Optimising service organisation for stroke patients

Lindsay Jane Walker Govan

A Dissertation Submitted to the

University of Glasgow

for the Degree of

Doctor of Philosophy

Cardiovascular and Medical Sciences

Faculty of Medicine

May 2009

Abstract

Background

Stroke is the leading cause of long-term neurological disability in adults and the third most common cause of death in Britain. It is well known that in addition to the patient characteristics of age and severity, the treatment a stroke patient receives in hospital significantly affects outcome. The effectiveness of complex service interventions, how the benefits of these interventions are achieved and the economic impact of different types of service delivery were explored.

Methods

The Stroke Unit Trialists' Collaboration systematic review was updated and currently contains 31 clinical trials (6936 subjects). The aims were explored using various basic frequentist and Bayesian meta-analysis techniques as well as more complex meta-analysis ideas. These more complex ideas include: meta-regression where covariate information is incorporated into the model; and network meta-analysis where direct and indirect information is used in a mixed treatment comparisons model while also incorporating covariate information.

Results

Organised inpatient (stroke unit) care showed reductions in death, death or dependency and death or institutional care compared to general medical wards. Stroke unit care appears to reduce the risk of adverse outcomes through prevention and treatment of complications. Acute, comprehensive and rehabilitation stroke unit care appeared to be most effective and acute stroke unit care appeared to be the most cost-effective. However, acute followed by rehabilitation stroke unit care, if required, appears to be the most cost-effective pathway of care compared to the other pathways analysed.

Discussion

Future research should focus on rehabilitation, acute and comprehensive systems of inpatient care, and explore the best ways of preventing and managing specific complications. Effort should be made to make individual patient data and information on the care pathway of a stroke patient available for meta-analysis.

Acknowledgements

Declaration

This thesis has been composed by myself and it has not been submitted in any previous application for a degree. The work reported within was executed by myself, with the exceptions noted in the Acknowledgements.

Several presentations of work described in this thesis have been made at UK and international meetings, including the 2006 and 2007 European Stroke Conferences (ESC) (Govan, Langhorne, Weir and for the Stroke Unit Trialists' Collaboration, 2006; Govan, Weir, Muir, Lees and for the IMAGES Study Group, 2006; Govan et al., 2007b; Govan and for the Stroke Monitoring Trialists, 2007; Govan, Weir, Langhorne, Ades and for the Stroke Unit Trialists' Collaboration, 2008).

Additionally, some of this work has already been published. The updated Stroke Unit Trialists' Collaboration's systematic review (Stroke Unit Trialists' Collaboration, 2007), the results of which are described in Chapter 4, was also summarised in Govan, Weir, Langhorne and for the Stroke Unit Trialists' Collaboration (2008). The analysis in Chapter 5 investigates whether organised inpatient care improves patient outcomes by exploring the use of interventions to prevent complications was published in Govan et al. (2007a).

Contents

2 Consistency and validity of the IMAGES Stroke Scale 27

3 Categorising stroke severity using different stroke scales 43

List of Tables

xiii

xiv

List of Figures

xviii

List of abbreviations

AD   Aggregate data

ADL   Activities of Daily Living

AI   Activity Index

ASU   Acute stroke unit

AUC   Area under curve

BI   Barthel Index

CCT   Controlled clinical trial

CI   Confidence interval

CrI   Credible interval

DIC   Deviance information criterion

ESD   Early supported discharge

GCS   Glasgow Coma Scale

ICER   Incremental cost-effectiveness ratio

IMAGES   Intravenous Magnesium Efficacy in Stroke

IPD   Individual patient data

LACS   Lacunar stroke

mNIHSS   Modified NIH Stroke Scale

mRS   Modified Rankin Scale

MTC	Mixed treatment comparisons
NHS	National Health Service
NIHSS	National Institutes of Health Stroke Scale
OCSP	Oxfordshire Community Stroke Project Classification
OR	Odds ratio
PACS	Partial anterior circulation stroke
POCS	Posterior circulation stroke
QALY	Quality-adjusted life-year
RCT	Randomised controlled trial
ROC	Receiver operating characteristic
SMD	Standardised mean difference
SP	Stroke progression
SSS	Scandinavian Stroke Scale
SU	Stroke unit
SUTC	Stroke Unit Trialists' Collaboration
TACS	Total anterior circulation stroke
WHO	World Health Organisation
WMD	Weighted mean difference

Summary

This thesis assesses the effectiveness of complex service interventions and the impact of different types of service delivery. Chapter 1 introduces stroke, different models of inpatient service delivery and current methods of assessing severity in stroke patients. Additionally, meta-analysis and some basic meta-analysis theory are described and finally the data sources used throughout this thesis are summarised.

Chapters 2 and 3 investigate methods of measuring stroke severity and their ability to predict outcome. Chapter 2 introduces a novel stroke severity scale and its internal consistency, construct validity and predictive ability are assessed with comparisons made to current methods of measuring severity. Chapter 3 investigates the amount of information retained when stroke severity is categorised and assesses the equivalence of four stroke scale categorisations in current use.

Chapters 4 to 8 investigate the effectiveness of organised inpatient (stroke unit) care by use of increasingly advanced meta-analysis techniques.

Chapter 4 uses standard frequentist meta-analysis to compare broadly "more" organised care with "less" organised care to determine if organised inpatient care is more beneficial than less organised care. Chapter 5 introduces a simple

Bayesian approach to meta-analysis to examine how organised inpatient care improves patient outcomes by exploring the use of interventions to prevent complications. Chapter 6 expands the Bayesian meta-analysis model used in Chapter 5 by re-creating individual patient data from available grouped data and introduces covariate effects into the model. This Chapter examines whether routine automated monitoring for and treatment of physiological complