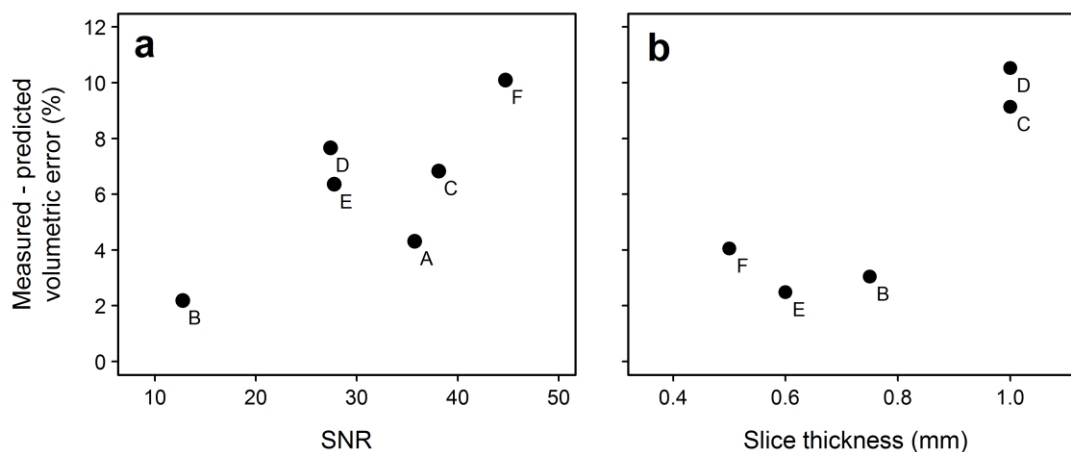


Figure 6.5 Influence of SNR and slice thickness in volumetric errors. Plots of the difference in measured and predicted volumetric error against (a) signal-to-noise ratio (SNR) and (b) slice thickness are shown. Errors were measured in axial scans acquired with (a) the standardised sequence and (b) the stroke sequences in six and five centres respectively, indicated next to each point. Points are too few to reliably determine the type of association between variables.

slice thickness led to an increased difference between measured and predicted volumes (sequences 'c' and 'd' with a thickness of 1 mm).

Figure 6.6 shows sample axial and coronal images through the centre of the phantom acquired at all centres using the standardised and stroke sequences. Images are characterised by various forms of bias field produced by RF inhomogeneity (for example, two-channel versus four-channel phased array surface coil in centres 'A' and 'B' respectively), Gibbs ringing artefact across phase encoding (most prevalent in images from centre 'D') and some geometric distortion. To visualise the geometric distortions more clearly, MRI images were used to create colour maps showing the Euclidean displacement per pixel to deform MRI images to match corresponding CT images of the phantom (Figure 6.7). Both raw images and deformation maps show two different patterns of geometric distortion in participating scanners; images from scanner 'A' are expanded in an isotropic way, whereas images from the rest of the scanners are characterised by minor nonlinearities that lie across phase encoding (a line across this direction would appear bent). Distortion effects were not only more prevalent but were also more variable in scans acquired using stroke sequences, in agreement with linear scaling and volumetric errors in Figure 6.4.

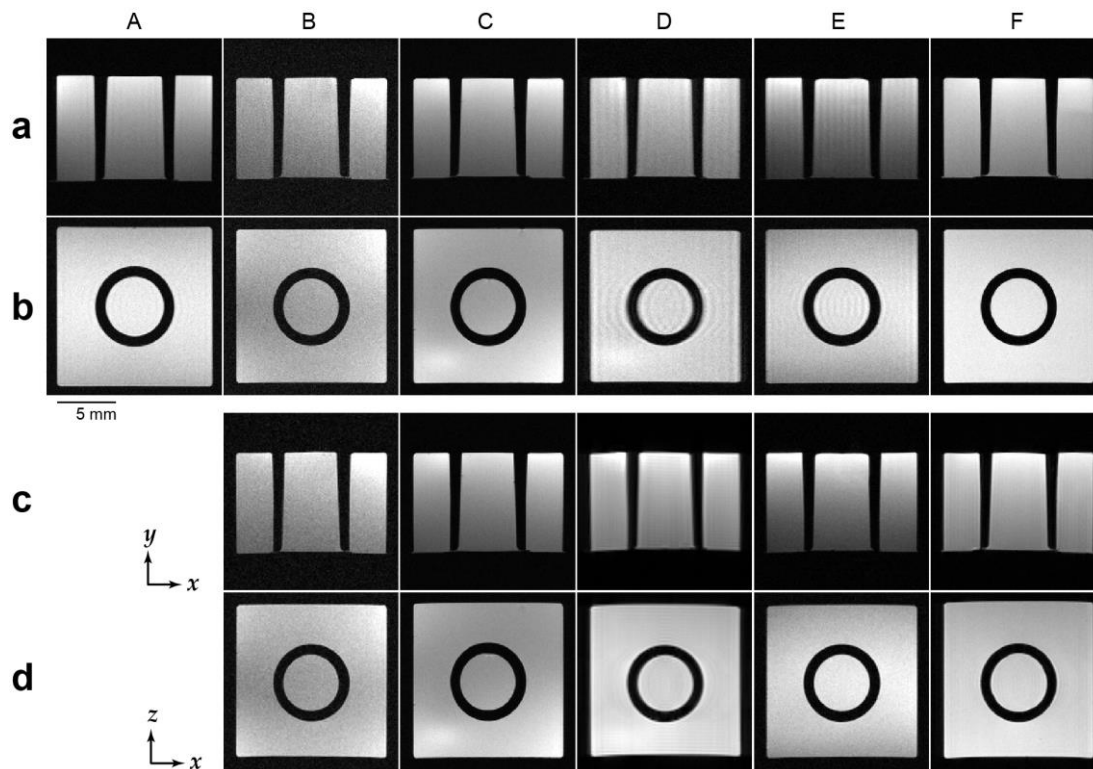


Figure 6.6 Sample images acquired using six different MRI scanners. These are slices through the centre of the phantom in (a, c) the axial (x - y) and (b, d) the coronal (x - z) plane acquired using (a, b) the standardised and (c, d) the stroke sequences. 'A'-'F' at the top of the figure indicate the scanner; 'A' was evaluated only with the standardised sequence 'a'. Phase encoding is in the horizontal direction (x) in both planes. Images taken using the stroke sequences for scanners 'B', 'D', 'E', and 'F' were scaled to match the in-plane resolution of the standardised sequence for direct comparison. Corresponding colour maps visualising the deformation required to recover the true shape of the phantom are shown in Figure 6.7.

Considering only the area covered by the phantom, the Euclidean displacement induced to MRI images to match corresponding CT images using non-rigid registration ranged from 0 to 0.45 mm; the maximum value was observed in the axial image taken using scanner 'A'. In images acquired in centres 'B'-'F', the maximum displacement was 0.22 mm with the standardised sequence and 0.38 mm with stroke sequences, both in the coronal plane (scanners 'E' and 'F' respectively; Figure 6.7 b, d).

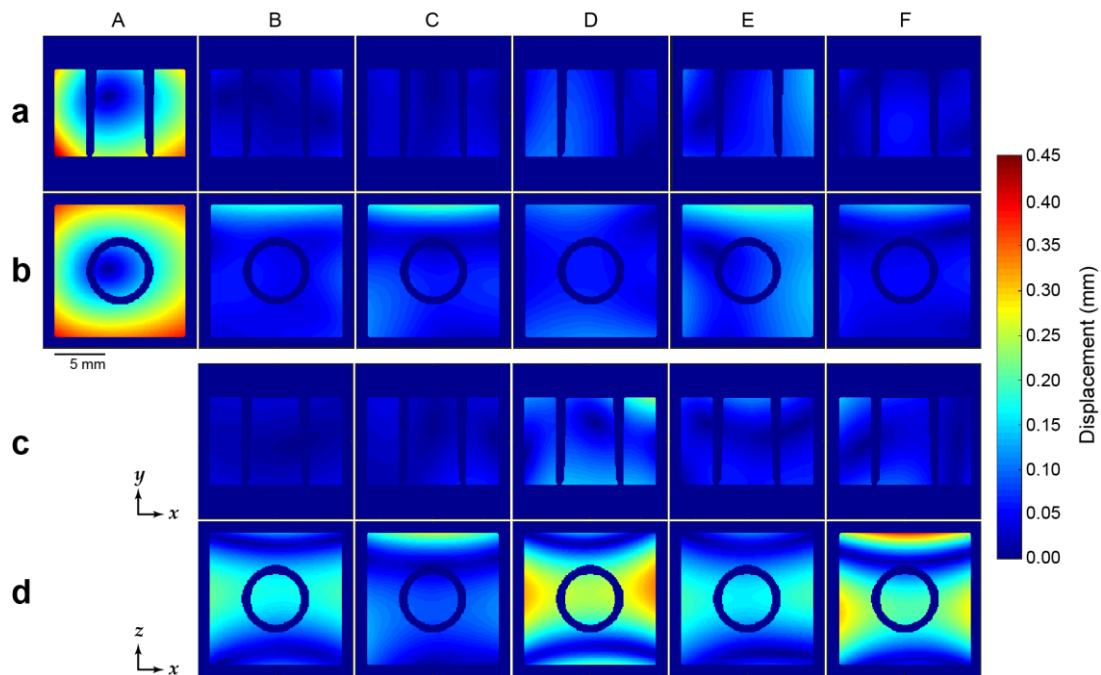


Figure 6.7 Deformation in images acquired with six different MRI scanners. The 2D maps show the in-plane Euclidean displacement applied to (a, c) axial (x - y) and (b, d) coronal (x - z) MRI images acquired using (a, b) the standardised and (c, d) the stroke sequences, to recover the true shape of the phantom in template CT scans. 'A'-'F' at the top of the figure indicate the scanner; 'A' was evaluated only with the standardised sequence 'a'. Phase encoding is in the horizontal direction (x) in both planes. Corresponding MRI images are shown in Figure 6.6.

6.3.2 Within-scanner variability

The gradient coils in system 'A' that performed poorly were calibrated based on measurements shown in Figure 6.4; then, drifts in its performance were recorded longitudinally over six months. Figure 6.8 shows plots of linear scaling and volumetric errors in this system. At baseline, median scaling errors were reduced to -0.33%, -0.50% and 0.25% in x , y and z direction respectively ('d0' time point in Figure 6.8 a). These figures did not change significantly during the subsequent longitudinal assessment, with median errors remaining within $\pm 1.00\%$ in all three directions over time. The apparent size of the phantom in the MRI images fluctuated similarly during daily, weekly and monthly imaging, with the exception of a small but apparent drift in the performance of the y gradient from the second to the last month (median -0.41% to -0.86% at 'm2' to 'm6' time points respectively; Figure 6.8 a).

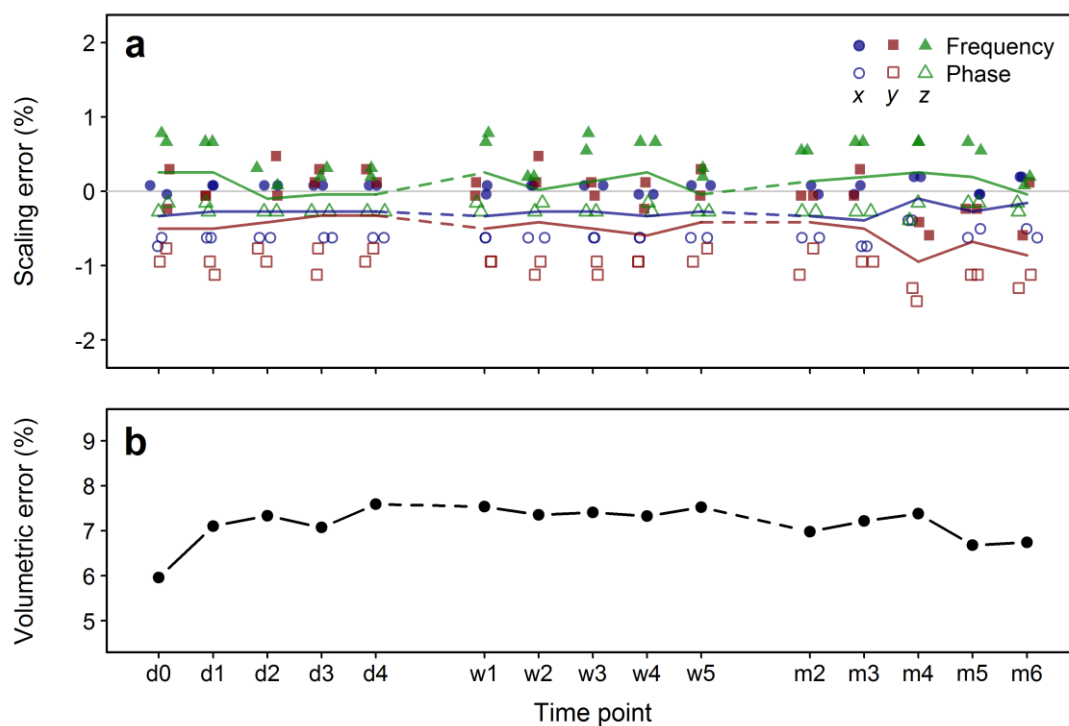


Figure 6.8 Within-scanner variability in scaling and volumetric error. Scanner ‘A’ and the standardised sequence ‘a’ were used to acquire images after gradient calibration and measure (a) percentage linear scaling errors across the three orthogonal directions and (b) percentage volumetric errors. Scaling errors were measured in scans in all imaging planes and volumetric errors in axial scans alone. The coloured lines in (a) follow the median values at each time point. d indicates day; w, week; m, month. ‘d0’ is the baseline time point immediately after calibration.

The volume of the phantom’s cylindrical compartment was overestimated at all time points (Figure 6.8 b). Across the three time intervals, volumetric error varied differently, but the overall variability across the whole six-month period was very small compared to between-scanner variability (median = 7.33%, IQR = 0.43%).

Considering all data in the within-scanner variability study, it was found that linear scaling errors across the frequency encoding direction were significantly larger than errors across phase encoding in scanner ‘A’ (frequency: median = 0.12%, IQR = 0.35%; phase: median = -0.63%, IQR = 0.67%; Mann–Whitney U test: $U = 136$, $p < 0.001$).

Images acquired during the three time intervals were used to create colour maps showing the Euclidean displacement per pixel that deforms the phantom in these images to match the phantom in the baseline images (‘d0’ time point in Figure 6.8);

pixel-wise mean and SD maps were created using the maps in each time interval (Figure 6.9). In agreement with previous observations, the mean displacement was very small in each time interval, with maximum displacement over the phantom area being 0.06 mm, 0.07 mm and 0.11 mm in daily, weekly and monthly scans respectively. However, SD deformation maps revealed an increasing variance in geometric accuracy from short-term to long-term imaging, particularly in y direction (Figure 6.9 c), in agreement with the heterogeneity observed in linear scaling errors across the same direction (Figure 6.8 a).

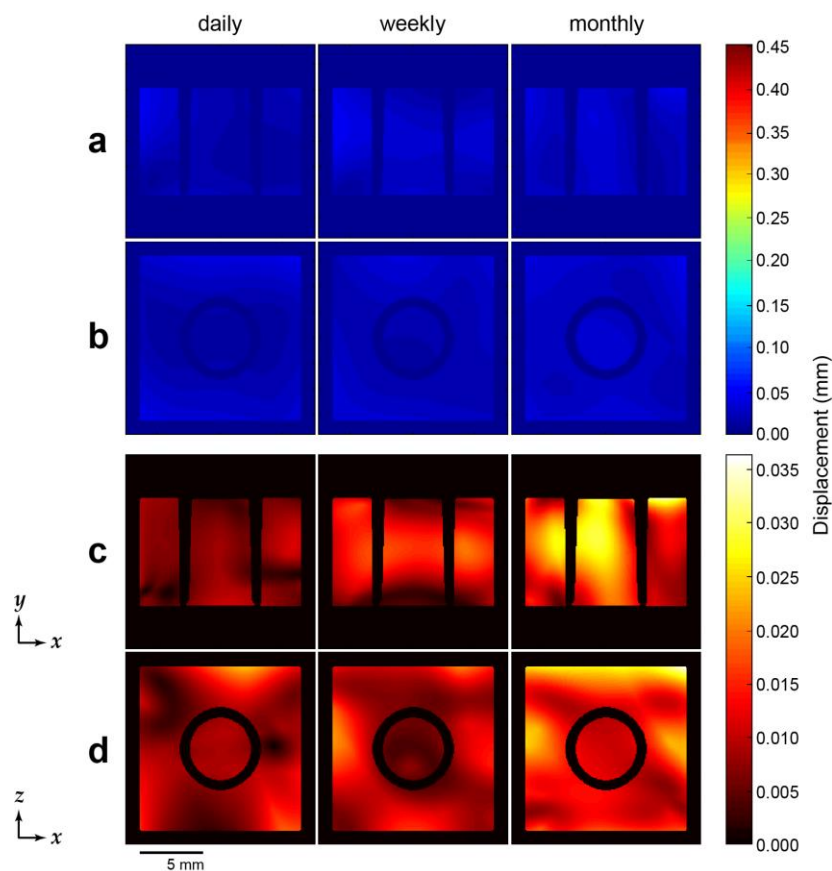


Figure 6.9 Deformation in repeated MRI.

The 2D maps show the pixel-wise (a, b) mean and (c, d) SD of the in-plane Euclidean displacement applied to (a, c) axial (x - y) and (b, d) coronal (x - z) MRI images acquired over time to recover the shape of the phantom in baseline scans ('d0' time point in Figure 6.8). As indicated at the top of the figure, images acquired 'daily' (four scans), 'weekly' (five scans) and 'monthly' (five scans) were used to create these maps. The colour bar of the mean displacement was scaled to that in Figure 6.7. Phase encoding is in the horizontal direction (x) in both planes.

6.3.3 Post hoc imaging

6.3.3.1 Comparison of imaging coils

In this multicentre study, centre 'E' used a mouse head surface coil for imaging, which expectedly produced images with a more severe bias field compared to other scanners that used rat head surface coils, at least along the y direction (Figure 6.6). To determine whether this effect could have affected measurements of QA criteria, both types of surface coils were used to scan the phantom in centre 'A' consecutively. Median scaling errors for the rat head surface coil were -0.10%, -0.86% and 0.25% along the x , y and z direction respectively; similarly, for the mouse head surface coil they were -0.10%, -0.95% and 0.25% respectively. Volumetric errors were 8.15% and 7.40% for the rat and mouse head surface coil respectively. Figure 6.10 shows sample images and corresponding deformation maps indicating that there is only a small difference in distortion effects produced by the rat and mouse head surface coils, with maximum displacement being 0.08 mm and 0.09 mm in images acquired with each, respectively. These observations suggest that these coils can be used interchangeably.

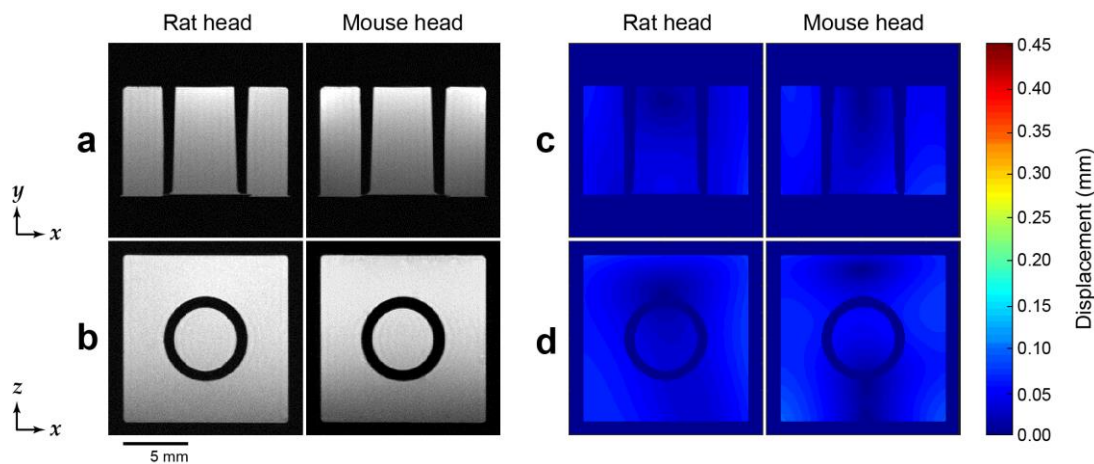


Figure 6.10 Effect of imaging coil on geometric distortion.

The figure shows sample slices through the phantom in (a) the axial (x - y) and (b) coronal (x - z) planes acquired using two different surface RF coils in centre 'A', as indicated at the top. Corresponding 2D deformation maps showing the displacement needed to recover the true shape of the phantom in template CT scans are given in (c) the axial and (d) the coronal plane. The colour bar of the displacement was scaled to that in Figure 6.7. Phase encoding is in the horizontal direction (x) in both planes.

6.3.3.2 Influence of magnetic susceptibility

Sequences 'a', 'b' (Table 6.2) and a modified version of the latter ('b_{modified}') which theoretically amplifies magnetic susceptibility effects were used to scan the phantom in centre 'A' consecutively and determine whether nonlinearity effects similar to those in systems 'B'-'F' (Figure 6.7) can be observed in 'A' as well. Sample images and corresponding deformation maps are given in Figure 6.11.

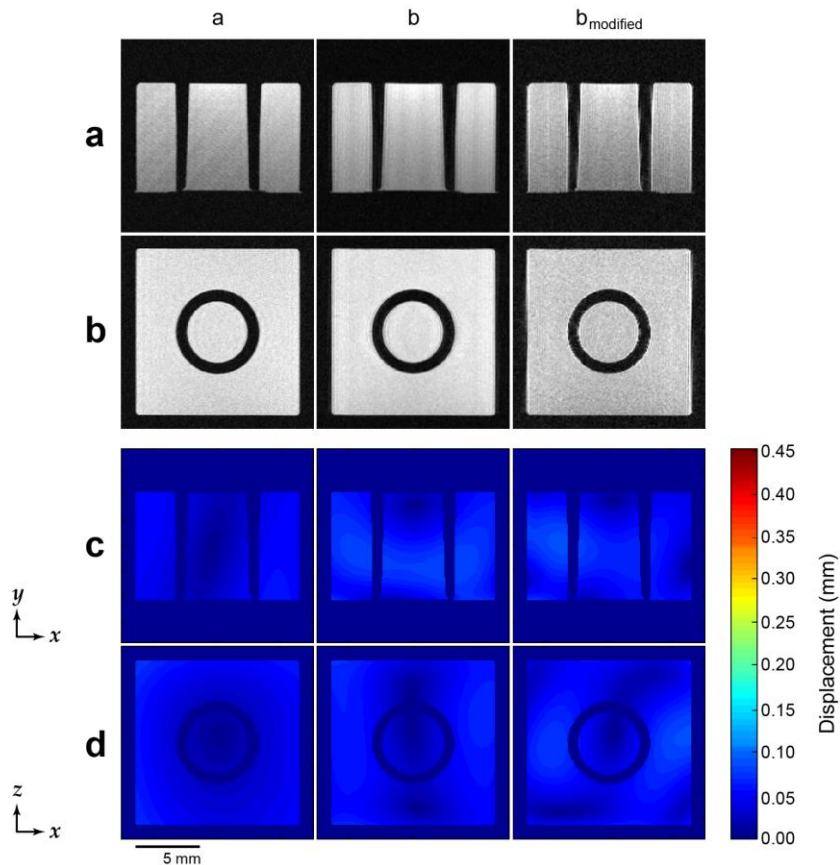


Figure 6.11 Magnetic susceptibility effects on geometric distortion.

The figure shows sample slices through the phantom in (a) the axial (x - y) and (b) coronal (x - z) planes acquired using three different sequences in centre 'A', as indicated at the top. Corresponding 2D deformation maps showing the displacement needed to recover the true shape of the phantom in template CT scans are given in (c) the axial and (d) the coronal plane. The colour bar of the displacement was scaled to that in Figure 6.7. Phase encoding is in the horizontal direction (x) in both planes. Images taken using sequences 'b' and 'b_{modified}' were scaled to match the in-plane resolution of the standardised sequence 'a' for direct comparison. These MRI images have improved intensity uniformity compared to images from the same scanner shown in Figures 6.6 and 6.10 as they were taken following maintenance of the system and imaging coils.

Apart from some minute geometric variation that could arise from the difference in noise levels in acquired images, neither ‘b’ nor ‘b_{modified}’ produced nonlinearities in this system similar to those in scanners ‘B’–‘F’; overall distortion was similar to that in images taken using the standardised sequence ‘a’. The maximum displacement was 0.08 mm, 0.08 mm and 0.09 mm in maps for sequence ‘a’, ‘b’ and ‘b_{modified}’ respectively.

6.3.4 Simulations

6.3.4.1 Scaling factors

Table 6.4 lists all the pooled scaling factors used for data simulation. As shown in the table, the reference scans acquired using the uncalibrated scanner ‘B’ and sequence ‘b’ were used in the first simulation case. For the second case, scaling factors across scanners were similar, as the standardised sequence ‘a’ was modelled for each system (Figure 6.4 a). Factors across scanners differed significantly in cases 1 and 3 due to the use of sequences with differing voxel size. Datasets for scanners ‘A’ and ‘C’ in case 3 were identical due to the use of sequences with the same voxel size.

Table 6.4 Pooled scaling factors for simulating datasets.

Case	Direction in scanner	A	B	C	D	E	F
1	<i>x</i>	1.35	1.00	1.30	0.62	1.00	0.97
	<i>y</i>	1.34	1.00	1.29	0.62	0.99	0.96
	<i>z</i>	0.78	1.00	0.76	0.74	1.24	1.51
2	<i>x</i>	1.35	1.30	1.30	1.29	1.32	1.30
	<i>y</i>	1.34	1.30	1.28	1.30	1.30	1.29
	<i>z</i>	0.78	0.75	0.76	0.74	0.75	0.75
3	<i>x</i>	1.29	0.99	1.29	0.62	0.99	0.97
	<i>y</i>	1.28	0.99	1.28	0.62	0.99	0.96
	<i>z</i>	0.74	0.99	0.74	0.74	1.24	1.49

These factors are used to simulate scans for cases where (1) scanners are not calibrated and sequences ‘a’–‘f’ are used at corresponding centres, (2) scanners are not calibrated and the standardised sequence ‘a’ is used, and (3) scanners are calibrated and sequences ‘a’–‘f’ are used at corresponding centres. ‘A’–‘F’ indicate the scanner/centre.

An example reference scan and corresponding produced scans for cases 2 and 3 for scanner 'A' are shown in Figure 6.12. The brain and infarct appear slightly larger in the scan incorporating the large scaling errors in the uncalibrated system 'A', compared to those in the scan that does not simulate scaling errors (Figure 6.12 b and c respectively); at the same time, scans are comprised by 13 and 12 slices respectively, leading to marked difference in the number of voxels within the infarct.

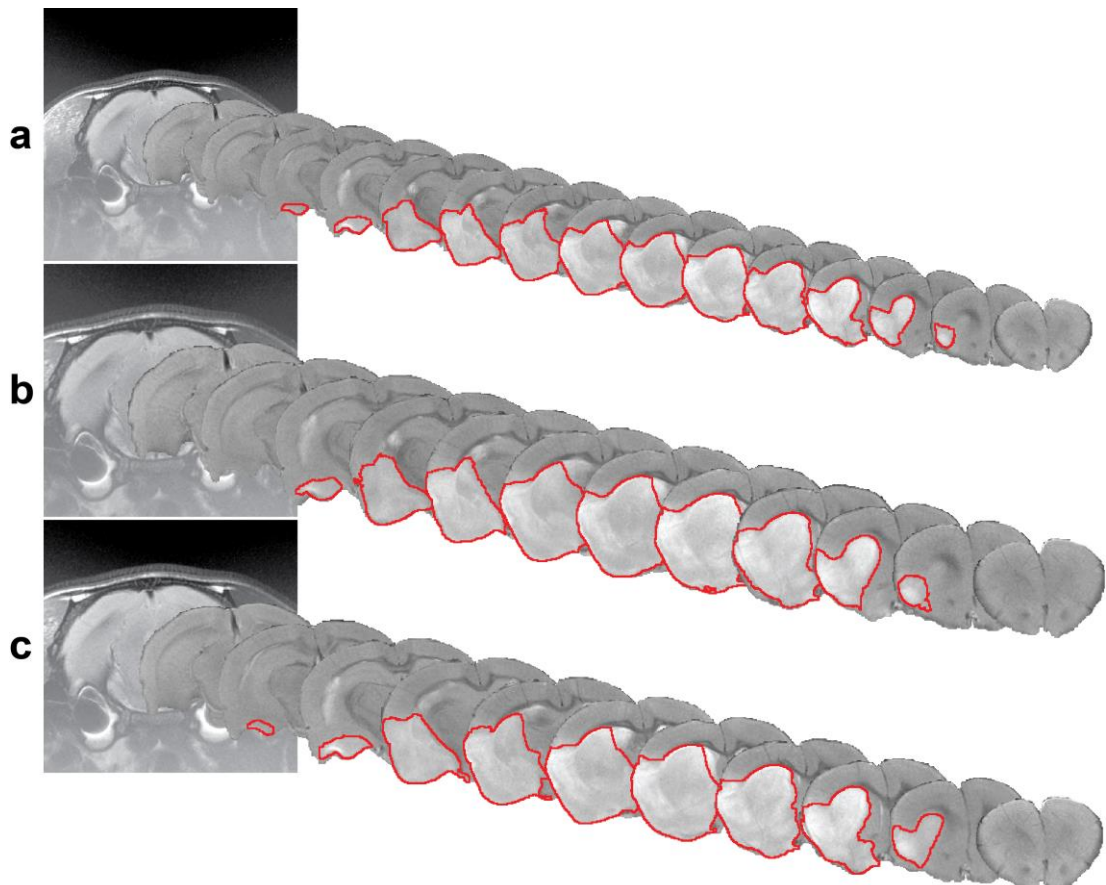


Figure 6.12 Reference and scaled datasets.

The brain and infarct in all slices (caudal to rostral from left to right) in (a) an example bias field corrected reference scan are shown, as well as in corresponding scaled scans simulating (b) errors in the uncalibrated scanner 'A' while the standardised sequence 'a' is used, and (c) no errors in the calibrated scanner 'A' while the standardised sequence 'a' is used. The produced true field of view in each scan is shown on the left. The infarct defined using the semi-automated tool is shown in red.

6.3.4.2 Infarct volume between scanners

Table 6.5 shows the results of repeated measures ANOVA between scanners when infarct volume is expressed as an absolute or relative estimate for all three simulation cases. Infarct volume was corrected using V_{IH} and V_{CH} (equation 5.10), calculated according to two different approaches reported in literature: using all slices covering the MCA territory (standardised across animals), or only the slices including the infarct (differing across animals). Then, V_{Ic} was expressed as the percentage of V_{IH} , V_{CH} or the volume of the whole brain. The results of pairwise comparisons using Bonferroni's post hoc tests are given in Table C.3 and are indicated in Figures 6.13 and 6.14.

Infarct volume measures differ significantly between scanners, even when the same sequence with uncalibrated scanners (case 2) or different sequences with perfectly calibrated scanners are used (case 3); the only exception was for V_{Ic} expressed as a percentage of V_{CH} measured in slices including the infarct for case 3 ($p = 0.059$). Between-scanner variability was often lower for case 3 compared to the other two in all comparisons, and post hoc tests sometimes failed to detect any significant difference between scanners even when ANOVA did (Figures 6.13 and 6.14). Despite a better concordance between scanners when the infarct volume was corrected for ipsilateral swelling and expressed as a relative estimate based on MRI slices including the infarct (p -values ranged from 0.003 to 0.059), the proportion of variance in measurements explained by variation between scanners was small (η^2 ranged from 0.12 to 0.16). In contrast, variation between scanners explained a significant portion of the overall measurement variance when infarct volume was corrected and expressed as a relative estimate based on all slices (η^2 ranged from 0.42 to 0.67). The use of uncalibrated scanners was the main contributor to the overall measurement variance (cases 1 and 2). No significant differences existed between scanners 'A' and 'C' in case 3 (using the same voxel size) for any of the examined infarct volume measures.

Table 6.5 Repeated measures analyses of variance between scanners.

Infarct volume measure	Case	df	F	η^2	p-value
Uncorrected, in mm ³	1	1.97	129.83	0.87	< 0.001
	2	1.75	120.26	0.86	< 0.001
	3	3.37	3.71	0.16	0.012
Corrected based on all slices					
in mm ³	1	2.77	52.79	0.72	< 0.001
	2	2.17	59.77	0.75	< 0.001
	3	2.89	12.50	0.38	< 0.001
as % of brain	1	5.00	15.30	0.43	< 0.001
	2	2.45	40.40	0.67	< 0.001
	3	2.83	17.11	0.46	< 0.001
as % of ipsilateral hemisphere	1	5.00	14.59	0.42	< 0.001
	2	2.47	39.73	0.66	< 0.001
	3	2.64	16.99	0.46	< 0.001
as % of contralateral hemisphere	1	5.00	16.15	0.45	< 0.001
	2	2.41	41.08	0.67	< 0.001
	3	3.01	16.82	0.46	< 0.001
Corrected based on slices including the infarct					
in mm ³	1	2.80	60.71	0.75	< 0.001
	2	2.47	69.17	0.78	< 0.001
	3	2.55	9.95	0.33	< 0.001
as % of brain	1	5.00	3.79	0.16	0.004
	2	2.24	3.54	0.15	0.033
	3	2.62	3.28	0.14	0.033
as % of ipsilateral hemisphere	1	5.00	3.88	0.16	0.003
	2	2.25	3.56	0.15	0.032
	3	2.53	3.76	0.16	0.022
as % of contralateral hemisphere	1	5.00	3.67	0.15	0.004
	2	2.24	3.50	0.15	0.034
	3	2.74	2.71	0.12	<i>0.059</i>

Case refers to simulations of scans when (1) scanners are not calibrated and sequences 'a'–'f' are used at corresponding centres, (2) scanners are not calibrated and the standardised sequence 'a' is used, and (3) scanners are calibrated and sequences 'a'–'f' are used at corresponding centres. Where $df = 5$, sphericity was assumed, otherwise the Greenhouse–Geisser correction was applied. p -values less than 0.01 are shown in bold and above 0.05 in italic. Results of corresponding pairwise comparisons using Bonferroni's post hoc tests are given in Table C.3 and are indicated in Figures 6.13 and 6.14. η^2 (partial eta squared) is the effect size indicating the proportion of variance in measurements explained by variation between scanners. df indicates degrees of freedom.

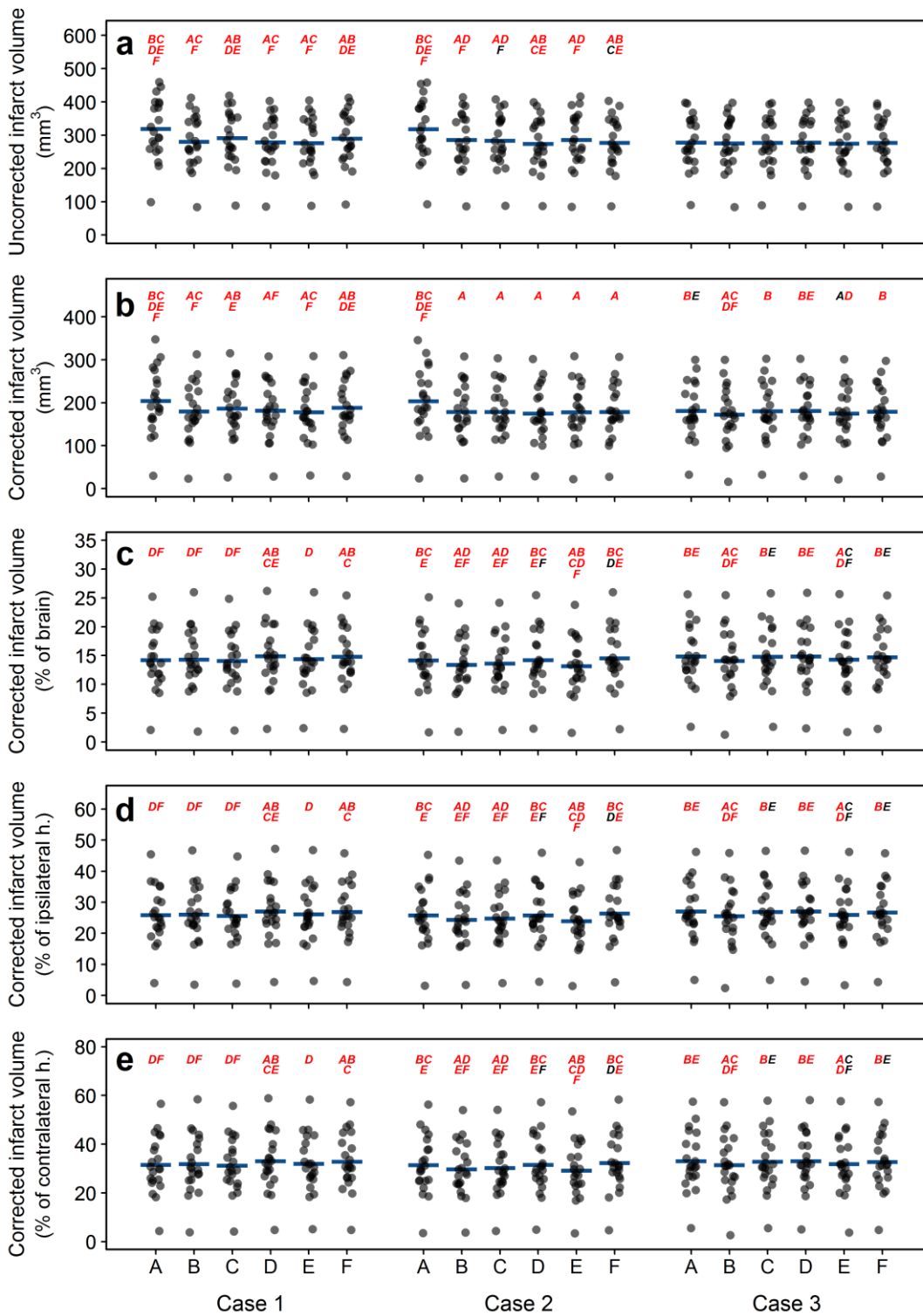


Figure 6.13 Infarct volume corrected based on all slices and expressed as an absolute or relative estimate.

Measurements for all three cases for scanners ‘A’–‘F’ as (a) an absolute uncorrected and (b) an absolute corrected infarct volume are shown, as well as the latter expressed as (c) the percentage of the brain, (d) the ipsilateral and (e) the contralateral hemisphere (h.). The thick blue line for each group of measurements indicates the mean value. Scanners with which there are significant pairwise differences are indicated above each distribution; red labels indicate $p < 0.01$.

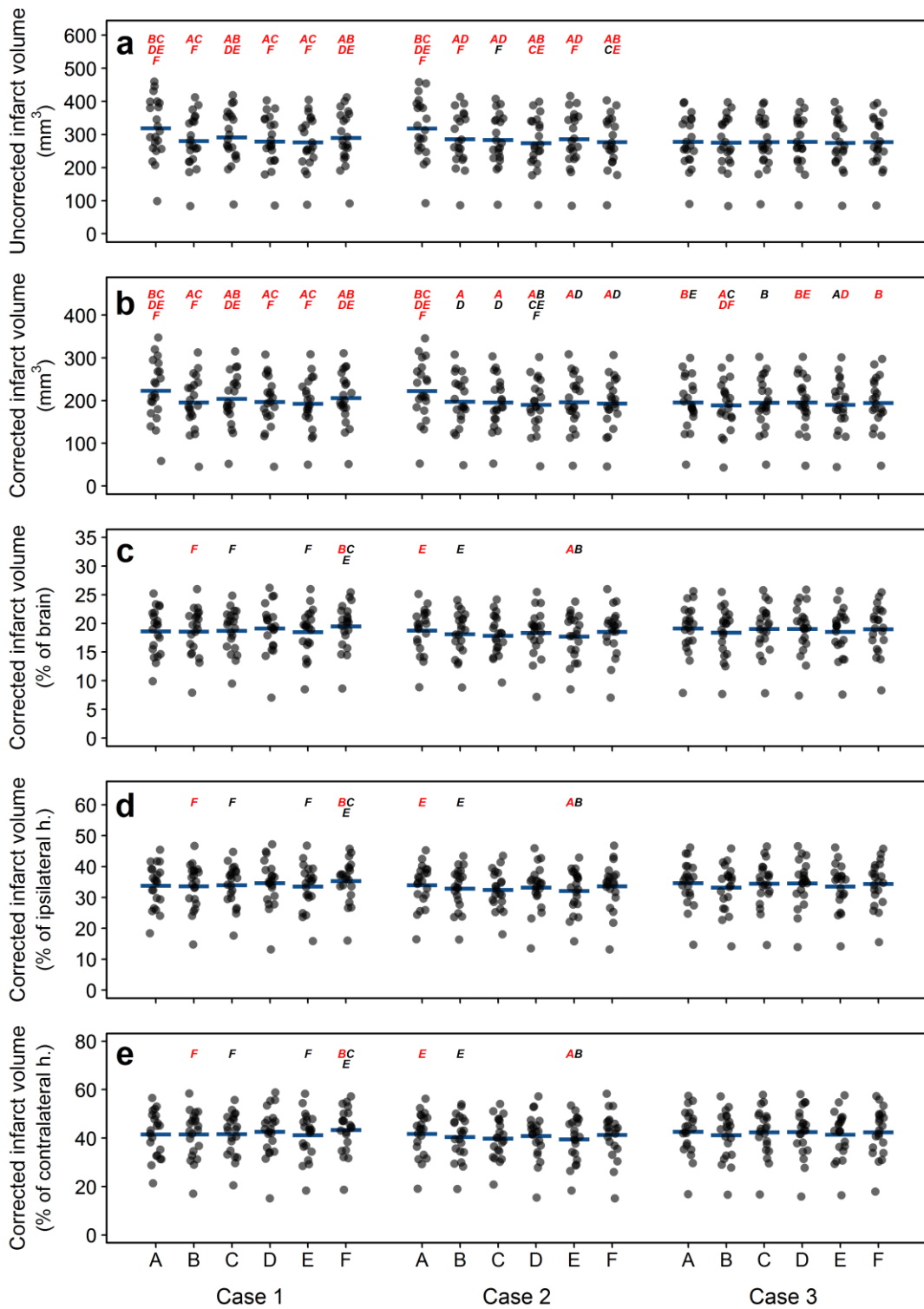


Figure 6.14 Infarct volume corrected based on slices including the infarct and expressed as an absolute or relative estimate. Measurements for all three cases for scanners 'A'–'F' as (a) an absolute uncorrected and (b) an absolute corrected infarct volume are shown, as well as the latter expressed as (c) the percentage of the brain, (d) the ipsilateral and (e) the contralateral hemisphere (h.). The thick blue line for each group of measurements indicates the mean value. Scanners with which there are significant pairwise differences are indicated above each distribution; red labels indicate $p < 0.01$.

6.3.4.3 Group comparisons

Two of the male rats died before completion of the experiment and in one female rat MCAO was incomplete; these were excluded from group comparisons in the original study and from this study as well. Tables 6.6 and 6.7 show the results of infarct volume comparison between the remaining eight male and 10 female rats; Table 6.6 lists the percentage median differences between groups and Table 6.7 lists the corresponding p -values from exact two-tailed Mann–Whitney U tests performed for each simulated dataset.

The group percentage median difference for absolute V_{I_u} was within 26.68–31.93% across scanners for all examined cases; the range was smaller for calibrated systems using different sequences (case 3; 2.73%). For absolute volumes corrected using either method, the range of percentage median differences was similar to that for the absolute V_{I_u} . Comparing groups using relative estimates of infarct volume increased the variability between scanners. Relative estimates of infarct volume corrected based on all slices varied similarly when expressed as a percentage of the brain or either hemisphere, with the overall variability being within 21.81–34.11%. Expectedly, relative estimates of infarct volume corrected based on a different number of slices per animal varied in a greater extent, specifically for cases 1 and 2 for all types of relative measures (Table 6.6). Furthermore, the percentage median differences for these measures was much smaller than for the absolute V_{I_u} (overall range is 5.84–21.05%). When volumes were corrected based on all slices, variability was smaller between uncalibrated systems using a standardised sequence (case 2), but when volumes were corrected based on slices including the infarct, variability was smaller between calibrated systems using different sequences (case 3).

Table 6.6 Percentage median differences in infarct volume between male and female rats.

Infarct volume measure	Case	A	B	C	D	E	F	Range
Uncorrected, in mm ³	1	31.43	28.45	29.67	27.65	28.35	28.71	3.78
	2	31.93	31.87	31.04	31.33	30.97	28.17	3.76
	3	29.14	28.10	28.68	27.16	29.41	26.68	2.73
Corrected based on all slices								
in mm ³	1	35.39	34.50	30.20	29.86	34.33	28.94	6.44
	2	35.25	35.31	33.57	35.40	32.53	30.26	5.14
	3	29.51	32.07	29.39	27.46	33.29	28.70	5.83
as % of brain	1	31.47	30.56	25.58	25.82	30.33	24.89	6.58
	2	29.97	31.29	29.57	31.48	28.49	27.16	4.32
	3	26.30	27.97	26.57	23.31	29.27	24.76	5.96
as % of ipsilateral hemisphere	1	28.83	28.81	23.72	24.00	28.47	23.56	5.27
	2	27.36	28.58	26.98	29.73	25.84	24.96	4.77
	3	25.05	25.51	25.32	21.81	27.45	22.82	5.64
as % of contralateral hemisphere	1	33.68	32.73	27.88	28.07	32.62	27.31	6.37
	2	33.66	34.11	31.80	33.64	31.00	29.87	4.23
	3	28.41	30.31	28.74	25.70	31.54	27.44	5.84
Corrected based on slices including the infarct								
in mm ³	1	32.51	30.40	28.00	26.97	30.49	26.77	5.74
	2	29.66	31.66	30.70	29.59	29.35	27.31	4.35
	3	26.44	27.43	25.43	25.14	29.24	27.58	4.10
as % of brain	1	16.65	14.46	13.85	9.29	7.24	11.45	9.41
	2	11.32	16.24	18.07	8.33	16.37	7.06	11.01
	3	9.52	12.12	12.49	8.73	10.32	13.40	4.67
as % of ipsilateral hemisphere	1	15.30	14.60	12.06	7.46	6.14	10.39	9.16
	2	11.20	14.74	15.65	6.95	14.94	5.84	9.81
	3	9.10	11.00	9.64	7.69	8.70	12.57	4.88
as % of contralateral hemisphere	1	19.50	15.36	16.43	10.73	9.93	12.90	9.58
	2	13.71	18.09	21.05	10.83	18.11	9.18	11.87
	3	12.14	13.93	16.03	13.15	12.49	16.41	4.28

All values are percentages (%) except for 'Case'. Eight male and 10 female rats comprise the compared groups. Case refers to simulations of scans when (1) scanners are not calibrated and sequences 'a'-'f' are used at corresponding centres, (2) scanners are not calibrated and the standardised sequence 'a' is used, and (3) scanners are calibrated and sequences 'a'-'f' are used at corresponding centres. The range of the percentage median difference across scanners is also shown. 'A'-'F' indicate the scanner.

In concordance with percentage median differences, statistical comparisons between infarct volume measurements in male and female rats show that the method used for correcting for brain swelling and expressing the volume as a relative estimate leads to large deviations in p -values (Table 6.7).

The p -value for absolute V_{I_u} was within 0.008–0.021 across scanners for all examined cases. In contrast with percentage median differences, the range was smaller for uncalibrated systems using the standardised sequence ‘a’ (case 2; 0.007). For absolute volumes corrected using either method, the range of p -values was similar to that for the absolute V_{I_u} (all below 0.027). For some scanners, group differences were highly significant ($p < 0.01$); this was often the case for scanner ‘D’.

Similarly to percentage median differences, comparing groups using relative estimates increased the variability between scanners. p -values for relative estimates of infarct volume corrected based on all slices were below 0.083 for all types of relative estimates and cases (overall range is 0.012–0.083). Relative estimates of infarct volume corrected based on a different number of slices per animal varied in a greater extent, with all p -values being larger than the maximum measured for the absolute V_{I_u} (overall range is 0.034–0.460). For the majority of infarct volume measures shown in Table 6.7, between-scanner variability in p -values was smaller when the standardised sequence ‘a’ was used (case 2). Group comparisons led to almost identical results for scanners ‘A’ and ‘C’ in case 3 for any of the examined infarct volume measures.

Figures 6.15 and 6.16 present plots of infarct volume in the two animal groups drawn from corresponding distributions of measurements shown in Figures 6.13 and 6.14 respectively; bootstrap estimates of the 95% CIs of each median value and results of group comparisons according to Table 6.7 are also indicated. Results obtained by Mann–Whitney U tests agreed with CIs; overlapping of the latter normally suggests that no significant difference exists between groups.

Table 6.7 Mann–Whitney *U* tests comparing infarct volume between male and female rats.

Infarct volume measure	Case	A	B	C	D	E	F	Range
Uncorrected, in mm ³	1	0.016	0.009	0.021	0.012	0.016	0.021	0.012
	2	0.012	0.016	0.016	0.009	0.016	0.009	0.007
	3	0.012	0.021	0.012	0.008	0.012	0.021	0.013
Corrected based on all slices								
in mm ³	1	0.012	0.006	0.016	0.009	0.016	0.016	0.009
	2	0.012	0.016	0.012	0.004	0.016	0.004	0.011
	3	0.016	0.027	0.012	0.009	0.021	0.016	0.018
as % of brain	1	0.021	0.012	0.021	0.016	0.027	0.021	0.015
	2	0.016	0.021	0.021	0.012	0.021	0.012	0.009
	3	0.021	<i>0.083</i>	0.016	0.016	0.027	0.021	0.068
as % of ipsilateral hemisphere	1	0.021	0.009	0.021	0.016	0.043	0.021	0.035
	2	0.016	0.021	0.021	0.012	0.021	0.021	0.009
	3	0.021	<i>0.083</i>	0.016	0.016	0.034	0.027	0.068
as % of contralateral hemisphere	1	0.021	0.016	0.021	0.016	0.027	0.021	0.011
	2	0.012	0.027	0.021	0.012	0.027	0.016	0.015
	3	0.016	0.027	0.016	0.016	0.016	0.021	0.011
Corrected based on slices including the infarct								
in mm ³	1	0.016	0.006	0.016	0.006	0.016	0.004	0.011
	2	0.016	0.016	0.012	0.006	0.016	0.004	0.011
	3	0.006	0.021	0.006	0.006	0.012	0.016	0.014
as % of brain	1	<i>0.068</i>	<i>0.055</i>	<i>0.101</i>	<i>0.068</i>	<i>0.408</i>	0.043	0.365
	2	<i>0.101</i>	<i>0.101</i>	0.034	<i>0.083</i>	<i>0.122</i>	<i>0.122</i>	0.088
	3	<i>0.055</i>	<i>0.146</i>	<i>0.068</i>	0.034	<i>0.237</i>	<i>0.068</i>	0.203
as % of ipsilateral hemisphere	1	<i>0.068</i>	<i>0.068</i>	<i>0.146</i>	<i>0.101</i>	<i>0.460</i>	0.043	0.416
	2	<i>0.146</i>	<i>0.101</i>	0.043	<i>0.122</i>	<i>0.122</i>	<i>0.146</i>	0.102
	3	<i>0.055</i>	<i>0.237</i>	<i>0.055</i>	0.043	<i>0.315</i>	<i>0.083</i>	0.272
as % of contralateral hemisphere	1	0.043	0.043	<i>0.101</i>	0.043	<i>0.274</i>	0.043	0.231
	2	<i>0.083</i>	<i>0.083</i>	0.043	<i>0.083</i>	<i>0.122</i>	<i>0.101</i>	0.079
	3	<i>0.055</i>	<i>0.101</i>	<i>0.055</i>	0.027	<i>0.173</i>	<i>0.068</i>	0.146

All values are *p*-values except for 'Case'. Corresponding *U* statistics are given in Table C.4. Eight male and 10 female rats comprise the compared groups. Case refers to simulations of scans when (1) scanners are not calibrated and sequences 'a'–'f' are used at corresponding centres, (2) scanners are not calibrated and the standardised sequence 'a' is used, and (3) scanners are calibrated and sequences 'a'–'f' are used at corresponding centres. *p*-values less than 0.01 are shown in bold and above 0.05 in italic. The range of *p*-values across scanners is also shown. 'A'–'F' indicate the scanner.

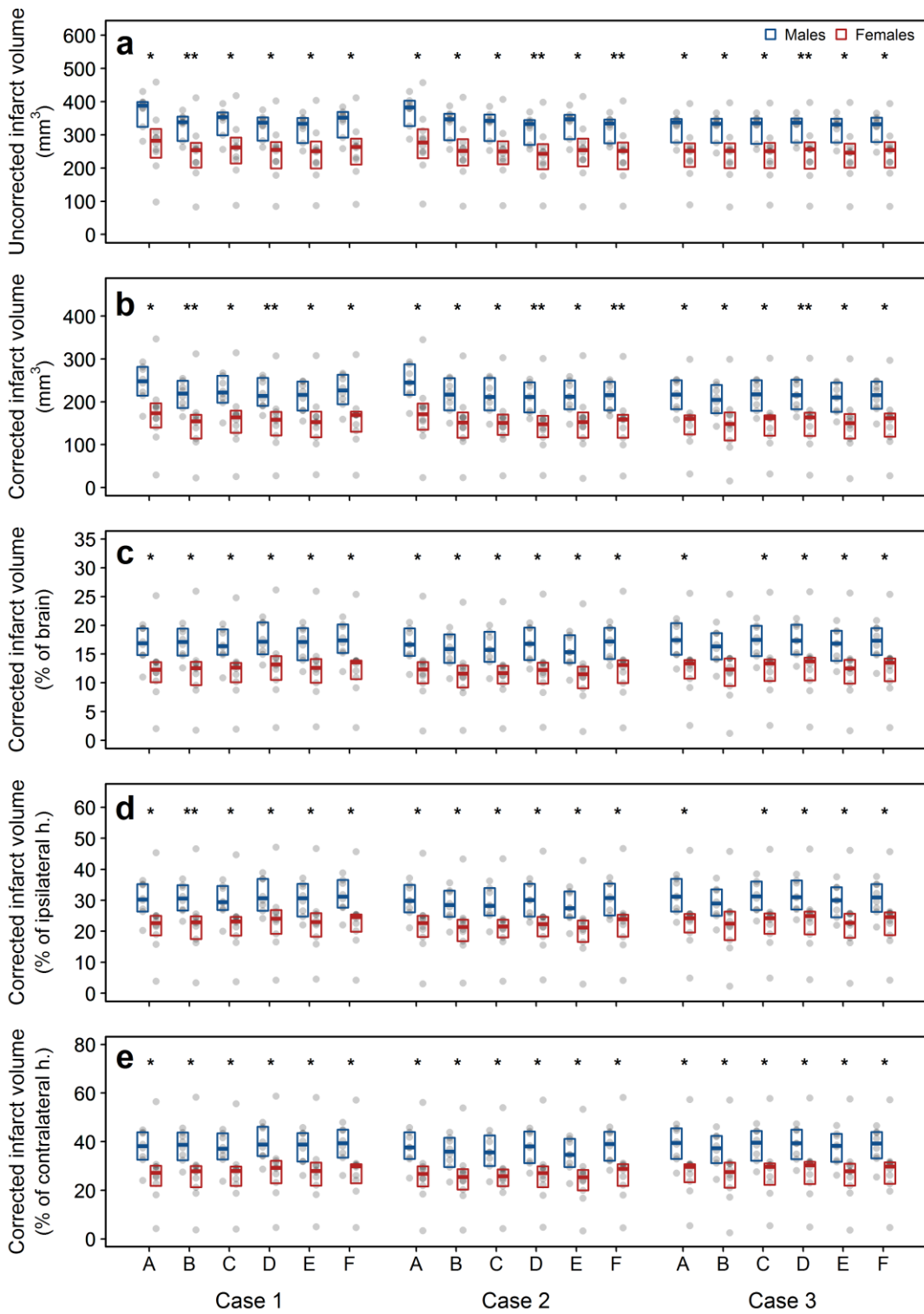


Figure 6.15 Infarct volume in male and female rats corrected based on all slices and expressed as an absolute or relative estimate.

Measurements for all three cases for scanners ‘A’–‘F’ as (a) an absolute uncorrected and (b) an absolute corrected infarct volume are shown, as well as the latter expressed as (c) the percentage of the brain, (d) the ipsilateral and (e) the contralateral hemisphere (h.). Each coloured box indicates the median and its 95% confidence intervals. Significant difference between groups is indicated above each pair (based on Mann–Whitney U test); * indicates $p < 0.05$; **, $p < 0.01$.

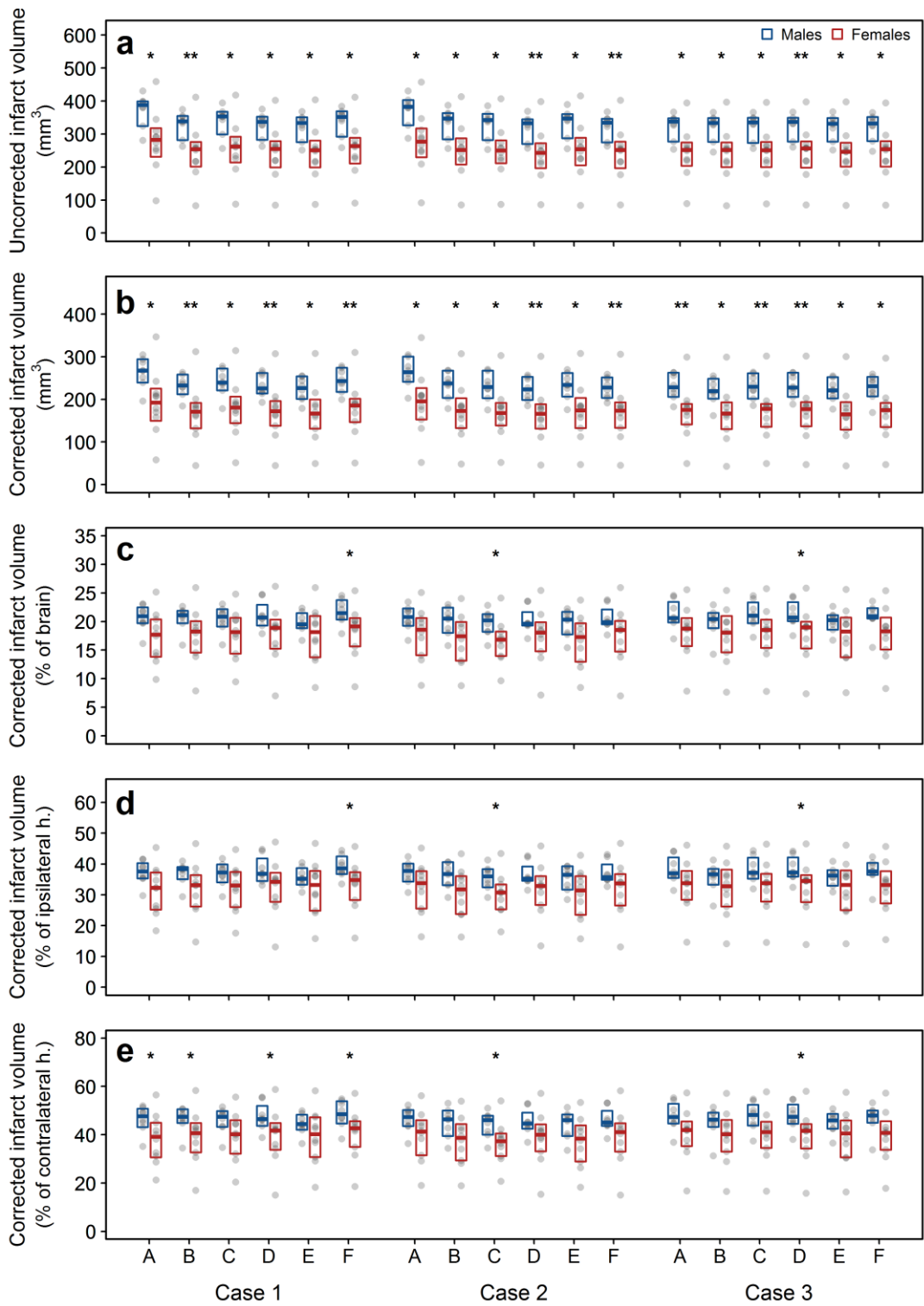


Figure 6.16 Infarct volume in male and female rats corrected based on slices including the infarct and expressed as an absolute or relative estimate. Measurements for all three cases for scanners ‘A’–‘F’ as (a) an absolute uncorrected and (b) an absolute corrected infarct volume are shown, as well as the latter expressed as (c) the percentage of the brain, (d) the ipsilateral and (e) the contralateral hemisphere (h.). Each coloured box indicates the median and its 95% confidence intervals. Significant difference between groups is indicated above each pair (based on Mann–Whitney *U* test); * indicates $p < 0.05$; **, $p < 0.01$.

6.4 Discussion

Despite that the use of preclinical MRI systems in experimental research was intensified long ago by the need of *in vivo* assessment of injury over extensive periods of time, individual centres still rely on annual maintenance of their systems (O'Callaghan et al., 2014), which is performed at different times across centres. At the same time, imaging hardware and protocols for *in vivo* imaging are highly heterogeneous, as shown in Chapter 3 (Milidonis et al., 2015). The impact of this variability in preclinical stroke studies remained unexplored, but its influence on effects derived using MRI lesion volumes can be profound, as recently hinted by findings of possibly the first multicentre preclinical stroke trial (Maysami et al., 2015). Using the simple QA approach presented in Chapter 4, a multicentre study was designed and carried out to specifically determine how imaging performance and heterogeneity between and within scanners can influence geometric accuracy when MRI is used as in the assessment of infarct size in rodent models of stroke. It was found that miscalibrated gradient coils have the most significant impact. To identify whether the use of different sequences increases variability compared to the use of an identical sequence, each centre's preferred *in vivo* sequence was used to scan the phantom and QA measurements were compared to those from a standardised sequence. Findings suggest that an identical sequence ensures better agreement between centres and could thus be more favourable in a multicentre setting.

Geometric accuracy was assessed by measurement of two QA criteria that were described in Chapter 4: linear scaling errors, estimated according to the internal dimensions of the phantom measured in scans in all imaging planes, and volumetric errors, based on the volume of the central compartment of the phantom measured in the axial plane. The axial plane in phantom imaging corresponds to the coronal plane in rodent brain MRI often used in experimental studies (x - y plane; Figure 2.3) and the phantom's central cylindrical compartment encompasses marked PVE in this plane that is characteristic of lesions in 2D MRI data.

Unfortunately, acceptance limits for geometric accuracy in preclinical systems are currently not available, but scaling errors in systems 'B'–'F' were comparable and mostly within the $\pm 1\%$ limit proposed by the American College of Radiology for clinical systems (American College of Radiology, 2015). Scaling errors in system 'A' were higher and similar in extent to those measured in a scanner from the same vendor in a previous study (O'Callaghan et al., 2014). Unexpectedly, a persistent difference in measurements between frequency and phase encoding directions in all imaging planes was observed in this system with sequence 'a', both before and after gradient coil calibration. Linewidths from manual global magnetic field shimming were well below 100 Hz for all scans and the rBW was 100 kHz; these are often sufficient to negate most distortion effects. Yet, the presence of this difference points to other sequence limitations and MRI artefacts that are prevalent in either of the two in-plane directions, such as ghosting or Gibbs ringing due to Fourier transform imperfections in phase encoding.

In contrast, volumetric errors were positive and inconsistent across centres. This emphasises that assessment of volume using 2D sequences can be influenced adversely by factors other than linear scaling errors and normal system performance drift. Firstly, a large voxel size can blur the boundary of an object due to PVE, leading to a seeming boundary outward shift and eventually volume overestimation. This is a well-known phenomenon and was observed in previous QA (Khan et al., 2012; Luft, Skalej, Welte, Kolb, & Klose, 1996) and in vivo studies (Ren et al., 2004). Stroke sequences 'c' and 'd' had the thickest slices and led to the two largest differences between measured and predicted volumes amongst centres. The use of a smaller slice thickness and therefore more slices through the volume of interest can reduce this artefact, explaining the higher overall measurement accuracy of stroke sequences compared to the standardised sequence in systems 'B'–'F'. Secondly, while a high SNR improves image quality and delineation of the ROI, in slices with significant PVE the apparent contrast between ROI and background may be enhanced, causing inclusion of false positives. Here, data from centre 'B' had a noticeably lower SNR

compared to data from other centres, leading to small differences between measured and predicted volumetric errors. Finally, slice positioning may affect volume estimation in a considerable extent, particularly when the ROI is small or the slices are very thick, and even small variations can introduce inconsistencies in PVE (Ashton et al., 2003) and gradient distortion effects (Jovicich et al., 2006). Phantom scanning in this study adhered to detailed instructions for slice positioning, among other, and acquired images were comparable; any influence of this factor could have only been minimal. These observations suggest that formulae currently used for estimating volumes from 2D in vivo sequences data could be modified to account for unwanted misestimations (Liu et al., 2009; Rosen & Harry, 1990; Sahin et al., 2003), but further studies are needed to determine the best approach in individual study designs.

Using slices through the middle of the phantom, maps depicting the deformation applied to the images to match distortion-free CT or baseline images were created, to provide aid in the interpretation of geometric errors. The geometric distortion seen in these maps is comparable to the expected 3D distortion in the MRI systems due to magnetic field inhomogeneity but might not be identical, as they are quantified using different methods; a phantom with a large number of equidistant control points (grid structure) and 3D registration methods are often required for the latter (O'Callaghan et al., 2014; Wang, Strugnell, Cowin, Doddrell, & Slaughter, 2004). The 2D maps created here clearly revealed that, further to factors described previously, the measurement variability in scanners 'B'-'F' was influenced by subtle spatial distortion effects that could have arisen due to a number of possible reasons. Firstly, design limitations of gradient coils in high-field systems, such as short bore or short gradient rise times that induce eddy currents in nearby conducting materials, can produce nonlinearities in the magnetic field (Caramanos et al., 2010; Wang et al., 2004). The pattern of distortion was similar for systems 'B'-'F' that make use of gradient coils from the same manufacturer, but could not be observed in system 'A' in any scan acquired, even after using in vivo sequence 'b' and a modified version of

it which theoretically amplifies distortion effects. However, the phantom's size is not sufficient to allow deduction of rational acceptance decisions, as manufacturers for these coil systems specify the maximum permitted linearity over a much larger volume than that of this phantom (in the coils used here this is between $\pm 4\text{-}5\%$ over an 80 mm diameter of spherical imaging volume). Secondly, additional local magnetic field inhomogeneities may have been produced due to magnetic susceptibility differences along the interfaces of the phantom's contrast solution, plastic and surrounding air. As the phantoms' plastic materials and water are both diamagnetic (Wapler et al., 2014) and the characteristic susceptibility effect of opposing regions with bright and dark signal was not apparent in the images, it is believed that influence of susceptibility along the plastic-solution interface has been insignificant. However, the presence of paramagnetic air in close proximity might have been influential, which raises further concerns over in vivo imaging across centres, as the air-filled sinuses and nasal cavity lie close to the brain and could cause heterogeneous susceptibility artefacts. Finally, it must be noted that a narrow rBW may introduce positional shifts along the frequency encoding direction as it decreases gradient amplitude and enhances susceptibility artefacts. This could partially explain why nonlinearities were more enhanced in images taken using the stroke sequences utilizing about half of the standardised sequence's rBW to improve SNR. In general, geometric distortion in small animal scanners increases rapidly as the distance increases from the magnet's isocentre, indicating that even relative measures based on a small and a larger volume—such as the lesion to brain size ratio—could be affected detrimentally. Methods for correcting geometric distortion should be used in preclinical MRI as well (Jack et al., 2008; Jovicich et al., 2006; O'Callaghan et al., 2014; Teh et al., 2016).

Despite that imaging performance is undoubtedly a major source of uncertainty in measurements, its impact on infarct size and associated statistical group comparisons has never been established; such comparisons typically determine the conclusions drawn and the success or failure of the study. Using in vivo data simulations, the

influence of the two largest sources of variability—scaling errors and voxel size differences—was examined. Representative scans for each of the six participating centres were generated and analysed by a single rater using the semi-automated tool described in Chapter 5, and infarct size was estimated using various approaches reported in literature. Simulations debunked the common misconception amongst researchers that relative estimates of morphometric outcomes allow efficient comparisons between scanners; in the case of infarct volume in experimental stroke MRI, at least, measurement variability and related group differences may increase due to the interplay of the various ROIs used to estimate these relative measures. Comparisons based on a repeated measures design suggested that infarct volume differs significantly between scanners when gradient coils are not calibrated and different sequences are used, when it is expressed either as an absolute or a relative estimate. The concordance between systems improves when they are characterised by high geometric accuracy or identical sequences are used. Concordance also improves when the infarct volume is corrected for brain swelling based only on slices showing the infarct and expressed as a relative estimate, probably because this method introduces a level of heterogeneity in measurements that eliminates biases in the direction of inter-animal volume differences across scanners, which normally determines the result of a repeated measures ANOVA test. The same reason could explain the small proportion of variance in these measurements that is explained by variation in scanners; variability is mostly within each scanner, attributed to the use of a different number of slices per animal when estimating this relative infarct volume. Nevertheless, group comparisons reveal that this method is inappropriate for correcting infarct volume as it leads to percentage differences that are much smaller than those measured in the absolute V_{I_u} ; accordingly, p -values from Mann–Whitney U tests suggest that no significant differences exist between groups, while significant differences exist between groups for the absolute V_{I_u} . Between-scanner variability in percentage group differences and p -values was smaller when a standardised sequence and uncalibrated systems were used, instead of different sequences and calibrated systems. Despite this, values differed for each examined

case and infarct volume measure, with p -values fluctuating above or below the significance level rather unpredictably. Semi-automated analysis is not to blame for this variability; measurements in identical datasets (scanners 'A' and 'C' for case 3) were not significantly different for any of the evaluated infarct volume measures, and corresponding group comparisons led to equivalent results for these scanners. These observations suggest that the only option in a multicentre scenario is the use of highly accurate systems and identical imaging protocols, agreeing with previous findings.

The percentage median differences are of particular importance in preclinical trials, as they are estimated in a similar fashion to the ES representing the efficacy of tested drugs (equation A.4 in Appendix A). While not common, researchers may choose to correct infarct volume for brain swelling by using V_{IH} and V_{CH} measured only on slices where the lesion is visible and then proceed to report it as a relative measure (Lestro Henriques et al., 2015). At a first glance, this seems a safe solution to the negating effect of formulae used to correct for brain swelling when it occupies a large part of the ipsilateral hemisphere (McBride et al., 2015). However, the space-occupying effect of oedema may extend beyond the slices where the infarct is visible, as indicated by the midline shift in the brain in the slice of the rostral end in Figure 6.12. In the dataset used here, for a 30% treatment effect based on the absolute V_{I_u} , the effect can be 15% or smaller based on the aforementioned relative estimate. This marked difference in ES could be detrimental for meta-analyses' validity, a primary hypothesis of which is the concordance of absolute and relative estimates of MRI infarct volume when pooling effects. It could therefore be important for such meta-analyses to first determine the presence of a potential bias in treatment effects before proceeding to data pooling. In the meantime, strict guidelines for the use of oedema-correcting formulae and the estimation of relative infarct volume must be defined for future use; ideally, hemispheric volumes based on the whole MCA territory should be used, covered by the same number of slices with an identical thickness across centres. Even though correction of the smallest infarct in this study did not lead to a negative estimate (from 82.8 mm³ to 22.6 mm³ in the reference scan), very small infarcts could

be excluded from analyses.

The transformations used a pooled scaling factor for each orthogonal direction in the evaluated MRI systems, combining factors accounting for the linear scaling errors in the scanner used to acquire the reference dataset, the linear scaling errors in each participating scanner and the required voxel size. It must be noted that scaling each reference scan three times using the individual factors could have caused cumulative transformation errors, such as significant blurring or production of more or less slices than those suggested by the pooled scaling factor and thus misestimations of infarct volume. The first task was to correct the reference scans for the linear scaling errors in the system that was used to acquire them ('B'; Table 6.1). Even if errors in this system were different at the time of scanning than those used here (Figure 6.4 b), statistical comparisons between simulated datasets would not be affected since the same correction was applied to all simulations; this could probably hold true even if this correction was entirely skipped. Furthermore, one could argue that infarct volume for each of the cases examined in this study could be measured by either scaling the binary infarct masks for the reference dataset, or multiplying the number of voxels in these masks with 'scaled' voxel sizes. Either approach would reduce analysis time significantly, but would fail to simulate changes in the shape and appearance of the infarct, as well as PVEs in the outermost slices.

A major limitation of the multicentre phantom study is that only 7 T scanners were evaluated, while just one of them was an Agilent scanner and longitudinal imaging was performed with this system alone. Inclusion of more preclinical systems would have provided more conclusive evidence for the role of field strength and vendor in geometric accuracy, but could not be achieved here due to logistical considerations. Furthermore, small deviations from the scanning protocol may have introduced small systematic variation in some of the analyses that could possibly render corresponding comparisons less effective. The most apparent deviation from the protocol was the use of a mouse head surface coil by one of the centres instead of a rat head surface coil, but post hoc imaging showed that this could not affect measurements

detrimentally. In any case, such deviations would be representative of the true heterogeneity in experimental imaging and the subsequent variation in measured biological effects. The limitations of the simulation study vary: Firstly, 2D MRI images were used as reference scans, provided by a collaborating institution performing stroke experiments routinely. High resolution 3D data with isotropic voxel size would have been ideal, because they would minimise interpolation errors in the transformation, provide more realistic PVEs in the generated slices and allow establishment of ground truth measurements. However, imaging protocols in experimental stroke remain 2D sequences with thicknesses that are significantly coarser than the in-plane resolution (0.35-3 mm in rat T2WI; Table 3.1). Furthermore, the purpose of the investigation was to compare infarct volumes and related study effects between centres, and not determine imaging accuracy by comparison with a ground truth. Secondly, the reference images were scaled using linear factors estimated by scanning a cuboid MRI phantom with a size similar to that of the rat brain. This fails to capture the exact geometric distortion in the FOV. 3D warping using deformation maps estimated by scanning dedicated whole coverage geometric phantoms at each MRI centre could be more useful but still not ideal, as such phantoms cannot model the exact characteristics of the rodent head and therefore the deformation maps would not capture associated MRI artefacts. Any potential nonlinearities and artefacts already present in the reference scans were transferred to the generated ones by the transformations used. Similarly, it was assumed that animal placement and orientation in the scanner, slice positioning, tissue contrast in each sequence and the method for image analysis were identical for each simulated dataset. SNR differences were not simulated as values for each scan and each centre were expected to be significantly above 20; any higher SNRs offer little advantage to image analysis (McRobbie et al., 2006). Nevertheless, the statistical comparisons performed might have even underestimated the true variability in effects between centres. Finally, the small size of the dataset used could have had a significant impact on statistical comparisons. Unfortunately, incorporating additional data was not possible due to the significant heterogeneity between animal studies in terms of

overall aims and experimental procedures (Milidonis et al., 2015; Rivers & Wardlaw, 2005). No such data are publicly available either. This study emphasises the need for larger and more effective studies, as well as the establishment of public repositories where MRI and other data from animal stroke studies are kept for use in post hoc exploratory analyses, such as the one performed here.

This study uncovers a widespread inconsistency in geometric accuracy of different preclinical MRI scanners, raising concerns regarding the comparability of measured outcomes across centres. Linear scaling and volumetric errors in scanners are overstated when MRI gradient coils are not properly calibrated, but even after calibration heterogeneity can be significant due to the use of 2D imaging protocols. Statistical comparisons were found to be prone to small differences in scanner errors, potentially leading to conflicting conclusions about the examined effect if the exact same animals were to be scanned across centres. Scanners of the same manufacturer have largely similar performance and imaging using an identical sequence ensures better agreement between measurements. While these might not be a concern in studies performed locally that estimate effects based on relative measurements and comparisons against controls assessed with the same method, multicentre studies could combine absolute measurements for one or more cohorts from different centres with a highly variable number of animals. In such studies, the conduct of MRI must be optimised by the development of standardised methods for routine scanner QA and in vivo imaging, similar in rigor to those utilized in clinical centres (Jack et al., 2008; Wintermark et al., 2008). Procedures such as those described here must be considered the minimum set of QA tests performed. Additional tests may be needed if scanners are used routinely for advanced stroke MRI, such as diffusion-weighted and functional MRI that place even higher demands on system performance. Characterization of the effects of these imaging methods across centres was beyond the scope of this study, but should be the focus of future work for the establishment of a comprehensive multicentre QA program.

Chapter 7

Discussion and Conclusions

This chapter summarises the main contributions of this thesis and describes potential implications for experimental focal cerebral ischaemia research. A list of recommendations based on major findings is provided and future directions are discussed.

7.1 Contributions

A known fact since the nineties, the failure in the translation of neuroprotective effects from animals to humans led to the development of recommendations to improve the quality and reporting of preclinical stroke research (Stroke Therapy Academic Industry Roundtable (STAIR), 1999). Later assessment of published records revealed that studies failing to report measures to reduce bias—randomisation and blinding—could falsely inflate estimates of efficacy; this forced an update to these criteria (Fisher et al., 2009) and the publication of further recommendations for good laboratory practice (Macleod et al., 2009). To complement STAIR criteria, specific instructions for the modelling of stroke and the measurement of physiological parameters and other outcomes were made available (Liu et al., 2009). Imitating standard recommendations for reporting clinical trials (Begg et al., 1996), the ‘Animals in Research: Reporting In Vivo Experiments’ (ARRIVE) guidelines were published to serve as the reference in preclinical publication practice (Kilkenny, Browne, Cuthill, Emerson, & Altman, 2010). More recently, generic guidelines for the description of basic information regarding the use of various imaging modalities in preclinical research were published (Stout et al., 2013). In the meantime, others suggested a more radical solution: implementation of multicentre animal stroke trials with the same rigour as randomised controlled clinical trials to improve the validity and generalisability of experimental findings (Bath et al., 2009; Dirnagl & Fisher, 2012; Dirnagl et al., 2013; Howells & Macleod, 2012). Such studies utilising the diagnostic and prognostic capacity of MRI could improve the understanding of stroke and help shape the future of its treatment.

While all these recommendations constitute a major step towards the improvement of preclinical stroke research, no consideration has been given specifically to the use and reporting of MRI in animal stroke studies. This could prove damaging for multicentre animal stroke studies. Large clinical imaging collaborations have investigated the use of MRI and devised standardised protocols for widespread use (Jack et al., 2008; Wintermark et al., 2008), but no preclinical consortium has ever

reached this milestone. Nevertheless, MRI is a versatile tool and its applications in stroke are ever-expanding, suggesting that definition of a standardised multimodal imaging protocol could be a far-fetched goal (Duong, 2012a, 2013; Farr & Wegener, 2010).

Instead, this thesis focussed on the most rudimentary application that is often taken for granted: the measurement of infarct volume using basic sMRI protocols. A coherent and systematic approach was followed to investigate approaches used for both image acquisition and analysis, identify factors contributing to variability between and within centres, and define strategies for its reduction to facilitate multicentre animal stroke studies. Firstly, a systematic review of relevant studies was carried out to determine the extent of heterogeneity in MRI acquisition and post-processing methods for the assessment of lesions. Extracted data were meta-analysed to identify the MRI contrast technique providing the most accurate estimate of late infarct-volume based treatment effect by comparison with gold standard post-mortem histological analysis. Secondly, a multicentre study involving six different MRI scanners across Europe was designed and carried out, in order to assess geometric accuracy in systems when used for stroke imaging. Imaging consisted of scanning a dedicated structural phantom using either a standardised sequence or centre-specific in vivo sequences used for the measurement of infarct size, and data were analysed using an automated tool developed specifically for this purpose. Finally, in vivo images in a rat model of stroke were simulated for each scanner by modelling the measured geometric errors and the voxel size for corresponding in vivo sequences. Infarct volume was measured in each scan using a custom-made threshold-based semi-automated tool and statistical tests were performed to determine the impact of MRI variability in both infarct volume and related group comparisons across centres.

7.1.1 Tools

Assessment of geometric accuracy in different MRI scanners was made possible by the use of sMRI phantoms. The phantoms had to be precisely constructed, have an overall size similar to that of the rat MCA territory, be longitudinally stable and be made with cheap, readily available and MRI-suitable materials. LEGO bricks were chosen and used to construct a number of identical phantoms (Chapter 4). Phantoms were filled with a Gd-DOTA solution with a T_2 relaxation time similar to that of the healthy striatum in the rat brain at 7 T (53.56 ms). Using callipers and high resolution CT imaging, phantoms were found to be structurally precise and suitable for multicentre QA. The internal dimensions and the volume of the central compartment of the phantom were used as criteria for geometric accuracy. Acquired phantom data had a high quality, which allowed the development of a simple and fully automated segmentation tool in MATLAB for their analysis. The main processing step in the analysis pipeline is the crude detection of the boundary of the phantom using Canny edge detection, which is a computationally light algorithm that performs accurate detection of local intensity variations without relying on assumptions of global intensity homogeneity (Canny, 1986). Intensity inhomogeneities were found to affect other tested algorithms such as basic thresholding and k -means clustering. Segmentation is aided by additional morphological processing steps that refine the boundary of the phantom. The tool was found to perform full assessment of each scan in just 5.3 seconds on average and with very high accuracy as compared to gold standard manual analysis (DC: median = 0.982, IQR = 0.975–0.983).

The work described in this thesis also led to the development of a second tool for the detection of the infarct and measurement of its volume on T2WI images in rodent models of stroke (Chapter 5). Similar to phantom data, the primary artefact in these images is the bias field arising from RF coil sensitivity. However, the abnormal shape of the infarct and its poor contrast compared to the surrounding healthy tissue did not allow implementation of a similar analysis pipeline. Estimating the bias field based on the voxels in the contralateral brain hemisphere proved pivotal and made

the use of a threshold-based approach viable. The tool performs binarisation of bias field corrected images using a threshold determined by the intensity in the healthy contralateral hemisphere of each animal, as used previously in analysis of quantitative MRI data (Jiang et al., 2006; Modo et al., 2009; Veldhuis et al., 2003). A series of morphological processing steps then refine the infarct boundary. The tool offered high segmentation accuracy at this stage (mean \pm SD of DC is 0.868 ± 0.052), but in many cases the hyperintense ventricles and global intensity variations in the slice direction were causing volume overestimations. Therefore, analysis is performed with a custom-made MATLAB GUI allowing manual post-processing of segmented regions where deemed necessary. The tool is independent of the severity of stroke, the extent and consequences of vasogenic oedema, the shape of the infarct or the alignment and positioning of the brain in the image that may confound unsupervised analysis techniques. Infarcts defined by semi-automated and gold standard manual analysis are alike (mean \pm SD of DC is 0.931 ± 0.022), but semi-automated analysis is significantly faster (mean \pm SD of analysis time per animal is 3.68 ± 0.88 and 10.54 ± 1.58 minutes) and more reliable when used by the same rater (mean infarct volume difference in repeated analyses is -0.6 and 4.3 mm³). Reproducibility in inter-rater analysis using this tool was not as high, despite equivalent training of both raters. These observations suggest that the tool could permit effective use in centralised data analysis in multicentre studies, which could be crucial for the execution of reliable statistical comparisons between cohorts.

7.1.2 Findings

Based partially on the aforementioned phantom and tools, studies described in this thesis led to important discoveries regarding the use of MRI for the measurement of infarct volume across centres; the main findings are summarised below.

Reporting of methods for lesion size measurement in published articles is lacking.

The systematic review presented in Chapter 3 revealed that the analysis method used for lesion size determination was mentioned for 65% of imaging protocols. Specific

lesion boundary criteria were specified for thresholding techniques only, accounting for 76% of the reported analysis methods. Methods for correcting infarct volume in T2WI due to the space-occupying effect of oedema were used in 25% of relevant studies. At the same time, methods to reduce bias and improve the validity of analysis were seldom reported; blinded assessment of lesion size was reported in 17% of studies and analysis by multiple independent raters in just 2%. In contrast, parameters pertinent to MRI hardware and scanning protocols were more often reported; yet, no study provided all imaging items considered.

The conduct of MRI is highly heterogeneous between studies. Systematic review findings indicate that various scanners, RF coils and types of pulse sequences are used. T2WI was the most commonly reported contrast technique (83% of studies) and was performed using SE, FSE, MSME and FLAIR sequences with a range of parameters; in rats, for example, slices ranged from six to 35, slice thicknesses from 0.35 mm to 3 mm, FOV from 22×22 mm² to 120×120 mm² and matrix sizes from 64×64 to 512×512. Lesions were defined in images using manual, semi-automated (thresholding) or fully automated methods. Various thresholds of abnormality were mentioned for quantitative T2WI, DWI and PWI, but a clear justification was often omitted. Lesion size was usually reported as an absolute estimate (58% of studies), while other studies chose a variety of relative estimates.

T2WI is equivalent to gold standard histology for estimating infarct-based ES. Meta-analysis of articles using both MRI and post-mortem histology in the same animals at corresponding time points suggests that T2WI is strongly associated with histology for estimating late infarct-volume based treatment effects ($\bar{R}^2 = 0.699$, $p < 0.001$). In fact, despite that T2WI gives a slightly more conservative estimate of pooled efficacy compared to histology (22.2% and 25.5% respectively), CIs are narrower for the former, suggesting that this MRI technique may allow improvements in the statistical power of a study. On the other hand, DWI is poorly associated with histology for the estimation of infarct volume-based treatment effects.

Stereological point counting is an accurate and reliable method for measuring infarct volume if the grid of points is dense enough. Nevertheless, it does not permit generation of masks of the infarct that are necessary in modern morphological studies that perform post-processing analyses, such as assessment of its shape and appearance across animals or time points, or template formation using registration methods. Threshold-based semi-automated analysis using the developed MATLAB tool is faster, has a higher reliability when used by the same rater and allows extraction of lesion masks (Chapter 5).

The infarct is represented by tissue in the ipsilateral hemisphere with an intensity higher than 2.1 times the mean intensity in healthy tissue in the contralateral hemisphere. Automatically segmented infarcts (pre-manual editing) had the highest DC for this threshold when compared to manually defined infarcts. The threshold can be used for semi-automated localisation of the infarct in qualitative T2WI data of—at least—Sprague Dawley rats scanned 24 hours after MCAO. It was similar to the threshold used in quantitative T₂ maps in previous studies using a different rat strain and different time points (Jiang et al., 2006; Li et al., 2009), which implies that it could be applicable to a range of study designs.

Geometric accuracy between preclinical MRI scanners varies. Currently, preclinical MRI centres rely on annual maintenance of systems by external service engineers. Chapter 6 highlighted the repercussions of this practice; linear scaling errors in each orthogonal direction were mostly positive in six different 7 T scanners. Five of the evaluated systems were made by Bruker and had smaller errors than the single Agilent system used in the study. Nevertheless, images acquired with Bruker systems demonstrated prevalent nonlinearities that were not observed in the Agilent system. RF coils do not influence geometric accuracy significantly.

The use of an identical pulse sequence across scanners reduces the variability in geometric errors. In systems routinely used in stroke studies, linear scaling errors were within 0.03% to 1.84% with a standardised sequence and within -0.27% to 2.19%

with different in vivo T2WI sequences. Similarly, volumetric error ranged from 5.71% to 11.67% and 3.96% to 12.51% respectively.

Volume estimation using 2D pulse sequences is affected by slice thickness and SNR. Linear scaling errors in MRI systems alone are not sufficient to explain volumetric errors in phantom images. The error in repeated scanning using the calibrated Agilent scanner was persistently high (median = 7.33%, IQR = 0.43%). The difference between measured errors and those predicted by linear scaling errors was positively associated with slice thickness and SNR. Careful inspection of images points out that a thick slice contributes to significant PVE, while a high SNR increases the apparent intensity of the ROI in areas where PVE is significant, causing inclusion of false positives.

Longitudinal variability in geometric accuracy over six months is low. Linear scaling errors remained within $\pm 1.00\%$ over six months of repeated phantom scanning using the standardised sequence and the calibrated Agilent scanner. The coefficient of variation in phantom's cylindrical volume was 0.40%, which is minute compared to that in infarct volume in rodents over much shorter periods (Liu et al., 2009). The fluctuation in the apparent size of the phantom was similar during daily, weekly and monthly imaging.

Infarct volume differs significantly between the evaluated six MRI scanners. The simulation study presented in Chapter 6 suggests that a variability in geometric errors or in voxel size of the extent shown by the multicentre phantom study, is sufficient to render infarct volumes across centres incomparable. The variance in measurements between scanners is attributed mostly to differences in geometric errors.

Statistical comparisons between groups lead to different conclusions across scanners. The variability in group comparisons between scanners decreases when a standardised sequence is used. Infarct volumes expressed as relative estimates (percentage of brain or hemispheric volume) inflate between-scanner variability. For the dataset used in the simulation study, p -values across scanners fluctuated around

the significance level of 0.05 for relative infarct volumes, but were always less than the significance level for absolute volumes.

Correcting infarct volume for brain swelling and expressing it as a relative estimate based on hemispheric volumes measured only in slices that include the infarct is erroneous. The variability in such relative estimates is attributed mostly to within-scanner variance rather than between-scanner variance. This method leads to smaller group percentage median differences than for the absolute V_{I_u} (values for our dataset were within 5.84–21.05% and 26.68–31.93% respectively), while p -values are considerably more variable (0.034–0.460 and 0.008–0.021 respectively). Volumes corrected for swelling and expressed as an absolute measure based on all acquired slices (identical brain region for all animals) have low between-scanner variability in percentage median differences and p -values, which are similar to those for the absolute V_{I_u} (27.46–35.40% and 0.004–0.027 respectively).

7.2 Recommendations

In alignment with previous knowledge, these findings allow recommendations to be made that could provide an optimal environment for multicentre studies in rodent models of stroke employing sMRI to measure infarct size, facilitate pooling of measurements and improve the statistical power of the study.

Before embarking on a multicentre study, each centre should test and optimise the performance of their system by scanning a structural phantom with size and signal characteristics similar to that of the rodent brain. The imaging protocol should consist of a single T2WI sequence that is intended for use in the actual in vivo experiment; some parameters may be tailored to each centre, as described below. Despite that radiologists tend to acquire and analyse 2D data with high in-plane resolution, PVE and therefore measurement errors can be minimised by using a more isotropic voxel size; this should be standardised. With current hardware and software capabilities, a

FSE sequence such as 'f' in Table 6.2 could be close to the ideal; this made use of the thinnest slices among protocols examined in the multicentre QA study and obtained data with great quality and reduced volumetric errors. This, however, necessitates investment in a phased array RF receive coil that allows improvement in SNR when small NSAs are used. A SNR of at least 20 is required (McRobbie et al., 2006) and the duration of a scan should normally not exceed 10 minutes. At least one scan in each of the three primary imaging planes should be acquired to measure the phantom's dimensions in each orthogonal direction and identify any bias in encoding direction. Phantom data should be analysed locally using an automated tool to ensure consistency across centres and measurements should be used to perform basic calibration of all three gradient coils. The phantom, QA protocol and data analysis tool presented in Chapter 4 could be used effectively for this purpose; minor modifications of the tool's source code can allow analysis of images of any cuboid phantom. If the phantom allows assessment of volumetric accuracy, such as the one used here, phantom data from all participating centres could be compared centrally to identify factors that contribute to variability. In the undesirable case that considerable differences exist, centres could attempt to harmonise image quality and signal characteristics further by either using similar hardware (particularly RF coils) and by adjusting sequence parameters other than those determining voxel size (for example, TR, TE, NSA and rBW). Gradient nonlinearities are normally minute at very small distances from the magnet's isocentre and will affect the size of infarcts in rodents very little when centred in the scanner. In some gradient coil systems, however, nonlinearities could increase rapidly at further distances and potentially confound larger ROIs; if phantom imaging suggests that this is the case, a more advanced—albeit cumbersome—method could be implemented to correct data for this distortion (Jovicich et al., 2006; O'Callaghan et al., 2014; Teh et al., 2016).

Assuming that no intermediate changes in scanner software and equipment take place, QA performed every six months should maintain geometric accuracy at acceptable levels. However, if multicentre studies are assessing infarct size at time

points when vasogenic oedema is prevalent (necessitates measurement of hemispheric volumes), scanners should be monitored weekly according to recommendations for clinical QA (American College of Radiology, 2015). In reality, the quality of MRI data can be influenced by a range of additional factors, therefore procedures described here must be considered the minimum set of QA tests performed. Additional tests may be needed if scanners are used routinely for advanced stroke MRI, such as DWI and functional MRI that place even higher demands on system performance, but characterization of their effects was beyond the scope of this thesis. QA procedures including phantom placement, slice positioning, scanning and preparation of data should be clearly defined and disseminated across centres in advance. Multicentre studies could also benefit from staff training beforehand to reduce variability in the collection of data.

As emphasised numerous times in the past, a priori sample size calculations are essential in any study that draws conclusions based on statistical comparisons; studies that do not report a sample size calculation may overstate the reduction in infarct volume (Dirnagl, 2006; Fisher et al., 2009). This thesis demonstrated the dramatic effects of using few animals on group comparisons and suggests the use of an appropriate sample size to reduce the uncertainty in combined estimates of efficacy and group comparisons. The hardware, software and pulse sequences optimised in QA should be used for in vivo scanning as well. Slices should be positioned at the same location in each animal relative to an anatomical feature, such as the bregma or the rostral end of the rhinal fissure, consistently for each scan. Depending on the species and strain used, a sufficient number of slices should be used to cover the whole MCA territory; this should be identical for all centres. An imaging session including various MRI sequences should typically last about one hour, in order to minimise the confounding effects of anaesthesia and ensure the welfare of animals. Acquired images should be collected at one centre and evaluated by an experienced radiologist for concordance in quality and slice positioning; the same or a different rater should then analyse them to measure infarct volume, blinded

to the centre and group allocation. The data analysis software should ideally be semi-automated to improve reliability and speed, while allowing visual inspection of segmented lesions and manual refinement where deemed necessary. Work presented in this thesis indicates that such an approach is entirely viable, and Chapter 5 provides the grounds for its implementation.

Infarct volume should be estimated and combined in each animal group as an absolute estimate (for example, in mm³). For animals scanned at the subacute or the early chronic stage of stroke (typically 1 to 7 days post MCAO in the majority of rodent models of stroke), correction of infarct volume for brain swelling using published formulae is required (Gerriets et al., 2004). Swelling leads to apparent changes in the brain that are sufficient to indicate if a correction is necessary; effects include the shift of the midline between the hemispheres and enlargement of brain ventricles in the lesioned ipsilateral hemisphere.

Recommendations arising from this thesis are summarised in Table 7.1. Most of these could potentially favour use of MRI for the assessment of various types of abnormalities at different time points, such as areas of decreased diffusion of water and perfusion of blood using DWI and PWI respectively. Some recommendations may even apply to preclinical consortia focussing on other diseases where morphological MRI outcomes are primarily used, though relevant studies are needed to determine this. They should complement other contemporary guidelines for the conduct and reporting of experimental research (Kilkenny et al., 2010; Stout et al., 2013; Stroke Therapy Academic Industry Roundtable (STAIR), 1999). It is believed that careful consideration of these recommendations by scientists will be vital for the improvement of the validity and generalisability of experimental findings towards the identification of new and efficacious neuroprotective drugs, as well as the reduction of the resources and the number of animals used. This will also bring improvements to the validity in other post hoc exploratory analyses using acquired data and will permit exact replication of the experiments.

Table 7.1 Recommendations regarding the use of MRI for the assessment of infarct size in multicentre studies in rodent models of stroke.

Category	Recommendation
Hardware and software	<ul style="list-style-type: none"> Minimise hardware and software differences. Scanners of the same vendor and RF coils of the same type are ideal; RF reception should be performed preferably using phased array surface coils.
Quality assurance	<ul style="list-style-type: none"> Define standard QA procedures including phantom placement, slice positioning, scanning and data preparation. At least one scan per imaging plane should be acquired. Use an identical structural phantom allowing thorough assessment of geometric accuracy (distance; volume is complementary but useful). Use in vivo T2WI sequences of identical type, matrix size, FOV and slice thickness. Slice thickness should be small to reduce PVE, while other parameters should be adjusted so that SNR is 20 or higher and scanning time is 10 minutes or less. Perform reproducible positioning of slices. Analyse data in each centre using an identical automated software. Collect data at one centre and compare image quality. Hardware and/or sequence parameters (except those defining voxel size) could be modified for each centre to ensure consistency. Perform QA for optimising geometric accuracy weekly and after hardware and software updates. If the study does not involve assessment of brain size or other longitudinally stable structures, biannual QA is sufficient.
In vivo imaging	<ul style="list-style-type: none"> Estimate an appropriate sample size based on the expected difference in infarct volume between groups and the desired power. Define standard imaging procedures including animal placement, slice positioning, scanning and data preparation. Use the in vivo sequences optimised during QA. An identical number of slices covering the whole MCA territory should be acquired. Perform reproducible positioning of slices relative to an anatomical feature (bregma or the rostral end of the rhinal fissure).
Image analysis	<ul style="list-style-type: none"> Assemble all the data at one centre and perform analysis by a single experienced rater, blinded to the centre and group allocation. The data analysis software could be semi-automated to allow manual refinement of the segmented infarct. Correct infarct volume for brain swelling if scanning is performed 1–7 days after induction of stroke; Gerriets' formula is appropriate for T2WI (Gerriets et al., 2004). Visually inspect brains at all time points to determine the presence of swelling. The volume of brain hemispheres should be measured across all slices. Combine absolute corrected infarct volumes across centres to estimate treatment effect.

FOV indicates field of view; MCA, middle cerebral artery; PVE, partial volume effect; QA, quality assurance; RF, radiofrequency; SNR, signal-to-noise ratio; T2WI, T₂-weighted imaging.

7.3 Limitations and Further Considerations

Further to the discussion at the end of each chapter of the thesis, other limitations that might have confounded the results and emerging considerations are discussed in this section.

7.3.1 Systematic review and meta-analysis

The systematic review described in Chapter 3 intended to provide an initial appreciation of the variability in the use of MRI in experimental stroke and aid in the design of subsequent studies performed as part of this project. Only a small number of representative articles were reviewed, that were identified in a single database (CAMARADES). However, potentially hundreds of such studies have been performed over the years and the systematic review might have obscured the true extent of variability. Including papers from more sources could have enabled the execution of effective stratified meta-analysis or multivariate meta-regression to explore the impact of individual imaging criteria on the relationship between MRI and histological outcomes. This could have helped, for example, in the identification of the optimal parameters for the standardised sequence used in the multicentre QA study and allow sequence-specific recommendations to be made, which are missing from Table 7.1.

7.3.2 Multicentre quality assurance

Designing and carrying out a large multicentre study involving six different MRI scanners located at six different institutions across Europe is very challenging, to say the least. 7 T scanners were selected to reduce field-specific variations (particularly in SNR). However, other variations were inevitable, including differences in scanner manufacturer, gradient strength, scanner software and version, and types of imaging coils, which arguably resulted in differences in SNR between centres and consequently differences in volumetric measurements (Figures 6.4 and 6.5).

Preclinical MRI scanners are still rather scarce, therefore identifying entirely identical systems or homogenising selected systems was practically and financially impossible. Nevertheless, the primary aim of this study was to evaluate the variability in geometric accuracy in the context of experimental stroke imaging across centres and not the inherent variability between the systems themselves, which implies taking into account the use of centre-specific equipment.

To reduce the overall duration of the study, identical phantoms were constructed and sent to each participating centre to permit simultaneous scanning. I travelled and contributed to imaging in centres 'A' and 'B' (Table 6.1), but phantoms in all other centres were scanned by local staff without my supervision. Staff could have had various degrees of experience and the conduct of MRI at each centre might have deviated from the predefined QA protocol (Appendix D), contributing to variability in measured geometric errors. Instructions given were comprehensive and covered phantom placement, slice positioning, scanning, data preparation and naming, while visual assessment of images did not reveal any apparent deviations. Furthermore, imaging consisted of six scans using each pulse sequence in various imaging planes in order to perform multiple measurements per direction and improve the reliability of assessment.

The contrast solution used to fill the phantoms was a Gd-DOTA solution of an appropriate concentration to obtain a T_2 relaxation time similar to that in the rat brain at 7 T. While this contrast agent has been used before for the same purpose (Teh et al., 2016), its main effect is the reduction of T_1 relaxation time; this was not measured for the phantoms used, but was pivotal in the reduction of scanning time with the standardised protocol 'a' (Table 6.2). As a paramagnetic agent, Gd-DOTA might have contributed to an increase in magnetic susceptibility effects around the phantom, causing spatial distortions. These could have been further overstated by the cuboid shape of the phantoms themselves. An object placed in a scanner becomes magnetised and produces an induced field that perturbs B_0 . This perturbation should be as little as possible; ellipsoidal objects with uniform susceptibility produce a homogeneous

external field and are ideal for use as MRI phantoms (Schenck, 1996). Therefore, the shape of the phantom could partially explain the nonlinearities in images from Bruker systems and the bias in encoding directions in the Agilent scanner (Figures 6.6 and 6.7). A potential solution could have been to submerge the phantoms in cylindrical plastic or glass tubes filled with the same Gd-DOTA solution, but this would have resulted in aliasing artefacts for sequences using small FOV and would not allow automated analysis of images using the developed tool. None of these were a concern before the start of the multicentre study, simply because images acquired during preliminary tests using the Agilent scanner had great quality overall and were free of any nonlinearities, while the small difference in encoding directions could not be minimised by attempted sequence modifications (Figure 4.1 d). Furthermore, the cuboid phantom design permitted reproducible placement and alignment in each scanner. Despite the antithetical effect of nonlinearities on each side of the phantom, it may have contributed to measurement errors in the frequency direction. To the staff's knowledge, no software of any of the evaluated preclinical scanners applies a geometric distortion correction to the data, which poses serious questions about the underlying differences between systems or system-specific sequences that cause this discrepancy in geometric distortions.

7.3.3 Image processing

In this thesis, particular attention was paid to image analysis; automated or semi-automated tools were developed where possible to improve the accuracy and reliability of measurements and hence the validity of each study. An additional criterion was processing time; only simple and computationally inexpensive approaches were considered. Both major processing tools described in this thesis (Chapters 4 and 5) were developed for use as specified only. However, small changes in the source codes could permit analysis of similar T2WI data, such as those of other cuboid phantoms using the automated QA tool, or rodent brains with other types of lesions using the semi-automated infarct segmentation tool—tumours, global

ischaemic or multiple sclerosis lesions could be plausible examples.

Both tools were validated by comparison against gold standard manual tracing. Intra-rater reliability and inter-rater reproducibility was performed for the semi-automated infarct segmentation tool and for manual tracing. A large number of scans were used to validate the automated QA tool for the measurement of the internal dimensions of the phantom ($n = 60$; 40 measurements per direction). However, volumetric pipelines in both tools were validated based on 10 scans each; these are too few to make reliable population inferences. The considerable burden and time required for manual analysis were the main reasons for the small sample, which precluded analysis of more data within a reasonable time frame. It should be emphasised that the performance of the semi-automated infarct segmentation tool depends on the accuracy of manual analysis, the results of which were used as ground truth for optimisation. There is no specific and widely accepted approach to determine ground truth in biological data, other than by employing an experienced rater to perform manual analysis.

Correction of bias field in qualitative MRI data can be formidable, particularly when there are prominent abnormalities due to diseases that distort the global intensity homogeneity in the images. In our case, the hyperintense lesion in the ipsilateral hemisphere was causing significant correction issues when default methods were used. It was found that the contralateral hemisphere is a good reference area for the bias field correction algorithm, due to the fact that the signal in tissues within this hemisphere is largely homogeneous. However, the presence of the hyperintense ventricles was not taken into account, which may have slightly affected the correction and confounded further analyses.

The variability in infarct volume might have been overstated by the quality of images. Phantom scanning using the scanner and coils used to acquire the in vivo data point at a potentially poor quality (Figure 6.5 a). Using data from other centres could have improved the reproducibility of infarct segmentation. It should be noted that a clear

definition of the infarct boundary in MRI data is lacking from literature; descriptions of image analysis in articles are either too vague or entirely absent. The semi-automated segmentation tool was based upon criteria devised by myself based on previous observations and solely for the dataset examined, though this is representative of current preclinical stroke practice.

7.3.4 Statistics

Chapter 6 of this thesis described a simulation study assessing the impact of scanners' linear scaling errors and voxel size in the measurement of infarct volume between scanners. Distributions of infarct volumes were compared using repeated measures ANOVA, which takes into account the within-subject variability across scanners and has a higher power than cross-sectional designs; this allows detection of very small differences between means (Guo, Logan, Glueck, & Muller, 2013). It was observed that in comparisons of infarct volumes expressed as relative estimates ANOVA results were more conservative, despite an apparent difference in means. This is probably because within-subject variability was high, giving the impression that a specific bias in measurements does not exist. As done in this study, careful interpretation of the results requires consideration of the effect size of ANOVA (η^2) indicating the proportion of variance in measurements explained by variation between scanners.

The use of a small number of animals is typical in preclinical research (Vesterinen et al., 2011); inevitably, this was the case for the post hoc analyses described here. Ideally, analyses should have been performed on images acquired in a preclinical stroke study assessing the efficacy of a drug. However, collaborators could not provide such data, whereas carrying out an animal study in Edinburgh was not feasible due to the lack of research staff specialising in stroke models. Combining data from other studies was not possible either, due to the large heterogeneity in study design and imaging performance between and within centres. Regardless of the concordance between studies, no data are readily available for reuse. The

development of standardised and useful paradigms for structural and functional MRI in experimental stroke requires prospective collection and dissemination of data. Images and other recorded information could be collected using predetermined definitions and deposited in a publicly available repository. These can be used by researchers worldwide to evaluate the performance of analysis tools and perform post hoc exploratory analyses or sample size calculations, which will accelerate the establishment of standardised imaging methodologies.

7.4 Future Work

Many of the findings of this thesis were unsurprising, as the variability in MRI findings in the clinical setting is well characterised. Even so, the use of the modality in preclinical studies has been overlooked despite decades of use, rendering the quality of studies relatively poor. This thesis stresses the impact of a range of factors in the acquisition and processing of MRI data in experimental stroke, but at the same time raises additional concerns that warrant further investigation.

Instead of performing an assessment of studies employing a range of MRI methods, future systematic reviews could focus on individual imaging methods or endpoints and perform a thorough identification of relevant studies. This will increase meta-analysis flexibility and help refine the relationship between T2WI and gold standard histology, which can be decisive in the determination of optimal imaging parameters for future use.

This thesis focussed solely on the impact of T2WI variability on infarct volume measurements. Future studies could emulate the proposed approaches for multicentre QA and data simulation to a range of other MRI morphological outcomes in stroke or other diseases. An important outcome would be the increase in V_{IH} due to oedema; this is well correlated with infarct volume and is often used to assess the integrity of BBB and distinguish between the acute and subacute stages of stroke

(Gerriets et al., 2004; Macrae, 2011). Other outcomes of particular interest in rodent models of stroke include the volume of diffusion or perfusion abnormalities and the widely assessed ischaemic penumbra using the PWI/DWI mismatch (Duong, 2013; Farr & Wegener, 2010). The accuracy and concordance of preclinical scanners when these outcomes are quantified will provide further insights into the validity of the current conduct of preclinical MRI in stroke and allow identification of refinement strategies. Simulation studies could further assess the effect of variability on statistical comparisons when different animals or stroke models are used, to provide the grounds for performing power calculations in local or multicentre MRI studies. However, it could be important to first refine the simulation procedure by using high resolution reference data and applying 3D non-rigid geometric transformations. Alternatively, development of computational methods for creating digital rodent brain phantoms with realistic stroke lesions and precisely known anatomical and signal characteristics could be the way forward.

The multicentre study presented in Chapter 6 hinted that volumes of structures in 2D sMRI data are overestimated due to PVE. In addition to oedema correction, formulae should be defined to correct infarct volume for PVE, such as those used in *ex vivo* (Liu et al., 2009; Rosen & Harry, 1990) and *in vivo* determination of various biological structures (Sahin et al., 2003). This can be accomplished by scanning animals using basic 2D and high-resolution 3D sequences within the same session; infarct volume quantified in the former can then be optimised based on ground truth estimates determined with the latter, though 3D imaging with acceptable SNR and contrast could require very long acquisition times.

The semi-automated infarct segmentation tool used in this thesis lacks methods for automated segmentation of brain hemispheres and bias field correction. Future work must address this deficiency in order to bring the tool a step closer to routine preclinical utility. Accurate rodent brain detection constitutes a mission for relatively many research groups, though no available algorithm can segment each hemisphere separately or be applied on ischaemic brains (Chou et al., 2011; Oguz et al., 2014;

Zhang et al., 2009). Segmentation using computer vision approaches does not rely explicitly on assumptions of global intensity homogeneity and is a potential candidate for this task, but will require large datasets for model training and validation (Babalola et al., 2008; Wells et al., 1996).

7.5 Conclusions

Multicentre preclinical stroke studies utilising the diagnostic and prognostic strength of MRI are likely to prove vital in the translation of experimental findings to the clinic and the development of new and effective treatments. This thesis aimed to assess MRI variability in the measurement of infarct volume and highlight contributing methodological issues. Despite that the studies carried out towards this goal were simple in terms of design and implementation, the acquired data and the variability in measurements were more complex and multifaceted than expected. Findings should not be regarded as being prejudicial to experimental stroke MRI and should not discourage its use by any means. Rather, the thesis aims to enlighten researchers about factors that can influence the accuracy and comparability of infarct volume measurements and encourage them to take simple actions that can improve the use of the modality and allow effective comparison of data between centres. This thesis established strategies to improve the consistency in data acquisition and analysis, and maximize across-centre sensitivity to true biological change by identification of the sources of errors and understanding of the underlying mechanisms. Collectively, recommendations arising from this work will contribute to the replacement, reduction and refinement of the use of animals in experimental stroke (the 3Rs). However, a thorough appreciation of the issues involved requires further effort from the preclinical stroke community, but will help establish a comprehensive and rigorous multimodal MRI protocol for use in large multicentre studies and move experimental research forward.

References

- Acer, N., Sahin, B., Usanmaz, M., Tatoğlu, H., & Irmak, Z. (2008). Comparison of point counting and planimetry methods for the assessment of cerebellar volume in human using magnetic resonance imaging: a stereological study. *Surgical and Radiologic Anatomy, 30*(4), 335-339. doi:10.1007/s00276-008-0330-9
- Agilent Technologies. (2011). Digital Eddy Current Compensation. *Agilent VnmrJ 3 Imaging, User Guide* (1st ed., pp. 533-544). Santa Clara: Agilent Technologies.
- Albers, G. W., & Fisher, M. (2011). Improving the accuracy of perfusion imaging in acute ischemic stroke. *Annals of Neurology, 70*(3), 347-349. doi:10.1002/ana.22524
- Albers, G. W., Thijs, V. N., Wechsler, L., Kemp, S., Schlaug, G., Skalabrin, E., . . . Investigators, D. (2006). Magnetic resonance imaging profiles predict clinical response to early reperfusion: the diffusion and perfusion imaging evaluation for understanding stroke evolution (DEFUSE) study. *Annals of Neurology, 60*(5), 508-517. doi:10.1002/ana.20976
- American College of Radiology. (2015). *Magnetic Resonance Imaging Quality Control Manual*. Reston: American College of Radiology.
- Arganda-Carreras, I., Sorzano, C. O. S., Marabini, R., Carazo, J., Ortiz-de-Solorzano, C., & Kybic, J. (2006). Consistent and Elastic Registration of Histological Sections Using Vector-Spline Regularization. In R. R. Beichel & M. Sonka (Eds.), *Computer Vision Approaches to Medical Image Analysis* (Vol. 4241, pp. 85-95). Berlin Heidelberg: Springer.
- Ashton, E. A., Takahashi, C., Berg, M. J., Goodman, A., Totterman, S., & Ekholm, S. (2003). Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. *Journal of Magnetic Resonance Imaging, 17*(3), 300-308. doi:10.1002/jmri.10258
- Astrup, J., Siesjo, B. K., & Symon, L. (1981). Thresholds in cerebral ischemia - the ischemic penumbra. *Stroke, 12*(6), 723-725.
- Astrup, J., Symon, L., Branston, N. M., & Lassen, N. A. (1977). Cortical evoked potential and extracellular K⁺ and H⁺ at critical levels of brain ischemia. *Stroke, 8*(1), 51-57.
- Ay, H., Arsava, E. M., Vangel, M., Oner, B., Zhu, M., Wu, O., . . . Sorensen, A. G. (2008). Interexaminer difference in infarct volume measurements on MRI: a source of variance in stroke research. *Stroke, 39*(4), 1171-1176. doi:10.1161/STROKEAHA.107.502104

- Babalola, K. O., Cootes, T. F., Twining, C. J., Petrovic, V., & Taylor, C. (2008). 3D Brain Segmentation Using Active Appearance Models and Local Regressors. In D. Metaxas, L. Axel, G. Fichtinger, & G. Székely (Eds.), *Proceedings of the 11th Medical Image Computing and Computer-Assisted Intervention (MICCAI) International Conference. New York, USA* (pp. 401-408). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Bandera, E., Botteri, M., Minelli, C., Sutton, A., Abrams, K. R., & Latronico, N. (2006). Cerebral blood flow threshold of ischemic penumbra and infarct core in acute ischemic stroke: a systematic review. *Stroke*, 37(5), 1334-1339. doi:10.1161/01.STR.0000217418.29609.22
- Bar-Guy, A. (2005). RANDRAW. *MATLAB Central File Exchange*. Retrieved 30 March 2015 from <http://uk.mathworks.com/matlabcentral/fileexchange/7309-randraw>
- Barber, P. A., Zhang, J., Demchuk, A. M., Hill, M. D., & Buchan, A. M. (2001). Why are stroke patients excluded from TPA therapy? An analysis of patient eligibility. *Neurology*, 56(8), 1015-1020.
- Bardutzky, J., Shen, Q., Henninger, N., Bouley, J., Duong, T. Q., & Fisher, M. (2005). Differences in ischemic lesion evolution in different rat strains using diffusion and perfusion imaging. *Stroke*, 36(9), 2000-2005. doi:10.1161/01.STR.0000177486.85508.4d
- Baron, J. C., Bousser, M. G., Rey, A., Guillard, A., Comar, D., & Castaigne, P. (1981). Reversal of focal "misery-perfusion syndrome" by extra-intracranial arterial bypass in hemodynamic cerebral ischemia. A case study with 15O positron emission tomography. *Stroke*, 12(4), 454-459.
- Barone, F. C., Clark, R. K., Feuerstein, G., Lenkinski, R. E., & Sarkar, S. K. (1991). Quantitative comparison of magnetic resonance imaging (MRI) and histologic analyses of focal ischemic damage in the rat. *Brain Research Bulletin*, 26(2), 285-291.
- Baskerville, T. A., Macrae, I. M., Holmes, W. M., & McCabe, C. (2016). The influence of gender on 'tissue at risk' in acute stroke: A diffusion-weighted magnetic resonance imaging study in a rat model of focal cerebral ischaemia. *Journal of Cerebral Blood Flow and Metabolism*, 36(2), 381-386. doi:10.1177/0271678x15606137
- Bath, P. M., Macleod, M. R., & Green, A. R. (2009). Emulating multicentre clinical stroke trials: a new paradigm for studying novel interventions in experimental models of stroke. *International Journal of Stroke*, 4(6), 471-479. doi:10.1111/j.1747-4949.2009.00386.x
- Bauer, S., Gratz, P. P., Gralla, J., Reyes, M., & Wiest, R. (2014). Towards automatic MRI volumetry for treatment selection in acute ischemic stroke patients. *Proceedings of*

the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014, 1521-1524. doi:10.1109/EMBC.2014.6943891

- Bederson, J. B., Pitts, L. H., Germano, S. M., Nishimura, M. C., Davis, R. L., & Bartkowski, H. M. (1986). Evaluation of 2,3,5-triphenyltetrazolium chloride as a stain for detection and quantification of experimental cerebral infarction in rats. *Stroke*, 17(6), 1304-1308.
- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., . . . Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*, 276(8), 637-639.
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476), 307-310.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons.
- Brouns, R., & De Deyn, P. P. (2009). The complexity of neurobiological processes in acute ischemic stroke. *Clinical Neurology and Neurosurgery*, 111(6), 483-495. doi:10.1016/j.clineuro.2009.04.001
- Butcher, K., Parsons, M., Allport, L., Lee, S. B., Barber, P. A., Tress, B., . . . Davis, S. M. (2008). Rapid assessment of perfusion-diffusion mismatch. *Stroke*, 39(1), 75-81. doi:10.1161/STROKEAHA.107.490524
- Campbell, B. C., & Macrae, I. M. (2015). Translational perspectives on perfusion-diffusion mismatch in ischemic stroke. *International Journal of Stroke*, 10(2), 153-162. doi:10.1111/ijis.12186
- Canazza, A., Minati, L., Boffano, C., Parati, E., & Binks, S. (2014). Experimental models of brain ischemia: a review of techniques, magnetic resonance imaging, and investigational cell-based therapies. *Frontiers in Neurology*, 5, 19. doi:10.3389/fneur.2014.00019
- Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6), 679-698. doi:10.1109/TPAMI.1986.4767851
- Caramanos, Z., Fonov, V. S., Francis, S. J., Narayanan, S., Pike, G. B., Collins, D. L., & Arnold, D. L. (2010). Gradient distortions in MRI: characterizing and correcting for their effects on SIENA-generated measures of brain volume change. *Neuroimage*, 49(2), 1601-1611. doi:10.1016/j.neuroimage.2009.08.008
- Carmichael, S. T. (2005). Rodent models of focal stroke: size, mechanism, and purpose. *NeuroRx*, 2(3), 396-409. doi:10.1602/neurorx.2.3.396

- Chen, J., Liu, J., Calhoun, V. D., Arias-Vasquez, A., Zwiers, M. P., Gupta, C. N., . . . Turner, J. A. (2014). Exploration of scanning effects in multi-site structural MRI studies. *Journal of Neuroscience Methods*, 230, 37-50. doi:10.1016/j.jneumeth.2014.04.023
- Chou, N., Wu, J., Bai Bingren, J., Qiu, A., & Chuang, K. H. (2011). Robust automatic rodent brain extraction using 3-D pulse-coupled neural networks (PCNN). *IEEE Transactions on Image Processing*, 20(9), 2554-2564. doi:10.1109/TIP.2011.2126587
- Csonka, C., Kupai, K., Kocsis, G. F., Novak, G., Fekete, V., Bencsik, P., . . . Ferdinandy, P. (2010). Measurement of myocardial infarct size in preclinical studies. *Journal of Pharmacological and Toxicological Methods*, 61(2), 163-170. doi:10.1016/j.vascn.2010.02.014
- Dani, K. A., Thomas, R. G., Chappell, F. M., Shuler, K., MacLeod, M. J., Muir, K. W., & Wardlaw, J. M. (2011). Computed tomography and magnetic resonance perfusion imaging in ischemic stroke: definitions and thresholds. *Annals of Neurology*, 70(3), 384-401. doi:10.1002/ana.22500
- Dani, K. A., Thomas, R. G., Chappell, F. M., Shuler, K., Muir, K. W., & Wardlaw, J. M. (2012). Systematic review of perfusion imaging with computed tomography and magnetic resonance in acute ischemic stroke: heterogeneity of acquisition and postprocessing parameters: a translational medicine research collaboration multicentre acute stroke imaging study. *Stroke*, 43(2), 563-566. doi:10.1161/STROKEAHA.111.629923
- Davids, M., Zollner, F. G., Ruttorf, M., Nees, F., Flor, H., Schumann, G., . . . The Imagen Consortium. (2014). Fully-automated quality assurance in multi-center studies using MRI phantom measurements. *Magnetic Resonance Imaging*. doi:10.1016/j.mri.2014.01.017
- Davis, S. M., & Donnan, G. A. (2009). 4.5 hours: the new time window for tissue plasminogen activator in stroke. *Stroke*, 40(6), 2266-2267. doi:10.1161/STROKEAHA.108.544171
- Davis, S. M., Donnan, G. A., Parsons, M. W., Levi, C., Butcher, K. S., Peeters, A., . . . investigators, E. (2008). Effects of alteplase beyond 3 h after stroke in the Echoplanar Imaging Thrombolytic Evaluation Trial (EPITHET): a placebo-controlled randomised trial. *Lancet Neurology*, 7(4), 299-309. doi:10.1016/S1474-4422(08)70044-9
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177-188.
- Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3), 297-302. doi:10.2307/1932409

- Dijkhuizen, R. M., Beekwilder, J. P., van der Worp, H. B., Berkelbach van der Sprenkel, J. W., Tulleken, K. A., & Nicolay, K. (1999). Correlation between tissue depolarizations and damage in focal ischemic rat brain. *Brain Research*, 840(1-2), 194-205.
- Dijkhuizen, R. M., & Nicolay, K. (2003). Magnetic resonance imaging in experimental models of brain disorders. *Journal of Cerebral Blood Flow and Metabolism*, 23(12), 1383-1402. doi:10.1097/01.WCB.0000100341.78607.EB
- Ding, G., Jiang, Q., Li, L., Zhang, L., Zhang, Z. G., Panda, S., . . . Chopp, M. (2006). MRI of combination treatment of embolic stroke in rat with rtPA and atorvastatin. *Journal of the Neurological Sciences*, 246(1-2), 139-147. doi:10.1016/j.jns.2006.02.020
- Dirnagl, U. (2006). Bench to bedside: the quest for quality in experimental stroke research. *Journal of Cerebral Blood Flow and Metabolism*, 26(12), 1465-1478. doi:10.1038/sj.jcbfm.9600298
- Dirnagl, U., & Fisher, M. (2012). International, multicenter randomized preclinical trials in translational stroke research: It's time to act. *Journal of Cerebral Blood Flow and Metabolism*, 32(6), 933-935. doi:DOI 10.1038/jcbfm.2012.51
- Dirnagl, U., Hakim, A., Macleod, M., Fisher, M., Howells, D., Alan, S. M., . . . Meairs, S. (2013). A concerted appeal for international cooperation in preclinical stroke research. *Stroke*, 44(6), 1754-1760. doi:10.1161/STROKEAHA.113.000734
- Dirnagl, U., Iadecola, C., & Moskowitz, M. A. (1999). Pathobiology of ischaemic stroke: an integrated view. *Trends in Neurosciences*, 22(9), 391-397.
- Donnan, G. A., Fisher, M., Macleod, M., & Davis, S. M. (2008). Stroke. *Lancet*, 371(9624), 1612-1623. doi:10.1016/S0140-6736(08)60694-7
- Duong, T. Q. (2012a). Multimodal MRI of experimental stroke. *Translational Stroke Research*, 3(1), 8-15. doi:10.1007/s12975-011-0140-y
- Duong, T. Q. (2012b). Recent MRI advances in experimental stroke. *Translational Stroke Research*, 3(1), 1-3. doi:10.1007/s12975-012-0144-2
- Duong, T. Q. (2013). Magnetic resonance imaging of perfusion-diffusion mismatch in rodent and non-human primate stroke models. *Neurological Research*, 35(5), 465-469. doi:10.1179/1743132813Y.0000000211
- Durukan, A., & Tatlisumak, T. (2007). Acute ischemic stroke: overview of major experimental rodent models, pathophysiology, and therapy of focal cerebral ischemia. *Pharmacology, Biochemistry and Behavior*, 87(1), 179-197. doi:10.1016/j.pbb.2007.04.015

- Ebinger, M., De Silva, D. A., Christensen, S., Parsons, M. W., Markus, R., Donnan, G. A., & Davis, S. M. (2009). Imaging the penumbra - strategies to detect tissue at risk after ischemic stroke. *Journal of Clinical Neuroscience*, *16*(2), 178-187. doi:10.1016/j.jocn.2008.04.002
- Ewers, M., Teipel, S. J., Dietrich, O., Schonberg, S. O., Jessen, F., Heun, R., . . . Hampel, H. (2006). Multicenter assessment of reliability of cranial MRI. *Neurobiology of Aging*, *27*(8), 1051-1059. doi:10.1016/j.neurobiolaging.2005.05.032
- Fagan, A. J., Mullin, J. M., Gallagher, L., Hadley, D. M., Macrae, I. M., & Condon, B. (2008). Serial postmortem relaxometry in the normal rat brain and following stroke. *Journal of Magnetic Resonance Imaging*, *27*(3), 469-475. doi:10.1002/jmri.21246
- Farr, T. D., & Wegener, S. (2010). Use of magnetic resonance imaging to predict outcome after stroke: a review of experimental and clinical evidence. *Journal of Cerebral Blood Flow and Metabolism*, *30*(4), 703-717. doi:10.1038/jcbfm.2010.5
- Filippi, C. G., El-Ali, A. M., Miloushev, V. Z., Chow, D. S., Guo, X., & Zhao, B. (2014). Computer-Assisted Volumetric Measurement of Core Infarct Volume in Pediatric Patients: Feasibility for Clinical Use and Development of Quantitative Metrics for Outcome Prediction. *AJNR: American Journal of Neuroradiology*. doi:10.3174/ajnr.A4183
- Firbank, M. J., Harrison, R. M., Williams, E. D., & Coulthard, A. (2000). Quality assurance for MRI: practical experience. *British Journal of Radiology*, *73*(868), 376-383. doi:10.1259/bjr.73.868.10844863
- Fisher, M., Feuerstein, G., Howells, D. W., Hurn, P. D., Kent, T. A., Savitz, S. I., . . . Group, S. (2009). Update of the stroke therapy academic industry roundtable preclinical recommendations. *Stroke*, *40*(6), 2244-2250. doi:10.1161/STROKEAHA.108.541128
- Galinovic, I., Puig, J., Neeb, L., Guibernau, J., Kemmling, A., Siemonsen, S., . . . Fiebach, J. B. (2014). Visual and region of interest-based inter-rater agreement in the assessment of the diffusion-weighted imaging- fluid-attenuated inversion recovery mismatch. *Stroke*, *45*(4), 1170-1172. doi:10.1161/STROKEAHA.113.002661
- Garcia-Finana, M., Cruz-Orive, L. M., Mackay, C. E., Pakkenberg, B., & Roberts, N. (2003). Comparison of MR imaging against physical sectioning to estimate the volume of human cerebral compartments. *Neuroimage*, *18*(2), 505-516.
- Gerriets, T., Stolz, E., Walberer, M., Muller, C., Kluge, A., Bachmann, A., . . . Bachmann, G. (2004). Noninvasive quantification of brain edema and the space-

- occupying effect in rat stroke models using magnetic resonance imaging. *Stroke*, 35(2), 566-571. doi:10.1161/01.STR.0000113692.38574.57
- Goto, S., Xue, R., Sugo, N., Sawada, M., Blizzard, K. K., Poitras, M. F., . . . Hurn, P. D. (2002). Poly(ADP-ribose) polymerase impairs early and long-term experimental stroke recovery. *Stroke*, 33(4), 1101-1106.
- Gudbjartsson, H., & Patz, S. (1995). The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6), 910-914.
- Gunter, J. L., Bernstein, M. A., Borowski, B. J., Ward, C. P., Britson, P. J., Felmlee, J. P., . . . Jack, C. R. (2009). Measurement of MRI scanner performance with the ADNI phantom. *Medical Physics*, 36(6), 2193-2205. doi:10.1118/1.3116776
- Guo, Y., Logan, H. L., Glueck, D. H., & Muller, K. E. (2013). Selecting a sample size for studies with repeated measures. *BMC Medical Research Methodology*, 13, 100. doi:10.1186/1471-2288-13-100
- Hacke, W., Furlan, A. J., Al-Rawi, Y., Davalos, A., Fiebach, J. B., Gruber, F., . . . Warach, S. (2009). Intravenous desmoteplase in patients with acute ischaemic stroke selected by MRI perfusion-diffusion weighted imaging or perfusion CT (DIAS-2): a prospective, randomised, double-blind, placebo-controlled study. *Lancet Neurology*, 8(2), 141-150. doi:10.1016/S1474-4422(08)70267-9
- Harbord, R. M., & Higgins, J. P. T. (2008). Meta-regression in Stata. *Stata Journal*, 8(4), 493-519.
- Heidenreich, P. A., Trogon, J. G., Khavjou, O. A., Butler, J., Dracup, K., Ezekowitz, M. D., . . . Woo, Y. J. (2011). Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation*, 123(8), 933-944. doi:10.1161/CIR.0b013e31820a55f5
- Honma, T., Honmou, O., Iihoshi, S., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2006). Intravenous infusion of immortalized human mesenchymal stem cells protects against injury in a cerebral ischemia model in adult rat. *Experimental Neurology*, 199(1), 56-66. doi:10.1016/j.expneurol.2005.05.004
- Horita, Y., Honmou, O., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2006). Intravenous administration of glial cell line-derived neurotrophic factor gene-modified human mesenchymal stem cells protects against injury in a cerebral ischemia model in the adult rat. *Journal of Neuroscience Research*, 84(7), 1495-1504. doi:10.1002/jnr.21056
- Howells, D. W., & Macleod, M. R. (2012). Translational research. *International Journal of Stroke*, 7(5), 367-368. doi:10.1111/j.1747-4949.2012.00817_2.x

- Howells, D. W., Porritt, M. J., Rewell, S. S., O'Collins, V., Sena, E. S., van der Worp, H. B., . . . Macleod, M. R. (2010). Different strokes for different folks: the rich diversity of animal models of focal cerebral ischemia. *Journal of Cerebral Blood Flow and Metabolism*, 30(8), 1412-1431. doi:10.1038/jcbfm.2010.66
- Howells, D. W., Sena, E. S., O'Collins, V., & Macleod, M. R. (2012). Improving the efficiency of the development of drugs for stroke. *International Journal of Stroke*, 7(5), 371-377. doi:10.1111/j.1747-4949.2012.00805.x
- Ihalainen, T. M., Lonroth, N. T., Peltonen, J. I., Uusi-Simola, J. K., Timonen, M. H., Kuusela, L. J., . . . Sipila, O. E. (2011). MRI quality assurance using the ACR phantom in a multi-unit imaging center. *Acta Oncologica*, 50(6), 966-972. doi:10.3109/0284186X.2011.582515
- Ioannidis, J. P., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., . . . Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*, 383(9912), 166-175. doi:10.1016/S0140-6736(13)62227-8
- Jack, C. R., Jr., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., . . . Weiner, M. W. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4), 685-691. doi:10.1002/jmri.21049
- Jiang, Q., Zhang, Z. G., Ding, G. L., Silver, B., Zhang, L., Meng, H., . . . Chopp, M. (2006). MRI detects white matter reorganization after neural progenitor cell treatment of stroke. *Neuroimage*, 32(3), 1080-1089. doi:10.1016/j.neuroimage.2006.05.025
- Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., . . . Dale, A. (2006). Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage*, 30(2), 436-443. doi:10.1016/j.neuroimage.2005.09.046
- Kane, I., Carpenter, T., Chappell, F., Rivers, C., Armitage, P., Sandercock, P., & Wardlaw, J. (2007). Comparison of 10 different magnetic resonance perfusion imaging processing methods in acute ischemic stroke: effect on lesion size, proportion of patients with diffusion/perfusion mismatch, clinical scores, and radiologic outcomes. *Stroke*, 38(12), 3158-3164. doi:10.1161/STROKEAHA.107.483842
- Karki, K., Knight, R. A., Shen, L. H., Kapke, A., Lu, M., Li, Y., & Chopp, M. (2010). Chronic brain tissue remodeling after stroke in rat: a 1-year multiparametric magnetic resonance imaging study. *Brain Research*, 1360, 168-176. doi:10.1016/j.brainres.2010.08.098

- Kawaguchi, M., Furuya, H., & Patel, P. M. (2005). Neuroprotective effects of anesthetic agents. *Journal of Anesthesia*, 19(2), 150-156. doi:10.1007/s00540-005-0305-5
- Keller, S. S., & Roberts, N. (2009). Measurement of brain volume using MRI: software, techniques, choices and prerequisites. *Journal of Anthropological Sciences*, 87, 127-151.
- Khan, A. F., Drozd, J. J., Moreland, R. K., Ta, R. M., Borrie, M. J., Bartha, R., & The Alzheimer's Disease Neuroimaging Initiative. (2012). A novel MRI-compatible brain ventricle phantom for validation of segmentation and volumetry methods. *Journal of Magnetic Resonance Imaging*, 36(2), 476-482. doi:10.1002/jmri.23612
- Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., & Altman, D. G. (2010). Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. *PLOS Biology*, 8(6), e1000412. doi:10.1371/journal.pbio.1000412
- Kivrak, A. S., Paksoy, Y., Erol, C., Koplay, M., Ozbek, S., & Kara, F. (2013). Comparison of apparent diffusion coefficient values among different MRI platforms: a multicenter phantom study. *Diagnostic and Interventional Radiology*, 19(6), 433-437. doi:10.5152/dir.2013.13034
- Knight, R. A., Barker, P. B., Fagan, S. C., Li, Y., Jacobs, M. A., & Welch, K. M. (1998). Prediction of impending hemorrhagic transformation in ischemic stroke using magnetic resonance imaging in rats. *Stroke*, 29(1), 144-151.
- Koizumi, J., Yoshida, Y., Nakazawa, T., & Ooneda, G. (1986). Experimental studies of ischemic brain edema. I: A new experimental model of cerebral embolism in rats in which recirculation can be introduced in the ischemic area. *Japan Journal of Stroke*, 8, 1-8.
- Kranz, P. G., & Eastwood, J. D. (2009). Does diffusion-weighted imaging represent the ischemic core? An evidence-based systematic review. *AJNR: American Journal of Neuroradiology*, 30(6), 1206-1212. doi:10.3174/ajnr.A1547
- Kremer, H., & Weisstein, E. W. (2016). Isoperimetric Quotient. *MathWorld-A Wolfram Web Resource*. Retrieved 18 July 2016 from <http://mathworld.wolfram.com/IsoperimetricQuotient.html>
- Krongold, M., Almekhlafi, M. A., Demchuk, A. M., Coutts, S. B., Frayne, R., & Eilaghi, A. (2015). Final infarct volume estimation on 1-week follow-up MR imaging is feasible and is dependent on recanalization status. *Neuroimage: Clinical*, 7, 1-6. doi:10.1016/j.nicl.2014.10.010
- Kroon D-J. (2009). Bias Field Corrected Fuzzy C-Means. *MATLAB Central File Exchange*. Retrieved 12 May 2014 from

<http://uk.mathworks.com/matlabcentral/fileexchange/25712-bias-field-corrected-fuzzy-c-means>

- Kruggel, F., Turner, J., Muftuler, L. T., & The Alzheimer's Disease Neuroimaging Initiative. (2010). Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. *Neuroimage*, 49(3), 2123-2133. doi:10.1016/j.neuroimage.2009.11.006
- Kudo, M., Aoyama, A., Ichimori, S., & Fukunaga, N. (1982). An animal model of cerebral infarction. Homologous blood clot emboli in rats. *Stroke*, 13(4), 505-508.
- Landis, S. C., Amara, S. G., Asadullah, K., Austin, C. P., Blumenstein, R., Bradley, E. W., . . . Silberberg, S. D. (2012). A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*, 490(7419), 187-191. doi:10.1038/nature11556
- Leach, M. J., Swan, J. H., Eisenthal, D., Dopson, M., & Nobbs, M. (1993). BW619C89, a glutamate release inhibitor, protects against focal cerebral ischemic damage. *Stroke*, 24(7), 1063-1067.
- Lee, Y. C., Fullerton, G. D., Baiu, C., Lescrenier, M. G., & Goins, B. A. (2011). Preclinical multimodality phantom design for quality assurance of tumor size measurement. *BMC Medical Physics*, 11, 1. doi:10.1186/1756-6649-11-1
- Leigh, R., & Krakauer, J. W. (2014). MRI-guided selection of patients for treatment of acute ischemic stroke. *Current Opinion in Neurology*, 27(4), 425-433. doi:10.1097/WCO.0000000000000110
- Leithner, C., Fuchtemeier, M., Jorks, D., Mueller, S., Dirnagl, U., & Royl, G. (2015). Infarct Volume Prediction by Early Magnetic Resonance Imaging in a Murine Stroke Model Depends on Ischemia Duration and Time of Imaging. *Stroke*, 46(11), 3249-3259. doi:10.1161/STROKEAHA.114.007832
- Lestro Henriques, I., Gutierrez-Fernandez, M., Rodriguez-Frutos, B., Ramos-Cejudo, J., Otero-Ortega, L., Navarro Hernanz, T., . . . Diez-Tejedor, E. (2015). Intralesional Patterns of MRI ADC Maps Predict Outcome in Experimental Stroke. *Cerebrovascular Diseases*, 39(5-6), 293-301. doi:10.1159/000381727
- Li, L., Jiang, Q., Ding, G., Zhang, L., Zhang, Z. G., Li, Q., . . . Chopp, M. (2009). MRI identification of white matter reorganization enhanced by erythropoietin treatment in a rat model of focal ischemia. *Stroke*, 40(3), 936-941. doi:10.1161/STROKEAHA.108.527713
- Li, W. Y., Choi, Y. J., Lee, P. H., Huh, K., Kang, Y. M., Kim, H. S., . . . Bang, O. Y. (2008). Mesenchymal stem cells for ischemic stroke: changes in effects after ex vivo culturing. *Cell Transplantation*, 17(9), 1045-1059.

- Lin, T. N., He, Y. Y., Wu, G., Khan, M., & Hsu, C. Y. (1993). Effect of brain edema on infarct volume in a focal cerebral ischemia model in rats. *Stroke*, *24*(1), 117-121.
- Liu, H., Honmou, O., Harada, K., Nakamura, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2006). Neuroprotection by PIGF gene-modified human mesenchymal stem cells after cerebral ischaemia. *Brain*, *129*(Pt 10), 2734-2745. doi:10.1093/brain/awl207
- Liu, S., Zhen, G., Meloni, B. P., Campbell, K., & Winn, H. R. (2009). Rodent Stroke Model Guidelines for Preclinical Stroke Trials (1st Edition). *Journal of Experimental Stroke & Translational Medicine*, *2*(2), 2-27.
- Lloyd, S. P. (1982). Least-Squares Quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129-137. doi:Doi 10.1109/Tit.1982.1056489
- Longa, E. Z., Weinstein, P. R., Carlson, S., & Cummins, R. (1989). Reversible middle cerebral artery occlusion without craniectomy in rats. *Stroke*, *20*(1), 84-91.
- Loubinoux, I., Volk, A., Borredon, J., Guirimand, S., Tiffon, B., Seylaz, J., & Meric, P. (1997). Spreading of vasogenic edema and cytotoxic edema assessed by quantitative diffusion and T2 magnetic resonance imaging. *Stroke*, *28*(2), 419-426; discussion 426-417.
- Luengo-Fernandez, R., Leal, J., & Gray, A. (2015). UK research spend in 2008 and 2012: comparing stroke, cancer, coronary heart disease and dementia. *BMJ Open*, *5*(4), e006648. doi:10.1136/bmjopen-2014-006648
- Luft, A. R., Skalej, M., Welte, D., Kolb, R., & Klose, U. (1996). Reliability and exactness of MRI-based volumetry: a phantom study. *Journal of Magnetic Resonance Imaging*, *6*(4), 700-704.
- Macleod, M. R., Fisher, M., O'Collins, V., Sena, E. S., Dirnagl, U., Bath, P. M., . . . Howells, D. W. (2009). Good laboratory practice: preventing introduction of bias at the bench. *Stroke*, *40*(3), e50-52. doi:10.1161/STROKEAHA.108.525386
- Macleod, M. R., O'Collins, T., Howells, D. W., & Donnan, G. A. (2004). Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke*, *35*(5), 1203-1208. doi:10.1161/01.STR.0000125719.25853.20
- Macleod, M. R., van der Worp, H. B., Sena, E. S., Howells, D. W., Dirnagl, U., & Donnan, G. A. (2008). Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke*, *39*(10), 2824-2829. doi:10.1161/STROKEAHA.108.515957
- Macrae, I. M. (2011). Preclinical stroke research – advantages and disadvantages of the most common rodent models of focal ischaemia. *British Journal of Pharmacology*, *164*(4), 1062-1078. doi:10.1111/j.1476-5381.2011.01398.x

- Macrae, I. M., Robinson, M. J., Graham, D. I., Reid, J. L., & McCulloch, J. (1993). Endothelin-1-induced reductions in cerebral blood flow: dose dependency, time course, and neuropathological consequences. *Journal of Cerebral Blood Flow and Metabolism*, *13*(2), 276-284. doi:10.1038/jcbfm.1993.34
- Marchal, G., Serrati, C., Rioux, P., Petit-Taboue, M. C., Viader, F., de la Sayette, V., . . . et al. (1993). PET imaging of cerebral perfusion and oxygen consumption in acute ischaemic stroke: relation to outcome. *Lancet*, *341*(8850), 925-927.
- Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLOS Medicine*, *3*(11), e442. doi:10.1371/journal.pmed.0030442
- Maysami, S., Wong, R., Pradillo, J. M., Denes, A., Dhungana, H., Malm, T., . . . Allan, S. M. (2015). A cross-laboratory preclinical study on the effectiveness of interleukin-1 receptor antagonist in stroke. *Journal of Cerebral Blood Flow and Metabolism*. doi:10.1177/0271678x15606714
- McBride, D. W., Klebe, D., Tang, J., & Zhang, J. H. (2015). Correcting for Brain Swelling's Effects on Infarct Volume Calculation After Middle Cerebral Artery Occlusion in Rats. *Translational Stroke Research*, *6*(4), 323-338. doi:10.1007/s12975-015-0400-3
- McCabe, C., Gallagher, L., Gsell, W., Graham, D., Dominiczak, A. F., & Macrae, I. M. (2009). Differences in the evolution of the ischemic penumbra in stroke-prone spontaneously hypertensive and Wistar-Kyoto rats. *Stroke*, *40*(12), 3864-3868. doi:10.1161/STROKEAHA.109.559021
- McRobbie, D. W., Moore, E. A., Graves, M. J., & Prince, M. R. (2006). *MRI from Picture to Proton* (2nd ed.). Cambridge: Cambridge University Press.
- Meng, X., Fisher, M., Shen, Q., Sotak, C. H., & Duong, T. Q. (2004). Characterizing the diffusion/perfusion mismatch in experimental focal cerebral ischemia. *Annals of Neurology*, *55*(2), 207-212. doi:10.1002/ana.10803
- Milidonis, X., Marshall, I., Macleod, M. R., & Sena, E. S. (2015). Magnetic resonance imaging in experimental stroke and comparison with histology: systematic review and meta-analysis. *Stroke*, *46*(3), 843-851. doi:10.1161/STROKEAHA.114.007560
- Modo, M., Beech, J. S., Meade, T. J., Williams, S. C., & Price, J. (2009). A chronic 1 year assessment of MRI contrast agent-labelled neural stem cell transplants in stroke. *Neuroimage*, *47 Suppl 2*, T133-142. doi:10.1016/j.neuroimage.2008.06.017
- Moseley, M. E., Mintorovitch, J., Cohen, Y., Asgari, H. S., Derugin, N., Norman, D., & Kucharczyk, J. (1990). Early detection of ischemic injury: comparison of

spectroscopy, diffusion-, T2-, and magnetic susceptibility-weighted MRI in cats. *Acta Neurochirurgica. Supplementum*, 51, 207-209.

- Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., . . . Turner, M. B. (2015). Heart Disease and Stroke Statistics—2016 Update: A Report From the American Heart Association. *Circulation*. doi:10.1161/cir.0000000000000350
- Neumann-Haefelin, T., Kastrup, A., de Crespigny, A., Yenari, M. A., Ringer, T., Sun, G. H., & Moseley, M. E. (2000). Serial MRI after transient focal cerebral ischemia in rats: dynamics of tissue injury, blood-brain barrier damage, and edema formation. *Stroke*, 31(8), 1965-1972; discussion 1972-1963.
- Nieuwstadt, H. A., Kassar, Z. A., van der Lugt, A., Breeuwer, M., van der Steen, A. F., Wentzel, J. J., & Gijzen, F. J. (2015). A computer-simulation study on the effects of MRI voxel dimensions on carotid plaque lipid-core and fibrous cap segmentation and stress modeling. *PLOS One*, 10(4), e0123031. doi:10.1371/journal.pone.0123031
- Noebauer-Huhmann, I. M., Kraff, O., Juras, V., Szomolanyi, P., Maderwald, S., Mlynarik, V., . . . Trattnig, S. (2008). MR Contrast Media at 7Tesla - Preliminary Study on Relaxivities. *Proceedings of the 16th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, Toronto, Canada.
- Nomura, T., Honmou, O., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2005). I.V. infusion of brain-derived neurotrophic factor gene-modified human mesenchymal stem cells protects against injury in a cerebral ischemia model in adult rat. *Neuroscience*, 136(1), 161-169. doi:10.1016/j.neuroscience.2005.06.062
- O'Callaghan, J., Wells, J., Richardson, S., Holmes, H., Yu, Y., Walker-Samuel, S., . . . Lythgoe, M. F. (2014). Is your system calibrated? MRI gradient system calibration for pre-clinical, high-resolution imaging. *PLOS One*, 9(5), e96568. doi:10.1371/journal.pone.0096568
- O'Collins, V. E., Macleod, M. R., Donnan, G. A., Horkey, L. L., van der Worp, B. H., & Howells, D. W. (2006). 1,026 experimental treatments in acute stroke. *Annals of Neurology*, 59(3), 467-477. doi:10.1002/ana.20741
- Oguz, I., Zhang, H., Rumple, A., & Sonka, M. (2014). RATS: Rapid Automatic Tissue Segmentation in rodent brain MRI. *Journal of Neuroscience Methods*, 221, 175-182. doi:10.1016/j.jneumeth.2013.09.021
- Omori, Y., Honmou, O., Harada, K., Suzuki, J., Houkin, K., & Kocsis, J. D. (2008). Optimization of a therapeutic protocol for intravenous injection of human mesenchymal stem cells after cerebral ischemia in adult rats. *Brain Research*, 1236, 30-38. doi:10.1016/j.brainres.2008.07.116

- Onda, T., Honmou, O., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2008). Therapeutic benefits by human mesenchymal stem cells (hMSCs) and Ang-1 gene-modified hMSCs after cerebral ischemia. *Journal of Cerebral Blood Flow and Metabolism*, 28(2), 329-340. doi:10.1038/sj.jcbfm.9600527
- Osswald, T. A., & Menges, G. (2010). Magnetic Properties. *Material Science of Polymers for Engineers* (3rd ed., pp. 510). Munich: Hanser.
- Panych, L. P., Chiou, J. G., Qin, L., Kimbrell, V. L., Bussolari, L., & Mulkern, R. V. (2015). On replacing the manual measurement of ACR phantom images performed by MRI technologists with an automated measurement approach. *Journal of Magnetic Resonance Imaging*. doi:10.1002/jmri.25052
- Paxinos, G., & Watson, C. (1997). *The rat brain in stereotaxic coordinates* (Compact third edition. ed.). San Diego: Academic Press.
- Peeling, J., Corbett, D., Del Bigio, M. R., Hudzik, T. J., Campbell, T. M., & Palmer, G. C. (2001). Rat middle cerebral artery occlusion: correlations between histopathology, T2-weighted magnetic resonance imaging, and behavioral indices. *Journal of Stroke and Cerebrovascular Diseases*, 10(4), 166-177. doi:10.1053/jscd.2001.26865
- Perel, P., Roberts, I., Sena, E., Wheble, P., Briscoe, C., Sandercock, P., . . . Khan, K. S. (2007). Comparison of treatment effects between animal experiments and clinical trials: systematic review. *BMJ*, 334(7586), 197. doi:10.1136/bmj.39048.407928.BE
- Pham, M., Kleinschnitz, C., Helluy, X., Bartsch, A. J., Austinat, M., Behr, V. C., . . . Bendszus, M. (2010). Enhanced cortical reperfusion protects coagulation factor XII-deficient mice from ischemic stroke as revealed by high-field MRI. *Neuroimage*, 49(4), 2907-2914. doi:10.1016/j.neuroimage.2009.11.061
- Price, R. R., Axel, L., Morgan, T., Newman, R., Perman, W., Schneiders, N., . . . Thomas, S. R. (1990). Quality assurance methods and phantoms for magnetic resonance imaging: report of AAPM nuclear magnetic resonance Task Group No. 1. *Medical Physics*, 17(2), 287-295.
- Prodanov, D., & Verstreken, K. (2012). Automated Segmentation and Morphometry of Cell and Tissue Structures. Selected Algorithms in ImageJ. In B. Schaller (Ed.), *Molecular Imaging* (pp. 183-208). Rijeka: InTech.
- Quercioli, F., Tiribilli, B., Mannoni, A., & Acciai, S. (1998). Optomechanics with LEGO. *Applied Optics*, 37(16), 3408-3416.
- Redgrave, J. N., Coutts, S. B., Schulz, U. G., Briley, D., & Rothwell, P. M. (2007). Systematic review of associations between the presence of acute ischemic lesions on diffusion-weighted imaging and clinical predictors of early stroke risk after

transient ischemic attack. *Stroke*, 38(5), 1482-1488.
doi:10.1161/STROKEAHA.106.477380

- Regan, H. K., Detwiler, T. J., Huang, J. C., Lynch, J. J., & Regan, C. P. (2007). An improved automated method to quantitate infarct volume in triphenyltetrazolium stained rat brain sections. *Journal of Pharmacological and Toxicological Methods*, 56(3), 339-343. doi:10.1016/j.vascn.2007.05.005
- Reid, E., Graham, D., Lopez-Gonzalez, M. R., Holmes, W. M., Macrae, I. M., & McCabe, C. (2012). Penumbra detection using PWI/DWI mismatch MRI in a rat stroke model with and without comorbidity: comparison of methods. *Journal of Cerebral Blood Flow and Metabolism*, 32(9), 1765-1777. doi:10.1038/jcbfm.2012.69
- Ren, H., Shen, Q., Bardutzky, J., Fisher, M., & Duong, T. Q. (2004). Partial-volume effect on ischemic tissue-fate delineation using quantitative perfusion and diffusion imaging on a rat stroke model. *Magnetic Resonance in Medicine*, 52(6), 1328-1335. doi:10.1002/mrm.20299
- Richter, S. H., Garner, J. P., Auer, C., Kunert, J., & Wurbel, H. (2010). Systematic variation improves reproducibility of animal experiments. *Nature Methods*, 7(3), 167-168. doi:10.1038/nmeth0310-167
- Richter, S. H., Garner, J. P., & Wurbel, H. (2009). Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nature Methods*, 6(4), 257-261. doi:10.1038/nmeth.1312
- Rivers, C. S., & Wardlaw, J. M. (2005). What has diffusion imaging in animals told us about diffusion imaging in patients with ischaemic stroke? *Cerebrovascular Diseases*, 19(5), 328-336. doi:10.1159/000084691
- Rivers, C. S., Wardlaw, J. M., Armitage, P. A., Bastin, M. E., Hand, P. J., & Dennis, M. S. (2007). Acute ischemic stroke lesion measurement on diffusion-weighted imaging--important considerations in designing acute stroke trials with magnetic resonance imaging. *Journal of Stroke and Cerebrovascular Diseases*, 16(2), 64-70. doi:10.1016/j.jstrokecerebrovasdis.2006.11.003
- Robinson, R. G., Shoemaker, W. J., Schlumpf, M., Valk, T., & Bloom, F. E. (1975). Effect of experimental cerebral infarction in rat brain on catecholamines and behaviour. *Nature*, 255(5506), 332-334.
- Rosen, G. D., & Harry, J. D. (1990). Brain volume estimation from serial section measurements: a comparison of methodologies. *Journal of Neuroscience Methods*, 35(2), 115-124.
- Rothwell, P. M. (2001). The high cost of not funding stroke research: a comparison with heart disease and cancer. *Lancet*, 357(9268), 1612-1616.

- Sacco, R. L., Kasner, S. E., Broderick, J. P., Caplan, L. R., Connors, J. J., Culebras, A., . . . Vinters, H. V. (2013). An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 44(7), 2064-2089. doi:10.1161/STR.0b013e318296aeca
- Sahin, B., Emirzeoglu, M., Uzun, A., Incesu, L., Bek, Y., Bilgic, S., & Kaplan, S. (2003). Unbiased estimation of the liver volume by the Cavalieri principle using magnetic resonance images. *European Journal of Radiology*, 47(2), 164-170.
- Salvado, O. (2006). Intensity inhomogeneity correction. *MATLAB Central File Exchange*. Retrieved 12 May 2014 from <http://uk.mathworks.com/matlabcentral/fileexchange/13411-intensity-inhomogeneity-correction>
- Saunders, D. E., Clifton, A. G., & Brown, M. M. (1995). Measurement of infarct size using MRI predicts prognosis in middle cerebral artery infarction. *Stroke*, 26(12), 2272-2276.
- Schabitz, W. R., Hoffmann, T. T., Heiland, S., Kollmar, R., Bardutzky, J., Sommer, C., & Schwab, S. (2001). Delayed neuroprotective effect of insulin-like growth factor- α after experimental transient focal cerebral ischemia monitored with mri. *Stroke*, 32(5), 1226-1233.
- Schenck, J. F. (1996). The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. *Medical Physics*, 23(6), 815-850.
- Schiemanck, S. K., Kwakkel, G., Post, M. W., & Prevo, A. J. (2006). Predictive value of ischemic lesion volume assessed with magnetic resonance imaging for neurological deficits and functional outcome poststroke: A critical review of the literature. *Neurorehabilitation and Neural Repair*, 20(4), 492-502. doi:10.1177/1545968306289298
- Schwamm, L. H., Ali, S. F., Reeves, M. J., Smith, E. E., Saver, J. L., Messe, S., . . . Fonarow, G. C. (2013). Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With The Guidelines-Stroke hospitals. *Circulation: Cardiovascular Quality and Outcomes*, 6(5), 543-549. doi:10.1161/CIRCOUTCOMES.111.000303
- Sena, E., van der Worp, H. B., Howells, D., & Macleod, M. (2007). How can we improve the pre-clinical development of drugs for stroke? *Trends in Neurosciences*, 30(9), 433-439. doi:10.1016/j.tins.2007.06.009
- Sena, E. S. (2010). *Systematic Review and Meta-Analysis of Animal Models of Acute Ischaemic Stroke*. (PhD Thesis), The University of Edinburgh.

- Sena, E. S., Currie, G. L., McCann, S. K., Macleod, M. R., & Howells, D. W. (2014). Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. *Journal of Cerebral Blood Flow and Metabolism*. doi:10.1038/jcbfm.2014.28
- Sena, E. S., van der Worp, H. B., Bath, P. M., Howells, D. W., & Macleod, M. R. (2010). Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLOS Biology*, 8(3), e1000344. doi:10.1371/journal.pbio.1000344
- Shackelford, J. F., Han, Y.-H., Kim, S., & Kwon, S.-H. (2015). *Polymers In: CRC Materials Science and Engineering Handbook* (4th ed., pp. 461-559). Boca Raton: CRC Press.
- Sharkey, J., Ritchie, I. M., & Kelly, P. A. (1993). Perivascular microapplication of endothelin-1: a new model of focal cerebral ischaemia in the rat. *Journal of Cerebral Blood Flow and Metabolism*, 13(5), 865-871. doi:10.1038/jcbfm.1993.108
- Shen, Q., Meng, X., Fisher, M., Sotak, C. H., & Duong, T. Q. (2003). Pixel-by-pixel spatiotemporal progression of focal ischemia derived using quantitative perfusion and diffusion imaging. *Journal of Cerebral Blood Flow and Metabolism*, 23(12), 1479-1488. doi:10.1097/01.WCB.0000100064.36077.03
- Sicard, K. M., & Fisher, M. (2009). Animal models of focal brain ischemia. *Experimental & Translational Stroke Medicine*, 1, 7. doi:10.1186/2040-7378-1-7
- Sommer, C. (2010). Histology and infarct volume determination. In U. Dirnagl (Ed.), *Rodent Models of Stroke* (pp. 213-226). New York: Humana Press.
- Stout, D., Berr, S. S., LeBlanc, A., Kalen, J. D., Osborne, D., Price, J., . . . Wall, J. (2013). Guidance for methods descriptions used in preclinical imaging papers. *Molecular Imaging*, 12(7), 1-15.
- Stroke Therapy Academic Industry Roundtable (STAIR). (1999). Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke*, 30(12), 2752-2758.
- Swanson, R. A., Morton, M. T., Tsao-Wu, G., Savalos, R. A., Davidson, C., & Sharp, F. R. (1990). A semiautomated method for measuring brain infarct volume. *Journal of Cerebral Blood Flow and Metabolism*, 10(2), 290-293. doi:10.1038/jcbfm.1990.47
- Swanson, R. A., & Sharp, F. R. (1994). Infarct measurement methodology. *Journal of Cerebral Blood Flow and Metabolism*, 14(4), 697-698. doi:10.1038/jcbfm.1994.88
- Takao, H., Hayashi, N., & Ohtomo, K. (2011). Effect of scanner in longitudinal studies of brain volume changes. *Journal of Magnetic Resonance Imaging*, 34(2), 438-444. doi:10.1002/jmri.22636

- Takao, H., Hayashi, N., & Ohtomo, K. (2013). Effects of the use of multiple scanners and of scanner upgrade in longitudinal voxel-based morphometry studies. *Journal of Magnetic Resonance Imaging*, 38(5), 1283-1291. doi:10.1002/jmri.24038
- Tamura, A., Graham, D. I., McCulloch, J., & Teasdale, G. M. (1981). Focal cerebral ischaemia in the rat: 1. Description of technique and early neuropathological consequences following middle cerebral artery occlusion. *Journal of Cerebral Blood Flow and Metabolism*, 1(1), 53-60. doi:10.1038/jcbfm.1981.6
- Teh, I., Maguire, M. L., & Schneider, J. E. (2016). Efficient gradient calibration based on diffusion MRI. *Magnetic Resonance in Medicine*. doi:10.1002/mrm.26105
- Thevenaz, P., Ruttimann, U. E., & Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Transactions on Image Processing*, 7(1), 27-41. doi:10.1109/83.650848
- Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, 18(20), 2693-2708.
- Thrippleton, M. J., Munro, K. I., McKillop, G., Newby, D. E., Marshall, I., Roberts, N., & Critchley, H. O. (2014). Unbiased and Efficient Estimation of the Volume of the Fibroid Uterus Using the Cavalieri Method and Magnetic Resonance Imaging. *Reproductive Sciences*. doi:10.1177/1933719114553451
- Toyama, K., Honmou, O., Harada, K., Suzuki, J., Houkin, K., Hamada, H., & Kocsis, J. D. (2009). Therapeutic benefits of angiogenetic gene-modified human mesenchymal stem cells after cerebral ischemia. *Experimental Neurology*, 216(1), 47-55. doi:10.1016/j.expneurol.2008.11.010
- Tureyen, K., Vemuganti, R., Sailor, K. A., & Dempsey, R. J. (2004). Infarct volume quantification in mouse focal cerebral ischemia: a comparison of triphenyltetrazolium chloride and cresyl violet staining techniques. *Journal of Neuroscience Methods*, 139(2), 203-207. doi:10.1016/j.jneumeth.2004.04.029
- Tustison, N., & Gee, J. (2010). N4ITK: Nick's N3 ITK Implementation For MRI Bias Field Correction. *The Insight Journal*. doi:http://hdl.handle.net/10380/3053
- Ukai, R., Honmou, O., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2007). Mesenchymal stem cells derived from peripheral blood protects against ischemia. *Journal of Neurotrauma*, 24(3), 508-520. doi:10.1089/neu.2006.0161
- van der Worp, H. B., Howells, D. W., Sena, E. S., Porritt, M. J., Rewell, S., O'Collins, V., & Macleod, M. R. (2010). Can animal models of disease reliably inform human studies? *PLOS Medicine*, 7(3), e1000245. doi:10.1371/journal.pmed.1000245

- van der Worp, H. B., & Macleod, M. R. (2011). Preclinical studies of human disease: time to take methodological quality seriously. *Journal of Molecular and Cellular Cardiology*, 51(4), 449-450. doi:10.1016/j.yjmcc.2011.04.008
- Veldhuis, W. B., van der Stelt, M., Delmas, F., Gillet, B., Veldink, G. A., Vliegenthart, J. F., . . . Bar, P. R. (2003). In vivo excitotoxicity induced by ouabain, a Na⁺/K⁺-ATPase inhibitor. *Journal of Cerebral Blood Flow and Metabolism*, 23(1), 62-74.
- Vesterinen, H. M., Egan, K., Deister, A., Schlattmann, P., Macleod, M. R., & Dirnagl, U. (2011). Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the *Journal of Cerebral Blood Flow and Metabolism*. *Journal of Cerebral Blood Flow and Metabolism*, 31(4), 1064-1072. doi:10.1038/jcbfm.2010.217
- Vesterinen, H. M., Sena, E. S., Egan, K. J., Hirst, T. C., Churolov, L., Currie, G. L., . . . Macleod, M. R. (2014). Meta-analysis of data from animal studies: a practical guide. *Journal of Neuroscience Methods*, 221(0), 92-102. doi:10.1016/j.jneumeth.2013.09.010
- Walberer, M., Stolz, E., Muller, C., Friedrich, C., Rottger, C., Blaes, F., . . . Gerriets, T. (2006). Experimental stroke: ischaemic lesion volume and oedema formation differ among rat strains (a comparison between Wistar and Sprague-Dawley rats using MRI). *Laboratory Animals*, 40(1), 1-8. doi:10.1258/002367706775404426
- Wang, D., Strugnell, W., Cowin, G., Doddrell, D. M., & Slaughter, R. (2004). Geometric distortion in clinical MRI systems Part I: evaluation using a 3D phantom. *Magnetic Resonance Imaging*, 22(9), 1211-1221. doi:10.1016/j.mri.2004.08.012
- Wapler, M. C., Leupold, J., Dragonu, I., von Elverfeld, D., Zaitsev, M., & Wallrabe, U. (2014). Magnetic properties of materials for MR engineering, micro-MR and beyond. *Journal of Magnetic Resonance*, 242, 233-242. doi:10.1016/j.jmr.2014.02.005
- Wardlaw, J. M. (2010). Neuroimaging in acute ischaemic stroke: insights into unanswered questions of pathophysiology. *Journal of Internal Medicine*, 267(2), 172-190. doi:10.1111/j.1365-2796.2009.02200.x
- Warlow, C., Sudlow, C., Dennis, M., Wardlaw, J., & Sandercock, P. (2003). Stroke. *Lancet*, 362(9391), 1211-1224. doi:10.1016/S0140-6736(03)14544-8
- Wegener, S., Weber, R., Ramos-Cabrera, P., Uhlenkueken, U., Sprenger, C., Wiedermann, D., . . . Hoehn, M. (2006). Temporal profile of T2-weighted MRI distinguishes between pannecrosis and selective neuronal death after transient focal cerebral ischemia in the rat. *Journal of Cerebral Blood Flow and Metabolism*, 26(1), 38-47. doi:10.1038/sj.jcbfm.9600166

- Wells, W. M., Grimson, W. L., Kikinis, R., & Jolesz, F. A. (1996). Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4), 429-442. doi:10.1109/42.511747
- Wintermark, M., Albers, G. W., Alexandrov, A. V., Alger, J. R., Bammer, R., Baron, J. C., . . . Warach, S. (2008). Acute stroke imaging research roadmap. *Stroke*, 39(5), 1621-1628. doi:10.1161/STROKEAHA.107.512319
- Wintermark, M., Sanelli, P. C., Albers, G. W., Bello, J., Derdeyn, C., Hetts, S. W., . . . Meltzer, C. C. (2013). Imaging Recommendations for Acute Stroke and Transient Ischemic Attack Patients: A Joint Statement by the American Society of Neuroradiology, the American College of Radiology, and the Society of NeuroInterventional Surgery. *AJNR: American Journal of Neuroradiology*, 34(11), E117-127. doi:10.3174/ajnr.A3690
- Wurnig, M. C., Rath, J., Klinger, N., Hollinger, I., Geissler, A., Fischmeister, F. P., . . . Beisteiner, R. (2013). Variability of clinical functional MR imaging results: a multicenter study. *Radiology*, 268(2), 521-531. doi:10.1148/radiol.13121357
- Yoshimaru, E., Totenhagen, J., Alexander, G. E., & Trouard, T. P. (2014). Design, manufacture, and analysis of customized phantoms for enhanced quality control in small animal MRI systems. *Magnetic Resonance in Medicine*, 71(2), 880-884. doi:10.1002/mrm.24678
- Zanier, E. R., Pischiutta, F., Villa, P., Paladini, A., Montinaro, M., Micotti, E., . . . De Simoni, M. G. (2013). Six-month ischemic mice show sensorimotor and cognitive deficits associated with brain atrophy and axonal disorganization. *CNS Neuroscience & Therapeutics*, 19(9), 695-704. doi:10.1111/cns.12128
- Zhang, S. T., Zhou, J. H., Wang, X. X., Chang, S., Metaxas, D. N., Pappas, G., . . . Kambhamettu, C. (2009). 3D Segmentation of Rodent Brains Using Deformable Models and Variational Methods. *Proceedings of the Annual International Conference of the IEEE Computer Vision and Pattern Recognition Society*, 157-163.
- Zille, M., Farr, T. D., Przesdzing, I., Muller, J., Sommer, C., Dirnagl, U., & Wunder, A. (2012). Visualizing cell death in experimental focal cerebral ischemia: promises, problems, and perspectives. *Journal of Cerebral Blood Flow and Metabolism*, 32(2), 213-231. doi:10.1038/jcbfm.2011.150

Appendix A

Meta-Analysis Methodology

The formulae described below are used for the meta-analysis of continuous data, and were used in Chapter 3 to analyse infarct size measurements extracted from systematically reviewed studies. They were adapted from Vesterinen et al. (2014) and Borenstein, Hedges, Higgins, and Rothstein (2009), who provide guidance for meta-analysis of additional types of data.

A.1 Estimating Effect Size

A meta-analysis aims to pool systematically extracted data by weighting individual study measurements (in this case the mean infarct volume) according to the sample size of each study and the variance of the measurement. Each study may compare measurements between various treatment and control groups to determine the ES representing the impact of a treatment on measured outcome. For deriving this, the correct number of animals used in each comparison is calculated according to

$$n'_c = \frac{n_c}{\text{number of treatment groups served}} \quad (\text{A.1})$$

where n_c is the total number of animals in the control group. The total number of animals involved in the comparison is then

$$N = n_{rx} + n'_c \quad (\text{A.2})$$

where n_{rx} is the number of animals in the treatment group. If a study reports the SEM of a measurement for the control or treatment groups, SE_c or SE_{rx} respectively, this is converted to SD according to

$$SD_c = SE_c \sqrt{n'_c} \quad \text{and} \quad SD_{rx} = SE_{rx} \sqrt{n_{rx}} \quad (\text{A.3})$$

respectively. The ES of a comparison is estimated as an NMD, describing the percentage improvement in the treated group compared to the control group:

$$ES_i = \frac{(\bar{x}_c - \bar{x}_{sham}) - (\bar{x}_{rx} - \bar{x}_{sham})}{\bar{x}_c - \bar{x}_{sham}} 100\% \quad (\text{A.4})$$

where \bar{x}_c and \bar{x}_{rx} are the mean values of measurements in the control and treatment groups respectively, \bar{x}_{sham} is the mean value for a normal and untreated group of animals (sham group), if used, and i the index of each individual comparison. NMD was preferred over alternative approaches used in—the highly heterogeneous—studies in animal models, such as the simple mean difference or the standardised mean difference, as it allows outcomes measured on different scales to be combined and it is not given in units of SD. NMD values range usually between -100% and 100%, denoting doubling (worsening of outcome) or halving (improvement of outcome) of the mean infarct size in the treated group compared to the control group, respectively. The standard deviations for both groups are also expressed as a percentage using the following equations:

$$SD_{c*} = \frac{SD_c}{\bar{x}_c - \bar{x}_{sham}} 100\% \quad \text{and} \quad SD_{rx*} = \frac{SD_{rx}}{\bar{x}_{rx} - \bar{x}_{sham}} 100\% \quad (\text{A.5})$$

which are used to estimate the SEM of the ES according to

$$SE_i = \sqrt{\frac{SD_{c*}^2}{n'_c} + \frac{SD_{rx*}^2}{n_{rx}}} \quad (\text{A.6})$$

A.2 Pooling Effect Sizes

Each ES is weighted based on its SEM, which effectively represents the precision of each estimate; the less variance in the estimate the more it will contribute to the pooled effect size. Each individual weight is calculated by

$$W_i = \frac{1}{SE_i^2} \quad (\text{A.7})$$

and the weighted ES by

$$WES_i = ES_i W_i \quad (\text{A.8})$$

Normally, the ES is calculated based on a single comparison from each study. However, if a study reports infarct size measurements by various methods in the same cohort of animals at the same time, these can be combined into a nested ES before proceeding to pooling estimates from different studies. In Chapter 3 of this thesis, this was used to examine the accuracy of multiparametric MRI for determining infarct size-based treatment effects using meta-regression (described in section A.3). The nested ES for a study is calculated using the following formula:

$$ES_i^* = \frac{\sum_{j=1}^m WES_j}{\sum_{j=1}^m W_j} \quad (\text{A.9})$$

where j is the index of each comparison/ES contributing to the nested ES and m their total number. The SEM of this estimate is calculated as follows:

$$SE_i^* = \sqrt{\frac{m}{\sum_{j=1}^m W_j}} \quad (\text{A.10})$$

The weight for the nested ES is

$$W_i^* = \frac{1}{SE_i^{*2}} \quad (\text{A.11})$$

and the new weighted ES is

$$WES_i^* = ES_i^* W_i^* \quad (\text{A.12})$$

There are two methods that can be used to pool ESs from different studies. The fixed effects model is the simplest method and assumes that ESs vary only slightly between studies due to random error (within-study variance), thus the effect is considered to be fixed. The random effects model assumes that ESs differ significantly between studies, mostly due to differences in study design (between-study variance); this model takes into consideration both the within- and between-study variance.

Firstly, the fixed effects estimate is calculated according to the following formula:

$$ES_{fixed} = \frac{\sum_{i=1}^k WES_i^*}{\sum_{i=1}^k W_i^*} \quad (\text{A.13})$$

where k is the total number of comparisons in the meta-analysis. This formula implies that for studies where non-nested ESs are used, equations A.7 and A.8 are used to estimate individual W_i^* and WES_i^* , respectively. The SEM and the 95% CI of the fixed effects estimate are calculated using the following two formulae respectively:

$$SE_{fixed} = \frac{1}{\sqrt{\sum_{i=1}^k W_i^*}} \quad (\text{A.14})$$

$$95\% CI_{fixed} = ES_{fixed} \pm 1.95996 SE_{fixed} \quad (\text{A.15})$$

To get an indication of how much the ESs vary between studies, the tau-squared value is calculated (τ^2), quantifying the amount of variance beyond that expected by random error alone (DerSimonian & Laird, 1986). This is derived using the following equation:

$$\tau^2 = \frac{Q - df}{C} \quad (\text{A.16})$$

where

$$Q = \sum_{i=1}^k W_i^* (ES_i^* - ES_{fixed})^2 \quad (\text{A.17})$$

$$df = k - 1 \quad (\text{A.18})$$

$$C = \sum_{i=1}^k W_i^* - \frac{\sum_{i=1}^k W_i^{*2}}{\sum_{i=1}^k W_i^*} \quad (\text{A.19})$$

If $\tau^2 = 0$, there is no between-study variance and the fixed and random effects models are equivalent; the larger the value the larger the between-study variance. The random effects model takes τ^2 into account to estimate more conservative weighted

ESs, the weights of which are determined by

$$W_i^+ = \frac{1}{SE_i^{*2} + \tau^2} \quad (\text{A.20})$$

The weighted ESs are therefore

$$WES_i^+ = ES_i^* W_i^+ \quad (\text{A.21})$$

and the random effects estimate is

$$ES_{random} = \frac{\sum_{i=1}^k WES_i^+}{\sum_{i=1}^k W_i^+} \quad (\text{A.22})$$

The SEM and the 95% CI of the random effects estimate are calculated using the following two formulae respectively:

$$SE_{random} = \frac{1}{\sqrt{\sum_{i=1}^k W_i^+}} \quad (\text{A.23})$$

$$95\% CI_{random} = ES_{random} \pm 1.95996 SE_{random} \quad (\text{A.24})$$

Q in equation A.17 follows a χ^2 distribution with df given by equation A.18. Therefore, the χ^2 statistic can be used to test the significance of differences between Q and the expected variation due to within-study variance. In Microsoft Excel (version 2010 or later), this is given by the right-tailed probability of the χ^2 distribution, calculated using the following function:

$$p = \text{CHISQ.DIST.RT}(Q, \text{df}) \quad (\text{A.25})$$

While Q can be used as an estimate of between-study heterogeneity (it is the sum of weighted individual effects without the random error contribution), its value is dependent on the number of comparisons included in the calculation.

An alternative but related approach is to estimate the proportion of total variance that comes from true heterogeneity in ES, I^2 , using the following equation:

$$I^2 = \frac{Q - df}{Q} 100\% \quad (\text{A.26})$$

This lies between 0% and 100%, denoting that all variance is due to chance alone and that all variance is due to between-study heterogeneity, respectively. A random effects model is normally used when $I^2 > 50\%$; this was the case for all combined ESs in the meta-analysis of Chapter 3.

A.3 Meta-Regression

Dedicated meta-analysis techniques can be used to examine the differences between studies or the impact of study characteristics on ES. Stratified meta-analysis is used when studies can be grouped according to a study design criterion in a number of strata with the criterion being, for example, the use of a different method for assessing outcome; strata can then be compared to examine the influence of that criterion on outcome. Unfortunately, the number of studies included in meta-analysis in Chapter 3 was small and their heterogeneity was very high, thus this technique was not employed. Instead, meta-regression was used. This is a type of weighted linear regression that takes into account the within- and between-measurement variance to estimate how much heterogeneity in the dependent variable is explained by one or more independent variables. In Stata software (version 8 or later), the function `metareg` can be used as shown:

$$\text{metareg } y \ x, \text{ wsse}(se_y) \quad (\text{A.27})$$

where y is the list of ESs (dependent variable), x the list/s of corresponding values of the independent variable/s and se_y the list of corresponding SEs estimated using equation A.10, the variance in each study (Harbord & Higgins, 2008). Independent variables can be other ESs calculated based on measurements of a different outcome

or measurements collected using a different method in the same cohort of animals, or absolute variables such as the time point following stroke. In Chapter 3, MRI-based ESs and the time of infarct size measurement following stroke were used as independent variables. For each value in y , metareg gives a weight equal to

$$W_i^R = \frac{1}{SE_i^* + \tau^2} \quad (\text{A.28})$$

where τ^2 is calculated using the restricted maximum likelihood approach to account for multiple covariates (Thompson & Sharp, 1999). The meta-regression model produces a coefficient of determination which represents the proportion of variance in the dependent variable that is accounted by the independent variables, and is estimated by the following equation:

$$\bar{R}^2 = 1 - \frac{\tau_{with\ covariates}^2}{\tau_{without\ covariates}^2} \quad (\text{A.29})$$

In addition, Stata outputs an F -ratio assessing the improvement in model predictability by the addition of covariates (larger values indicate better prediction), the meta-regression coefficient β for each independent variable (slope of the regression line), the 95% CI of β and a p -value from a t -test testing the null hypothesis that $\beta = 0$.

Appendix B

Characteristics of Systematic Review Articles

B.1 Extracted Information

Table B.1 Basic characteristics of articles included in the systematic review.

First author	Year	Species	Type of ischaemia	Intervention tested
Umemura	1994	Rat	Thrombotic	Tirilazad
Quast	1995	Rat	Transient	NOS inhibitor
Tatlisumak	1996	Rat	Permanent	Growth factor
Wei	1998	Rat	Transient	NOS inhibitor
Wiessner	1999	Mouse	Permanent	Transgenics
Mancuso	2000	Rat	Permanent	Hypothermia
Saarelainen	2000	Mouse	Transient	Transgenics
Yenari	2000	Rat	Transient	Hypothermia
Zhang	2000	Rat	Thrombotic	Growth factor
Cash	2001	Rat	Transient	Aminoguanidine
Lee	2001	Rat	Transient	Hypothermia
Schabitz*	2001	Rat	Transient	Growth factor
Shi	2001	Rat	Transient	Estradiol
Sinha	2001	Rat	Transient	Melatonin
Wiessner	2001	Mouse	Permanent	Transgenics
Goto*	2002	Mouse	Transient	Transgenics
Hughes	2002	Mouse	Permanent	Transgenics
Koistinaho	2002	Mouse	Permanent	Transgenics
Mack	2003	Baboon	Transient	Hypothermia
Sironi	2003	Rat	Permanent	Simvastatin
Veldhuis	2003	Rat	Permanent	Dizocilpine
Wagner	2003	Rat	Transient	Hypothermia
Kurozumi	2004	Rat	Transient	Stem cells
Shyu	2004	Rat	Transient	Growth factor
Cimino	2005	Rat	Permanent	Simvastatin
Koistinaho	2005	Mouse	Permanent	Transgenics
Kurozumi	2005	Rat	Transient	Stem cells
Nomura*	2005	Rat	Permanent	Stem cells
van der Weerd	2005	Mouse	Permanent	Transgenics
Yrjänheikki	2005	Rat	Permanent	Atorvastatin
Boltze	2006	Rat	Permanent	Stem cells
Ding*	2006	Rat	Permanent	Atorvastatin
Honma*	2006	Rat	Transient	Stem cells
Horita*	2006	Rat	Permanent	Stem cells
Jiang	2006	Rat	Transient	Stem cells
Liu*	2006	Rat	Permanent	Stem cells
Kameda	2007	Rat	Transient	Stem cells
Pialat	2007	Mouse	Permanent	Transgenics
Shimamura	2007	Rat	Permanent	Fluvastatin
Ukai*	2007	Rat	Permanent	Stem cells
Uno	2007	Mouse	Transient	Edaravone
Wei	2007	Rat	Transient	Stem cells

First author	Year	Species	Type of ischaemia	Intervention tested
Esneault	2008	Rat	Transient	Stem cells
Kim	2008	Rat	Transient	Stem cells
Koh	2008	Rat	Transient	Stem cells
Li*	2008	Rat	Transient	Stem cells
Omori*	2008	Rat	Permanent	Stem cells
Onda*	2008	Rat	Permanent	Stem cells
Schmerbach	2008	Rat	Transient	Pioglitazone
Yoo	2008	Rat	Transient	Stem cells
Li	2009	Rat	Thrombotic	Erythropoietin
Modo	2009	Rat	Transient	Stem cells
Toyama*	2009	Rat	Permanent	Stem cells
van der Weerd	2010	Mouse	Permanent	Transgenics

NOS indicates nitric oxide synthase.

*Studies that measured infarct size with both MRI and histology and were included in meta-analysis.

Table B.2 Reporting of study quality items.

First author	Year	1	2	3	4	5	6	7	8	9	10	Score
Umemura	1994	+	+	+		+	+					5
Quast	1995	+	+				+			+		4
Tatlisumak	1996	+	+	+	+		+			+		6
Wei	1998	+	+				+	+				4
Wiessner	1999	+	+				+					3
Mancuso	2000	+	+	+			+					4
Saarelainen	2000	+	+				+			+		4
Yenari	2000	+	+			+	+					4
Zhang	2000	+	+			+	+			+		5
Cash	2001	+	+	+		+	+			+		6
Lee	2001		+				+					2
Schabitz*	2001	+	+	+		+	+					5
Shi	2001	+	+							+		3
Sinha	2001	+	+				+			+		4
Wiessner	2001	+	+				+			+		4
Goto*	2002	+	+	+		+	+			+		6
Hughes	2002	+	+				+			+		4
Koistinaho	2002	+	+				+			+		4
Mack	2003	+	+							+		3
Sironi	2003	+	+	+			+			+		5
Veldhuis	2003	+	+				+			+		4
Wagner	2003	+	+				+			+		4
Kurozumi	2004	+	+	+						+		4
Shyu	2004	+	+			+	+					4
Cimino	2005	+										1
Koistinaho	2005	+	+				+			+		4

First author	Year	1	2	3	4	5	6	7	8	9	10	Score
Kurozumi	2005	+		+								2
Nomura*	2005	+	+							+		3
van der Weerd	2005	+	+			+	+			+		5
Yrjänheikki	2005	+	+			+	+			+		5
Boltze	2006	+	+	+		+		+		+		6
Ding*	2006	+		+			+			+		4
Honma*	2006	+	+									2
Horita*	2006	+								+		2
Jiang	2006	+	+	+		+	+			+		6
Liu*	2006	+	+									2
Kameda	2007	+					+			+		3
Pialat	2007	+	+			+				+		4
Shimamura	2007	+									+	2
Ukai*	2007	+								+		2
Uno	2007											0
Wei	2007	+	+	+			+					4
Esneault	2008	+	+				+			+		4
Kim	2008	+	+	+	+					+	+	6
Koh	2008	+	+			+	+			+		5
Li*	2008	+	+	+		+						4
Omori*	2008	+		+						+		3
Onda*	2008	+								+		2
Schmerbach	2008	+		+		+	+			+		5
Yoo	2008	+	+	+			+			+		5
Li	2009	+	+			+	+			+		5
Modo	2009	+								+	+	3
Toyama*	2009	+								+		2
van der Weerd	2010	+	+			+	+			+		5

Studies were assessed according to a checklist by Macleod et al. (2004). The quality items are: (1) peer-reviewed publication, (2) control of temperature, (3) random allocation to treatment or control, (4) blinded induction of ischaemia, (5) blinded assessment of outcome, (6) use of anaesthetic without significant intrinsic neuroprotective activity, (7) use of comorbid animals (aged, diabetic, or hypertensive), (8) sample size calculation, (9) compliance with animal welfare regulations and (10) statement of potential conflict of interests.

*Studies that measured infarct size with both MRI and histology and were included in meta-analysis.

Table B.3 Reporting of basic MRI parameters per type.

Parameter	Reported, n/N (%)	Details of reported parameters, n/N (%)
Ischaemic brain side	42/54 (78)	Left: 12/42 (29) Right: 30/42 (71)
Animal positioning method in scanner	22/54 (41)	Carriage/cradle: 15/22 (68) Stereotaxic: 7/22 (32)
Anaesthetic used during scanning	36/54 (67)	Fentanyl-Fluanisone-Midazolame: 1/36 (3) Halothane: 10/36 (28) Isoflurane: 10/36 (28) Ketamine: 12/36 (33) Pentobarbital: 3/36 (8)
Monitoring of physiological parameters during scanning	16/54 (30)	Temperature only: 7/16 (44) Temperature & other: 6/16 (38) Other: 3/16 (19)
Scanner equipment manufacturer	51/54 (94)	Bruker: 14/51 (27) General Electric: 9/51 (18) Medinus: 2/51 (4) Otsuka Electronics: 1/51 (2) Philips: 1/51 (2) Spectroscopy Imaging Systems: 1/51 (2) Varian: 1/51 (2) Multiple: 22/51 (43)
Scanner field strength (T)	51/54 (94)	1.5: 5/51 (10) 2: 4/51 (8) 2.3: 4/51 (8) 3: 7/51 (14) 4.7: 15/51 (29) 7: 15/51 (29) 9.4: 1/51 (2)
Type of radiofrequency coil	31/54 (57)	Volume: 17/31 (55) Surface: 7/31 (23) Volume and surface: 7/31 (23)
Software for image analysis	40/54 (74)	Academic: 17/40 (43) Commercial: 16/40 (40) Custom-made: 7/40 (17)
Method of lesion size determination*	51/78 (65)	Manual tracing: 10/51 (20) Thresholding: 39/51 (76) Automated algorithm: 2/51 (4)

Parameter	Reported, n/N (%)	Details of reported parameters, n/N (%)
How the lesion size is presented*	77/78 (99)	Lesion volume: 45/77 (58) Lesion area at one slice: 1/77 (1) Change from baseline: 7/77 (9) Fraction/percent of ipsilateral side: 14/77 (18) Fraction/percent of contralateral side: 6/77 (8) Percent of brain: 4/77 (5)
Oedema correction method for infarct size†	11/44 (25)	Swanson's method: 9/11 (82) Leach's method: 2/11 (18)

n indicates the number of protocols for which the parameter is reported; N, total number of protocols for which the parameter is relevant.

*Post-processing parameters, applicable to each of the 78 imaging protocols.

†Applied on T₂-weighted images acquired at the subacute and early chronic stages of stroke (relevant to 44 out of 45 protocols).

B.2 Reviewed Articles References

Boltze, J., Kowalski, I., Forschler, A., Schmidt, U., Wagner, D., Lobsien, D., . . . Emmrich, F. (2006). The stairway: a novel behavioral test detecting sensorimotor stroke deficits in rats. *Artificial Organs*, 30(10), 756-763. doi:10.1111/j.1525-1594.2006.00297.x

Cash, D., Beech, J. S., Rayne, R. C., Bath, P. M., Meldrum, B. S., & Williams, S. C. (2001). Neuroprotective effect of aminoguanidine on transient focal ischaemia in the rat brain. *Brain Research*, 905(1-2), 91-103.

Cimino, M., Balduini, W., Carloni, S., Gelosa, P., Guerrini, U., Tremoli, E., & Sironi, L. (2005). Neuroprotective effect of simvastatin in stroke: a comparison between adult and neonatal rat models of cerebral ischemia. *Neurotoxicology*, 26(5), 929-933. doi:10.1016/j.neuro.2005.03.009

Ding, G., Jiang, Q., Li, L., Zhang, L., Zhang, Z. G., Panda, S., . . . Chopp, M. (2006). MRI of combination treatment of embolic stroke in rat with rtPA and atorvastatin. *Journal of the Neurological Sciences*, 246(1-2), 139-147. doi:10.1016/j.jns.2006.02.020

Esneault, E., Pacary, E., Eddi, D., Freret, T., Tixier, E., Toutain, J., . . . Bernaudin, M. (2008). Combined therapeutic strategy using erythropoietin and mesenchymal stem cells potentiates neurogenesis after transient focal cerebral ischemia in rats. *Journal of Cerebral Blood Flow and Metabolism*, 28(9), 1552-1563. doi:10.1038/jcbfm.2008.40

- Goto, S., Xue, R., Sugo, N., Sawada, M., Blizzard, K. K., Poitras, M. F., . . . Hurn, P. D. (2002). Poly(ADP-ribose) polymerase impairs early and long-term experimental stroke recovery. *Stroke*, 33(4), 1101-1106.
- Honma, T., Honmou, O., Iihoshi, S., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2006). Intravenous infusion of immortalized human mesenchymal stem cells protects against injury in a cerebral ischemia model in adult rat. *Experimental Neurology*, 199(1), 56-66. doi:10.1016/j.expneurol.2005.05.004
- Horita, Y., Honmou, O., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2006). Intravenous administration of glial cell line-derived neurotrophic factor gene-modified human mesenchymal stem cells protects against injury in a cerebral ischemia model in the adult rat. *Journal of Neuroscience Research*, 84(7), 1495-1504. doi:10.1002/jnr.21056
- Hughes, P. M., Allegrini, P. R., Rudin, M., Perry, V. H., Mir, A. K., & Wiessner, C. (2002). Monocyte chemoattractant protein-1 deficiency is protective in a murine stroke model. *Journal of Cerebral Blood Flow and Metabolism*, 22(3), 308-317. doi:10.1097/00004647-200203000-00008
- Jiang, Q., Zhang, Z. G., Ding, G. L., Silver, B., Zhang, L., Meng, H., . . . Chopp, M. (2006). MRI detects white matter reorganization after neural progenitor cell treatment of stroke. *Neuroimage*, 32(3), 1080-1089. doi:10.1016/j.neuroimage.2006.05.025
- Kameda, M., Shingo, T., Takahashi, K., Muraoka, K., Kurozumi, K., Yasuhara, T., . . . Date, I. (2007). Adult neural stem and progenitor cells modified to secrete GDNF can protect, migrate and integrate after intracerebral transplantation in rats with transient forebrain ischemia. *European Journal of Neuroscience*, 26(6), 1462-1478. doi:10.1111/j.1460-9568.2007.05776.x
- Kim, S. S., Yoo, S. W., Park, T. S., Ahn, S. C., Jeong, H. S., Kim, J. W., . . . Suh-Kim, H. (2008). Neural induction with neurogenin1 increases the therapeutic effects of mesenchymal stem cells in the ischemic brain. *Stem Cells*, 26(9), 2217-2228. doi:10.1634/stemcells.2008-0108
- Koh, S. H., Kim, K. S., Choi, M. R., Jung, K. H., Park, K. S., Chai, Y. G., . . . Kim, S. H. (2008). Implantation of human umbilical cord-derived mesenchymal stem cells as a neuroprotective therapy for ischemic stroke in rats. *Brain Research*, 1229, 233-248. doi:10.1016/j.brainres.2008.06.087
- Koistinaho, M., Kettunen, M. I., Goldsteins, G., Keinänen, R., Salminen, A., Ort, M., . . . Koistinaho, J. (2002). Beta-amyloid precursor protein transgenic mice that harbor diffuse A beta deposits but do not form plaques show increased ischemic vulnerability: role of inflammation. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3), 1610-1615. doi:10.1073/pnas.032670899

- Koistinaho, M., Malm, T. M., Kettunen, M. I., Goldsteins, G., Starckx, S., Kauppinen, R. A., . . . Koistinaho, J. (2005). Minocycline protects against permanent cerebral ischemia in wild type but not in matrix metalloprotease-9-deficient mice. *Journal of Cerebral Blood Flow and Metabolism*, 25(4), 460-467. doi:10.1038/sj.jcbfm.9600040
- Kurozumi, K., Nakamura, K., Tamiya, T., Kawano, Y., Ishii, K., Kobune, M., . . . Hamada, H. (2005). Mesenchymal stem cells that produce neurotrophic factors reduce ischemic damage in the rat middle cerebral artery occlusion model. *Molecular Therapy*, 11(1), 96-104. doi:10.1016/j.ymthe.2004.09.020
- Kurozumi, K., Nakamura, K., Tamiya, T., Kawano, Y., Kobune, M., Hirai, S., . . . Hamada, H. (2004). BDNF gene-modified mesenchymal stem cells promote functional recovery and reduce infarct size in the rat middle cerebral artery occlusion model. *Molecular Therapy*, 9(2), 189-197. doi:10.1016/j.ymthe.2003.10.012
- Lee, J. E., Yoon, E. J., Sun, G. H., Lee, S. H., Moseley, M. E., Steinberg, G. K., & Yenari, M. A. (2001). Mild hypothermia inhibits matrix metalloprotease generation and blood-brain barrier disruption in experimental stroke [Abstract]. *Journal of Cerebral Blood Flow and Metabolism*, 21 Suppl 1, S130.
- Li, L., Jiang, Q., Ding, G., Zhang, L., Zhang, Z. G., Li, Q., . . . Chopp, M. (2009). MRI identification of white matter reorganization enhanced by erythropoietin treatment in a rat model of focal ischemia. *Stroke*, 40(3), 936-941. doi:10.1161/STROKEAHA.108.527713
- Li, W. Y., Choi, Y. J., Lee, P. H., Huh, K., Kang, Y. M., Kim, H. S., . . . Bang, O. Y. (2008). Mesenchymal stem cells for ischemic stroke: changes in effects after ex vivo culturing. *Cell Transplantation*, 17(9), 1045-1059.
- Liu, H., Honmou, O., Harada, K., Nakamura, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2006). Neuroprotection by PlGF gene-modified human mesenchymal stem cells after cerebral ischaemia. *Brain*, 129(Pt 10), 2734-2745. doi:10.1093/brain/awl207
- Mack, W. J., Huang, J., Winfree, C., Kim, G., Oppermann, M., Dobak, J., . . . Connolly, E. S., Jr. (2003). Ultrarapid, convection-enhanced intravascular hypothermia: a feasibility study in nonhuman primate stroke. *Stroke*, 34(8), 1994-1999. doi:10.1161/01.STR.0000079813.31539.6D
- Mancuso, A., Derugin, N., Hara, K., Sharp, F. R., & Weinstein, P. R. (2000). Mild hypothermia decreases the incidence of transient ADC reduction detected with diffusion MRI and expression of c-fos and hsp70 mRNA during acute focal ischemia in rats. *Brain Research*, 887(1), 34-45.

- Modo, M., Beech, J. S., Meade, T. J., Williams, S. C., & Price, J. (2009). A chronic 1 year assessment of MRI contrast agent-labelled neural stem cell transplants in stroke. *Neuroimage*, 47 Suppl 2, T133-142. doi:10.1016/j.neuroimage.2008.06.017
- Nomura, T., Honmou, O., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2005). I.V. infusion of brain-derived neurotrophic factor gene-modified human mesenchymal stem cells protects against injury in a cerebral ischemia model in adult rat. *Neuroscience*, 136(1), 161-169. doi:10.1016/j.neuroscience.2005.06.062
- Omori, Y., Honmou, O., Harada, K., Suzuki, J., Houkin, K., & Kocsis, J. D. (2008). Optimization of a therapeutic protocol for intravenous injection of human mesenchymal stem cells after cerebral ischemia in adult rats. *Brain Research*, 1236, 30-38. doi:10.1016/j.brainres.2008.07.116
- Onda, T., Honmou, O., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2008). Therapeutic benefits by human mesenchymal stem cells (hMSCs) and Ang-1 gene-modified hMSCs after cerebral ischemia. *Journal of Cerebral Blood Flow and Metabolism*, 28(2), 329-340. doi:10.1038/sj.jcbfm.9600527
- Pialat, J. B., Cho, T. H., Beuf, O., Joye, E., Moucharrafié, S., Langlois, J. B., . . . Wiart, M. (2007). MRI monitoring of focal cerebral ischemia in peroxisome proliferator-activated receptor (PPAR)-deficient mice. *NMR in Biomedicine*, 20(3), 335-342. doi:10.1002/nbm.1157
- Quast, M. J., Wei, J., & Huang, N. C. (1995). Nitric oxide synthase inhibitor NG-nitro-L-arginine methyl ester decreases ischemic damage in reversible focal cerebral ischemia in hyperglycemic rats. *Brain Research*, 677(2), 204-212.
- Saarelainen, T., Lukkarinen, J. A., Koponen, S., Grohn, O. H., Jolkkonen, J., Koponen, E., . . . Castren, E. (2000). Transgenic mice overexpressing truncated trkB neurotrophin receptors in neurons show increased susceptibility to cortical injury after focal cerebral ischemia. *Molecular and Cellular Neurosciences*, 16(2), 87-96. doi:10.1006/mcne.2000.0863
- Schabitz, W. R., Hoffmann, T. T., Heiland, S., Kollmar, R., Bardutzky, J., Sommer, C., & Schwab, S. (2001). Delayed neuroprotective effect of insulin-like growth factor-i after experimental transient focal cerebral ischemia monitored with MRI. *Stroke*, 32(5), 1226-1233.
- Schmerbach, K., Schefe, J. H., Krikov, M., Muller, S., Villringer, A., Kintscher, U., . . . Thoene-Reineke, C. (2008). Comparison between single and combined treatment with candesartan and pioglitazone following transient focal ischemia in rat brain. *Brain Research*, 1208, 225-233. doi:10.1016/j.brainres.2008.02.032

- Shi, J., Bui, J. D., Yang, S. H., He, Z., Lucas, T. H., Buckley, D. L., . . . Simpkins, J. W. (2001). Estrogens decrease reperfusion-associated cortical ischemic damage: an MRI analysis in a transient focal ischemia model. *Stroke*, *32*(4), 987-992.
- Shimamura, M., Sato, N., Sata, M., Kurinami, H., Takeuchi, D., Wakayama, K., . . . Morishita, R. (2007). Delayed postischemic treatment with fluvastatin improved cognitive impairment after stroke in rats. *Stroke*, *38*(12), 3251-3258. doi:10.1161/STROKEAHA.107.485045
- Shyu, W. C., Lin, S. Z., Yang, H. I., Tzeng, Y. S., Pang, C. Y., Yen, P. S., & Li, H. (2004). Functional recovery of stroke rats induced by granulocyte colony-stimulating factor-stimulated stem cells. *Circulation*, *110*(13), 1847-1854. doi:10.1161/01.CIR.0000142616.07367.66
- Sinha, K., Degaonkar, M. N., Jagannathan, N. R., & Gupta, Y. K. (2001). Effect of melatonin on ischemia reperfusion injury induced by middle cerebral artery occlusion in rats. *European Journal of Pharmacology*, *428*(2), 185-192.
- Sironi, L., Cimino, M., Guerrini, U., Calvio, A. M., Lodetti, B., Asdente, M., . . . Tremoli, E. (2003). Treatment with statins after induction of focal ischemia in rats reduces the extent of brain damage. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *23*(2), 322-327.
- Tatlisumak, T., Takano, K., Carano, R. A., & Fisher, M. (1996). Effect of basic fibroblast growth factor on experimental focal ischemia studied by diffusion-weighted and perfusion imaging. *Stroke*, *27*(12), 2292-2297; discussion 2298.
- Toyama, K., Honmou, O., Harada, K., Suzuki, J., Houkin, K., Hamada, H., & Kocsis, J. D. (2009). Therapeutic benefits of angiogenetic gene-modified human mesenchymal stem cells after cerebral ischemia. *Experimental Neurology*, *216*(1), 47-55. doi:10.1016/j.expneurol.2008.11.010
- Ukai, R., Honmou, O., Harada, K., Houkin, K., Hamada, H., & Kocsis, J. D. (2007). Mesenchymal stem cells derived from peripheral blood protects against ischemia. *Journal of Neurotrauma*, *24*(3), 508-520. doi:10.1089/neu.2006.0161
- Umemura, K., Wada, K., Uematsu, T., Mizuno, A., & Nakashima, M. (1994). Effect of 21-aminosteroid lipid peroxidation inhibitor, U74006F, in the rat middle cerebral artery occlusion model. *European Journal of Pharmacology*, *251*(1), 69-74.
- Uno, H., Taguchi, A., Nakagomi, T., Saino, O., Fujikawa, M., & Matsuyama, T. (2007). A long-term effect of edaravone given at ischemia - reperfusion on ischemic stroke in mice [Abstract]. *Journal of Cerebral Blood Flow and Metabolism*, *BP*(10), 05H.
- van der Weerd, L., Lythgoe, M. F., Badin, R. A., Valentim, L. M., Akbar, M. T., de Belleruche, J. S., . . . Gadian, D. G. (2005). Neuroprotective effects of HSP70

- overexpression after cerebral ischaemia--an MRI study. *Experimental Neurology*, 195(1), 257-266. doi:10.1016/j.expneurol.2005.05.002
- van der Weerd, L., Tariq Akbar, M., Aron Badin, R., Valentim, L. M., Thomas, D. L., Wells, D. J., . . . de Belleruche, J. S. (2010). Overexpression of heat shock protein 27 reduces cortical damage after cerebral ischemia. *Journal of Cerebral Blood Flow and Metabolism*, 30(4), 849-856. doi:10.1038/jcbfm.2009.249
- Veldhuis, W. B., van der Stelt, M., Delmas, F., Gillet, B., Veldink, G. A., Vliegenthart, J. F., . . . Bar, P. R. (2003). In vivo excitotoxicity induced by ouabain, a Na⁺/K⁺-ATPase inhibitor. *Journal of Cerebral Blood Flow and Metabolism*, 23(1), 62-74.
- Wagner, S., Nagel, S., Kluge, B., Schwab, S., Heiland, S., Koziol, J., . . . Hacke, W. (2003). Topographically graded postischemic presence of metalloproteinases is inhibited by hypothermia. *Brain Research*, 984(1-2), 63-75.
- Wei, J., & Quast, M. J. (1998). Effect of nitric oxide synthase inhibitor on a hyperglycemic rat model of reversible focal ischemia: detection of excitatory amino acids release and hydroxyl radical formation. *Brain Research*, 791(1-2), 146-156.
- Wei, J. J., Wang, R. Z., Lu, J. J., Wang, Y., Fan, X. T., Feng, F., . . . Kong, Y. G. (2007). [In vivo tracking of bone marrow mesenchymal stem cells labeled with superparamagnetic iron oxide after cerebral ischemia in rats by magnetic resonance imaging]. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao. Acta Academiae Medicinae Sinicae*, 29(1), 73-77.
- Wiessner, C., Allegrini, P. R., Ekatodramis, D., Jewell, U. R., Stallmach, T., & Gassmann, M. (2001). Increased cerebral infarct volumes in polyglobulic mice overexpressing erythropoietin. *Journal of Cerebral Blood Flow and Metabolism*, 21(7), 857-864. doi:10.1097/00004647-200107000-00011
- Wiessner, C., Allegrini, P. R., Rupalla, K., Sauer, D., Oltersdorf, T., McGregor, A. L., . . . van der Putten, H. (1999). Neuron-specific transgene expression of Bcl-XL but not Bcl-2 genes reduced lesion size after permanent middle cerebral artery occlusion in mice. *Neuroscience Letters*, 268(3), 119-122.
- Yenari, M. A., Onley, D., Hedehus, M., deCrespigny, A., Sun, G. H., Moseley, M. E., & Steinberg, G. K. (2000). Diffusion- and perfusion-weighted magnetic resonance imaging of focal cerebral ischemia and cortical spreading depression under conditions of mild hypothermia. *Brain Research*, 885(2), 208-219.
- Yoo, S. W., Kim, S. S., Lee, S. Y., Lee, H. S., Kim, H. S., Lee, Y. D., & Suh-Kim, H. (2008). Mesenchymal stem cells promote proliferation of endogenous neural stem cells and survival of newborn cells in a rat stroke model. *Experimental and Molecular Medicine*, 40(4), 387-397. doi:10.3858/emm.2008.40.4.387

- Yrjanheikki, J., Koistinaho, J., Kettunen, M., Kauppinen, R. A., Appel, K., Hull, M., & Fiebich, B. L. (2005). Long-term protective effect of atorvastatin in permanent focal cerebral ischemia. *Brain Research*, 1052(2), 174-179. doi:10.1016/j.brainres.2005.06.004
- Zhang, Z. G., Zhang, L., Jiang, Q., Zhang, R., Davies, K., Powers, C., . . . Chopp, M. (2000). VEGF enhances angiogenesis and promotes blood-brain barrier leakage in the ischemic brain. *Journal of Clinical Investigation*, 106(7), 829-838. doi:10.1172/JCI9369

Appendix C

Statistical Tests

Table C.1 Shapiro–Wilk test for normality.

Variable	Dataset	W	<i>p</i> -value	
Chapter 4				
Analysis time, in seconds	Manual	0.97	0.902	
	Thresholding	0.88	0.140	
	<i>k</i> -means (2)	0.97	0.863	
	<i>k</i> -means (3)	0.91	0.299	
	Canny	0.89	0.168	
Chapter 5				
Analysis time, in minutes	Manual	0.97	0.686	
	Point counting	0.97	0.659	
	Semi-automated	0.95	0.281	
Uncorrected infarct volume, in mm ³	Manual	0.95	0.329	
	Point counting	0.96	0.576	
	Semi-automated	0.97	0.783	
Corrected infarct volume, in mm ³	Manual	0.98	0.882	
	Point counting	0.98	0.878	
	Semi-automated	0.99	0.988	
Chapter 6				
Linear scaling errors, in %	Frequency encoding	0.93	< 0.001	
	Phase encoding	0.93	< 0.001	
Uncorrected, in mm ³	1	A	0.96	0.569
		B	0.97	0.634
		C	0.96	0.507
		D	0.96	0.597
		E	0.97	0.703
		F	0.96	0.532
	2	A	0.96	0.565
		B	0.96	0.502
		C	0.96	0.475
		D	0.97	0.725
		E	0.96	0.571
		F	0.97	0.708
	3	A	0.96	0.530
		B	0.96	0.543
		C	0.96	0.578
		D	0.96	0.488
		E	0.96	0.605
		F	0.96	0.562
Corrected based on all slices, in mm ³	1	A	0.98	0.953
		B	0.98	0.901
		C	0.98	0.894
		D	0.98	0.872
		E	0.98	0.910
		F	0.97	0.713
	2	A	0.98	0.939

Variable		Dataset	W	p-value	
Corrected based on all slices, as % of brain	3	B	0.98	0.897	
		C	0.97	0.774	
		D	0.98	0.874	
		E	0.98	0.932	
		F	0.97	0.816	
		A	0.97	0.655	
	1	B	0.98	0.953	
		C	0.97	0.807	
		D	0.97	0.796	
		E	0.98	0.855	
		F	0.97	0.848	
		A	0.97	0.808	
	Corrected based on all slices, as % of ipsilateral hemisphere	2	B	0.97	0.717
			C	0.97	0.810
			D	0.97	0.775
			E	0.97	0.756
			F	0.97	0.676
			A	0.97	0.676
3		A	0.98	0.850	
		B	0.97	0.808	
		C	0.97	0.654	
		D	0.97	0.666	
		E	0.97	0.789	
		F	0.97	0.639	
Corrected based on all slices, as % of ipsilateral hemisphere	3	A	0.96	0.631	
		B	0.97	0.821	
		C	0.97	0.813	
		D	0.97	0.742	
		E	0.97	0.661	
		F	0.97	0.773	
	1	A	0.97	0.764	
		B	0.96	0.603	
		C	0.97	0.691	
		D	0.97	0.683	
		E	0.97	0.709	
		F	0.96	0.585	
	2	A	0.97	0.777	
		B	0.97	0.734	
		C	0.96	0.575	
		D	0.96	0.532	
		E	0.97	0.707	
		F	0.96	0.507	
3	A	0.96	0.527		
	B	0.97	0.792		
	C	0.97	0.684		
	D	0.96	0.611		
	E	0.96	0.586		
	F	0.97	0.671		

Variable	Dataset	W	<i>p</i> -value	
Corrected based on all slices, as % of contralateral hemisphere	1	A	0.98	0.897
		B	0.97	0.838
		C	0.98	0.886
		D	0.98	0.877
		E	0.97	0.825
		F	0.97	0.761
	2	A	0.98	0.901
		B	0.98	0.861
		C	0.97	0.714
		D	0.97	0.795
		E	0.98	0.871
		F	0.97	0.802
	3	A	0.97	0.761
		B	0.98	0.933
		C	0.98	0.892
		D	0.98	0.850
		E	0.97	0.763
		F	0.98	0.885
Corrected based on slices including the infarct, in mm ³	1	A	0.99	0.983
		B	0.98	0.956
		C	0.98	0.866
		D	0.98	0.885
		E	0.99	0.986
		F	0.97	0.725
	2	A	0.98	0.918
		B	0.97	0.848
		C	0.98	0.897
		D	0.98	0.957
		E	0.98	0.921
		F	0.98	0.878
	3	A	0.97	0.814
		B	0.98	0.961
		C	0.98	0.867
		D	0.97	0.847
		E	0.98	0.942
		F	0.98	0.881
Corrected based on slices including the infarct, as % of brain	1	A	0.97	0.682
		B	0.95	0.345
		C	0.96	0.461
		D	0.94	0.220
		E	0.97	0.650
		F	0.94	0.231
	2	A	0.95	0.374
		B	0.96	0.556
		C	0.98	0.960
		D	0.94	0.259
		E	0.96	0.438

Variable		Dataset	W	<i>p</i> -value
		F	0.94	0.194
	3	A	0.95	0.361
		B	0.96	0.452
		C	0.96	0.465
		D	0.95	0.348
		E	0.95	0.414
		F	0.96	0.548
Corrected based on slices including the infarct, as % of ipsilateral hemisphere	1	A	0.97	0.682
		B	0.95	0.316
		C	0.96	0.463
		D	0.94	0.210
		E	0.97	0.649
		F	0.93	0.151
	2	A	0.95	0.356
		B	0.95	0.412
		C	0.98	0.974
		D	0.94	0.202
		E	0.95	0.352
		F	0.93	0.149
	3	A	0.95	0.311
		B	0.96	0.454
		C	0.95	0.375
		D	0.95	0.285
		E	0.95	0.313
		F	0.96	0.542
Corrected based on slices including the infarct, as % of contralateral hemisphere	1	A	0.96	0.607
		B	0.95	0.419
		C	0.95	0.409
		D	0.94	0.275
		E	0.97	0.734
		F	0.95	0.283
	2	A	0.95	0.364
		B	0.96	0.602
		C	0.98	0.891
		D	0.95	0.404
		E	0.96	0.495
		F	0.95	0.282
	3	A	0.95	0.423
		B	0.95	0.407
		C	0.96	0.516
		D	0.95	0.425
		E	0.96	0.497
		F	0.96	0.551

p-values less than 0.05 are shown in bold. A set of measurements for each scanner 'A'–'F' and each case 1–3 is used in Chapter 6.

Table C.2 Mauchly's sphericity test.

Variable	Case	df	χ^2	<i>p</i> -value	ϵ
Chapter 4					
Analysis time, in seconds	N/A	9	96.07	< 0.001	0.43
Chapter 5					
Analysis time, in minutes	N/A	2	2.90	0.235	-
Uncorrected infarct volume, in mm ³	N/A	2	7.35	0.025	0.76
Corrected infarct volume, in mm ³	N/A	2	2.86	0.239	-
Chapter 6					
Uncorrected, in mm ³	1	14	51.78	< 0.001	0.39
	2	14	84.03	< 0.001	0.35
	3	14	24.53	0.041	0.67
Corrected based on all slices in mm ³	1	14	29.43	0.010	0.55
	2	14	63.69	< 0.001	0.43
	3	14	36.95	< 0.001	0.58
as % of brain	1	14	12.44	0.575	-
	2	14	44.52	< 0.001	0.49
	3	14	36.60	< 0.001	0.57
as % of ipsilateral hemisphere	1	14	13.13	0.520	-
	2	14	43.77	< 0.001	0.49
	3	14	40.80	< 0.001	0.53
as % of contralateral hemisphere	1	14	11.63	0.640	-
	2	14	45.71	< 0.001	0.48
	3	14	33.01	0.003	0.60
Corrected based on slices including the infarct in mm ³	1	14	35.22	< 0.001	0.56
	2	14	63.81	< 0.001	0.49
	3	14	48.29	< 0.001	0.51
as % of brain	1	14	22.10	0.079	-
	2	14	80.19	< 0.001	0.45
	3	14	49.64	< 0.001	0.52
as % of ipsilateral hemisphere	1	14	21.28	0.097	-
	2	14	79.38	< 0.001	0.45
	3	14	51.64	< 0.001	0.50
as % of contralateral hemisphere	1	14	22.88	0.064	-
	2	14	81.08	< 0.001	0.45
	3	14	47.31	< 0.001	0.55

p-values less than 0.05 are shown in bold. ϵ is the Greenhouse–Geisser factor used to correct the degrees of freedom (df) for analysis of variance; where *p* > 0.05 a correction was not used. N/A indicates not applicable.

Table C.3 *p*-values for Bonferroni pairwise comparisons.

Infarct volume measure	Case		A	B	C	D	E
Uncorrected, in mm ³	1	B	< 0.001				
		C	< 0.001	< 0.001			
		D	< 0.001	1.000	< 0.001		
		E	< 0.001	0.089	< 0.001	1.000	
		F	< 0.001	< 0.001	1.000	< 0.001	< 0.001
	2	B	< 0.001				
		C	< 0.001	0.639			
		D	< 0.001	< 0.001	< 0.001		
		E	< 0.001	1.000	0.595	< 0.001	
		F	< 0.001	0.003	0.024	0.082	< 0.001
	3	B	0.168				
		C	1.000	0.456			
		D	1.000	0.198	1.000		
		E	0.249	1.000	0.892	0.135	
		F	1.000	0.239	1.000	1.000	0.453
Corrected based on all slices, in mm ³	1	B	< 0.001				
		C	< 0.001	0.003			
		D	< 0.001	1.000	0.146		
		E	< 0.001	1.000	< 0.001	0.240	
		F	< 0.001	< 0.001	0.886	0.004	< 0.001
	2	B	< 0.001				
		C	< 0.001	1.000			
		D	< 0.001	0.182	0.269		
		E	< 0.001	1.000	1.000	0.876	
		F	< 0.001	1.000	1.000	0.082	1.000
	3	B	0.002				
		C	1.000	0.007			
		D	1.000	0.002	1.000		
		E	0.016	1.000	0.072	0.003	
		F	1.000	< 0.001	1.000	1.000	0.064
Corrected based on all slices, as % of brain	1	B	1.000				
		C	1.000	0.959			
		D	< 0.001	< 0.001	< 0.001		
		E	1.000	1.000	0.441	0.007	
		F	< 0.001	0.005	< 0.001	1.000	0.070
	2	B	< 0.001				
		C	0.003	0.068			
		D	1.000	< 0.001	0.004		
		E	< 0.001	0.003	< 0.001	< 0.001	
		F	0.300	< 0.001	< 0.001	0.036	< 0.001
	3	B	< 0.001				
		C	1.000	< 0.001			
		D	1.000	< 0.001	1.000		
		E	0.005	0.429	0.025	< 0.001	
		F	1.000	< 0.001	1.000	1.000	0.020

Infarct volume measure	Case		A	B	C	D	E
Corrected based on all slices, as % of ipsilateral hemisphere	1	B	1.000				
		C	1.000	1.000			
		D	< 0.001	< 0.001	< 0.001		
		E	1.000	1.000	0.519	0.008	
		F	< 0.001	0.006	< 0.001	1.000	0.090
	2	B	< 0.001				
		C	0.003	0.062			
		D	1.000	< 0.001	0.005		
		E	< 0.001	0.003	< 0.001	< 0.001	
		F	0.358	< 0.001	< 0.001	0.040	< 0.001
	3	B	< 0.001				
		C	1.000	0.002			
		D	1.000	< 0.001	1.000		
		E	0.004	0.327	0.023	< 0.001	
		F	1.000	< 0.001	1.000	1.000	0.019
Corrected based on all slices, as % of contralateral hemisphere	1	B	1.000				
		C	1.000	0.620			
		D	< 0.001	< 0.001	< 0.001		
		E	1.000	1.000	0.362	0.005	
		F	< 0.001	0.005	< 0.001	1.000	0.052
	2	B	< 0.001				
		C	0.003	0.077			
		D	1.000	< 0.001	0.003		
		E	< 0.001	0.003	< 0.001	< 0.001	
		F	0.238	< 0.001	< 0.001	0.032	< 0.001
	3	B	< 0.001				
		C	1.000	< 0.001			
		D	1.000	< 0.001	1.000		
		E	0.007	0.676	0.031	< 0.001	
		F	1.000	< 0.001	1.000	1.000	0.022
Corrected based on slices including the infarct, in mm ³	1	B	< 0.001				
		C	< 0.001	< 0.001			
		D	< 0.001	1.000	0.009		
		E	< 0.001	1.000	< 0.001	0.315	
		F	< 0.001	< 0.001	1.000	< 0.001	< 0.001
	2	B	< 0.001				
		C	< 0.001	1.000			
		D	< 0.001	0.010	0.038		
		E	< 0.001	1.000	1.000	0.049	
		F	< 0.001	0.607	1.000	0.031	1.000
	3	B	0.006				
		C	1.000	0.031			
		D	1.000	0.009	1.000		
		E	0.012	1.000	0.082	0.009	
		F	1.000	0.004	1.000	1.000	0.148

Infarct volume measure	Case		A	B	C	D	E
Corrected based on slices including the infarct, as % of brain	1	B	1.000				
		C	1.000	1.000			
		D	1.000	1.000	1.000		
		E	1.000	1.000	1.000	1.000	
		F	0.189	0.003	0.021	1.000	0.048
	2	B	0.325				
		C	0.063	1.000			
		D	1.000	1.000	1.000		
		E	< 0.001	0.023	1.000	1.000	
		F	1.000	1.000	0.989	0.459	0.625
	3	B	0.185				
		C	1.000	0.490			
		D	1.000	0.769	1.000		
		E	0.668	1.000	1.000	1.000	
		F	1.000	0.090	1.000	1.000	0.845
Corrected based on slices including the infarct, as % of ipsilateral hemisphere	1	B	1.000				
		C	1.000	1.000			
		D	1.000	1.000	1.000		
		E	1.000	1.000	1.000	1.000	
		F	0.189	0.003	0.025	1.000	0.049
	2	B	0.308				
		C	0.066	1.000			
		D	1.000	1.000	1.000		
		E	< 0.001	0.029	1.000	1.000	
		F	1.000	1.000	1.000	0.427	0.608
	3	B	0.128				
		C	1.000	0.353			
		D	1.000	0.530	1.000		
		E	0.546	1.000	1.000	1.000	
		F	1.000	0.059	1.000	1.000	0.816
Corrected based on slices including the infarct, as % of contralateral hemisphere	1	B	1.000				
		C	1.000	1.000			
		D	1.000	1.000	1.000		
		E	1.000	1.000	1.000	1.000	
		F	0.192	0.003	0.018	1.000	0.049
	2	B	0.358				
		C	0.060	1.000			
		D	1.000	1.000	1.000		
		E	< 0.001	0.017	1.000	1.000	
		F	1.000	1.000	0.894	0.507	0.644
	3	B	0.318				
		C	1.000	0.784			
		D	1.000	1.000	1.000		
		E	0.860	1.000	1.000	1.000	
		F	1.000	0.172	1.000	1.000	0.899

All values are *p*-values except for 'Case'. Corresponding analysis of variance statistics are given in Table 6.5. *p*-values less than 0.05 are shown in bold. 'A'–'F' indicate the scanner.

Table C.4 *U* statistics for Mann–Whitney *U* tests.

Infarct volume measure	Case	A	B	C	D	E	F
Uncorrected, in mm ³	1	13	11	14	12	13	14
	2	12	13	13	11	13	11
	3	12	14	12	11	12	14
Corrected based on all slices							
in mm ³	1	12	10	13	11	13	13
	2	12	13	12	9	13	9
	3	13	15	12	11	14	13
as % of brain	1	14	12	14	13	15	14
	2	13	14	14	12	14	12
	3	14	20	13	13	15	14
as % of ipsilateral hemisphere	1	14	11	14	13	17	14
	2	13	14	14	12	14	14
	3	14	20	13	13	16	15
as % of contralateral hemisphere	1	14	13	14	13	15	14
	2	12	15	14	12	15	13
	3	13	15	13	13	13	14
Corrected based on slices including the infarct							
in mm ³	1	13	10	13	10	13	9
	2	13	13	12	10	13	9
	3	10	14	10	10	12	13
as % of brain	1	19	18	21	19	30	17
	2	21	21	16	20	22	22
	3	18	23	19	16	26	19
as % of ipsilateral hemisphere	1	19	19	23	21	31	17
	2	23	21	17	22	22	23
	3	18	26	18	17	28	20
as % of contralateral hemisphere	1	17	17	21	17	27	17
	2	20	20	17	20	22	21
	3	18	21	18	15	24	19

All values are *U* statistics except for 'Case'. Corresponding *p*-values are given in Table 6.7. 'A'–'F' indicate the scanner.

Appendix D

Phantom Scanning Form

Form for Structural Phantom Scanning in Preclinical MRI Centres

Please follow the instructions below for scanning the structural MRI phantom and complete the fields in grey. The whole procedure should normally take around 2 hours. The completed form along with the sets of scout scans and the 12 sets of protocol scans in the standard DICOM format should be saved in a Dropbox folder named '*Phantom scanning, <your centre>, <date>*'. Share this folder with [email address].

Centre: Enter name of MRI centre.

Country: Enter country.

MR scanner manufacturer: Enter manufacturer.

Model: Enter model of the scanner.

Magnetic field strength (T): Enter field strength.

Software, version: Enter software and version.

Phantom positioning and alignment

1. The phantom should be tightly attached on the radiofrequency (RF) coil using tape or otherwise, with its four knobs facing upwards (anterior direction). It must be perfectly aligned so that the three orthogonal directions in the MRI system are perpendicular to the sides of the phantom. Do not place the phantom into another object with solution (e.g. a tube); only the phantom's internal volume must be visible in the images. Exemplar datasets are shown in Figures 1–3.
2. Place the RF coil in the gradient coil, so that the phantom is at the magnet's isocentre. Perform shimming and localisation (see recommended sequence below). If scout images indicate that the phantom is malposed or misaligned, adjust accordingly and re-scan.
3. Save the images in a separate folder for each scan. Name the folders '*Scout_<scan number>*'.

Correct positioning and alignment are necessary for accurate and reliable analysis of the acquired images.

Scout (localiser) imaging

A gradient echo sequence with the following parameters is proposed. However, you may use an alternative.

Type of sequence	Gradient echo		
TR (ms)	20	FOV (mm×mm)	25×25
TE (ms)	3	Matrix (pixels×pixels)	128×128
Flip angle (°)	20	Slice thickness (mm)	2 (gap 2)
NSA/NEX	1	Number of slices	3×3 planes
Receive bandwidth (kHz)	50	Scan time (m:s)	0:8

Imaging using two different protocols

1. The phantom must be in the centre of the acquired slices, which must be of a sufficient number to cover the whole phantom volume. For this reason, the middle slice (e.g. no. 9 for 17 slices) must be at the centre of the phantom (see Figures 1–3).
2. Scan the phantom using **two different protocols (A and B below)**. The sequence for each protocol must be repeated six times, corresponding to different combinations of imaging planes and signal encoding directions. Specifically, scan the phantom in the **axial**, **coronal**, and **sagittal** planes. Then, swap the frequency and phase encoding directions for each of these and scan the phantom again (here called **axial90**, **coronal90**, and **sagittal90** planes respectively). Figures 1–3 show exemplar images acquired using protocol A.
3. Save the images in a separate folder for each plane, named '*Protocol<A or B>_<plane>*'.

A. Standardised structural protocol

Scan the phantom using the type of RF coil and the exact sequence indicated below. If, for any reason, this type of coil or any of the prescribed parameters cannot be used, replace with a similar one. In this case, report the value used and briefly the reason for substitution, as well as the new scan time (if different) in the 'Comments' field at the bottom of this page. However, please retain the suggested imaging resolution (voxel size).

COILS: *Set-up the scanner for rat brain imaging using a surface RF coil. Provide the details below.*

Gradient

Maker, model: Enter maker and model.
 Outer/inter diameter (mm): Enter OD/ID.
 Max gradient amplitude (mT/m): Enter G_{max} .
 Max slew rate (T/m/s): Enter $SlewRate_{max}$.

Radiofrequency (RF)

Maker, model: Enter maker and model.
 Type and specifications (transmit and receive): Enter coil(s) details (e.g. 72mm ID birdcage transmit, 2 channel rat brain phased-array surface receive).

SEQUENCE:

Type of sequence	Fast Spin Echo (FSE) / Rapid acquisition with Relaxation Enhancement (RARE)		
TR (ms)	1600	FOV (mm×mm)	19.2×19.2
TE_{eff} (ms)	20	Matrix (pixels×pixels)	256×256
Echo train length	4	Slice thickness (mm)	1 (gapless)
NSA/NEX	2	Number of slices	17
Receive bandwidth (kHz)	100	Scan time (m:s)	3:28
Additional scan options:	—		

B. Centre's favourite protocol for infarct size assessment in a rat stroke model

Scan the phantom using a scanning protocol commonly used in your centre for assessing the infarct size in a rat stroke model. Use a different RF coil from the above, if relevant. Use a sufficient number of slices to cover the whole phantom; this number should be odd so that a middle slice is placed at the centre of the phantom.

COILS: Use your preferred set-up for rat brain imaging. Provide the details below where different from above.

Gradient

Maker, model: Enter maker and model.

Outer/inter diameter (mm): Enter OD/ID.

Max gradient amplitude (mT/m): Enter G_{max} .

Max slew rate (T/m/s): Enter $SlewRate_{max}$.

Radiofrequency (RF)

Maker, model: Enter maker and model.

Type and specifications (transmit and receive): Enter coil(s) details (e.g. 72mm ID birdcage transmit, 2 channel rat brain phased-array surface receive).

SEQUENCE:

Type of sequence

Enter type of sequence.

TR (ms)

Enter TR.

FOV (mm×mm)

Enter FOV.

TE or TE_{eff} (ms)

Enter TE.

Matrix (pixels×pixels)

Enter matrix size.

Echo train length

Enter ETL, if relevant.

Slice thickness (mm)

Enter slice thickness.

NSA/NEX

Enter averages.

Number of slices

Enter number of slices.

Receive bandwidth (kHz)

Enter bandwidth.

Scan time (m:s)

Enter scan time.

Additional scan options: Enter other important information (e.g. flip angle), if relevant.

Comments: Enter your comments, if any.

Scanned by: Enter your full name.

Date: Select date of scanning.

Email address: Enter your email address.

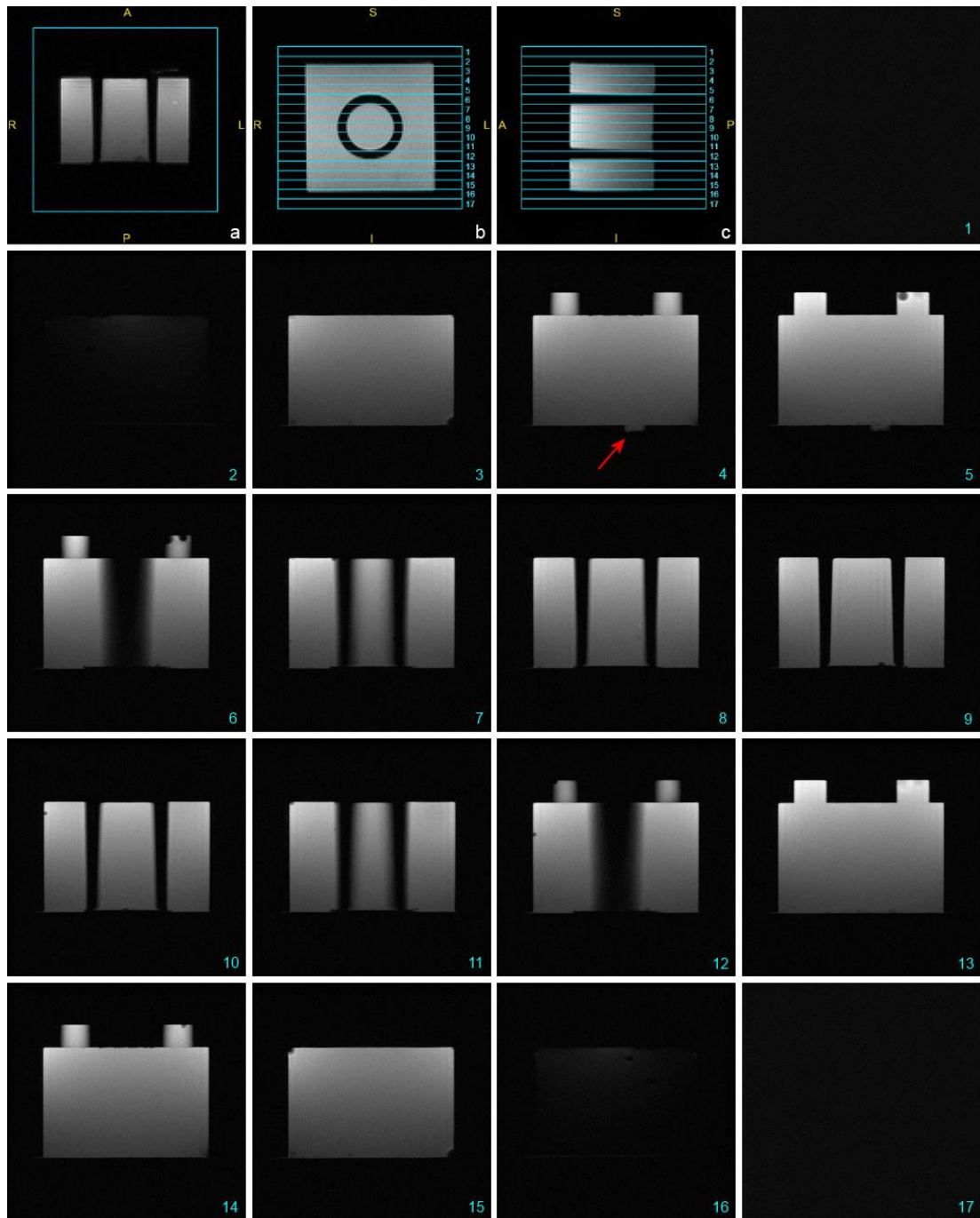


Figure 1. Scout images showing the slice selection for imaging in the **axial** and **axial90** planes using **protocol A (a-c)**, and an example set of acquired slices (**1-17**). The middle slice (**9**) should intersect the centre of the phantom. The red arrow shows the position of the orientation mark. Due to possible gradient scaling differences between scanners the phantom volume might look slightly different, while small air bubbles inevitably can be present in the phantom.

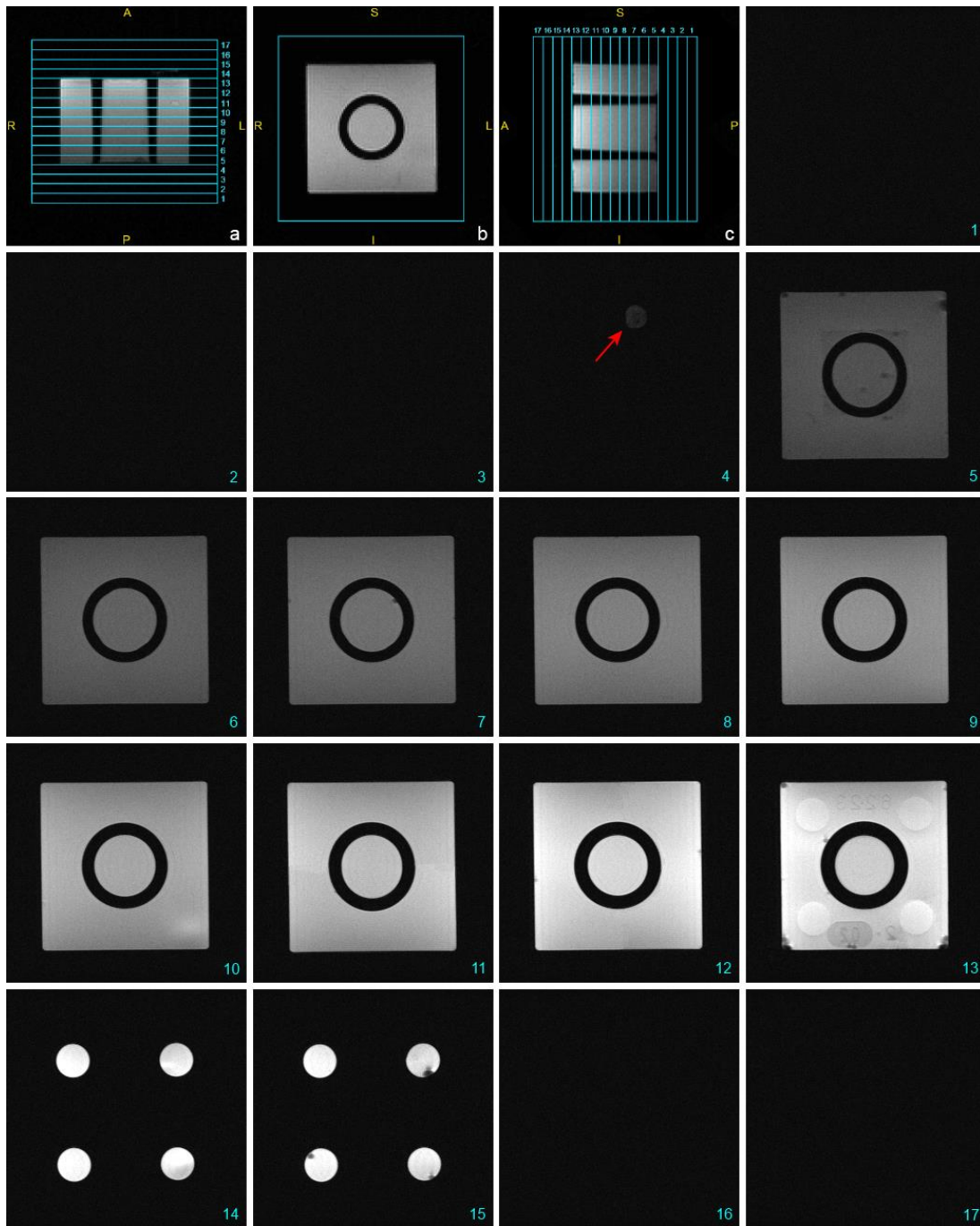


Figure 2. Scout images showing the slice selection for imaging in the **coronal** and **coronal90** planes using **protocol A (a-c)**, and an example set of acquired slices (**1-17**). The middle slice (**9**) should intersect the centre of the phantom. The red arrow shows the position of the orientation mark. Due to possible gradient scaling differences between scanners the phantom volume might look slightly different, while small air bubbles inevitably can be present in the phantom.

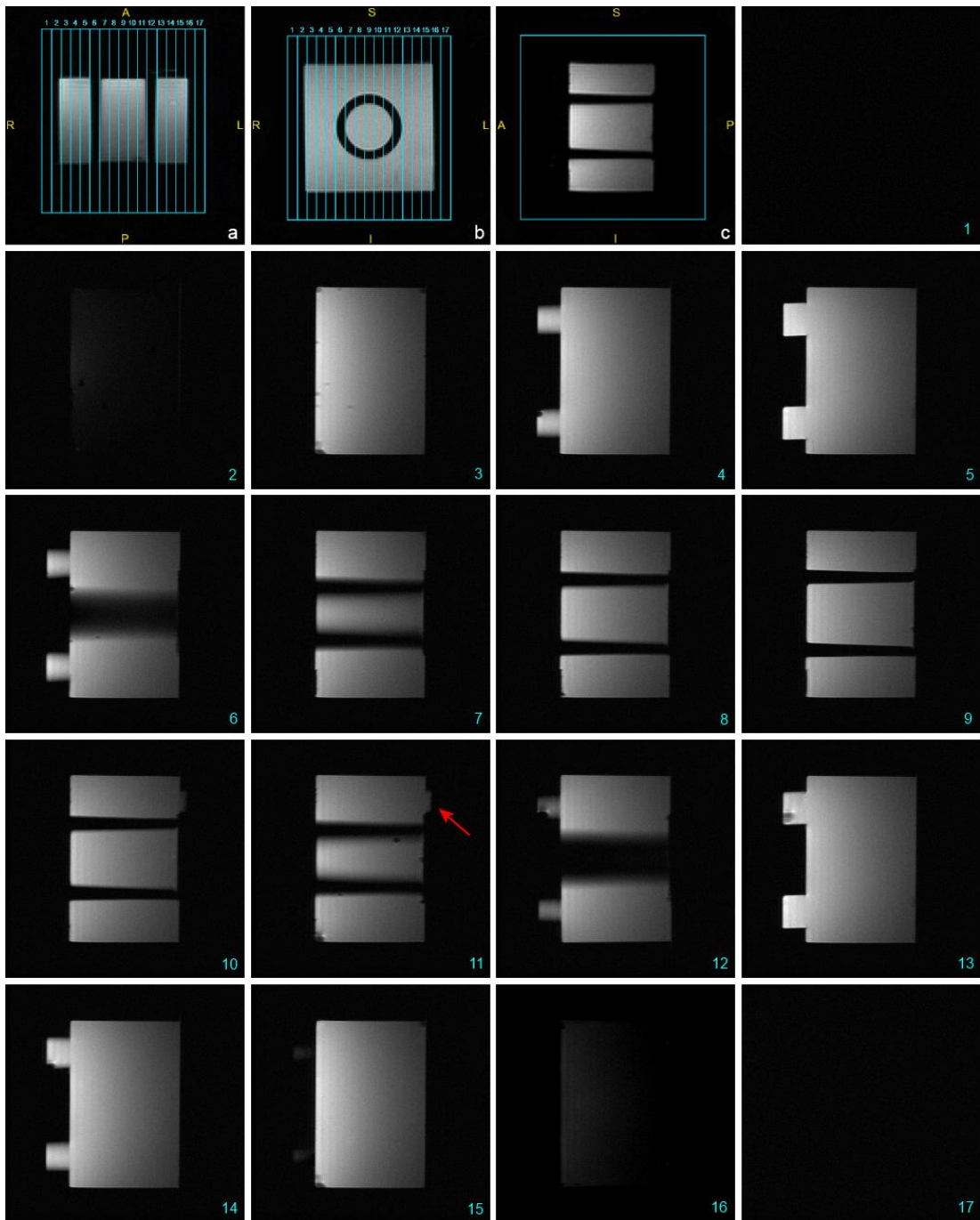


Figure 3. Scout images showing the slice selection for imaging in the **sagittal** and **sagittal90** planes using **protocol A (a-c)**, and an example set of acquired slices (**1-17**). The middle slice (**9**) should intersect the centre of the phantom. The red arrow shows the position of the orientation mark. Due to possible gradient scaling differences between scanners the phantom volume might look slightly different, while small air bubbles inevitably can be present in the phantom.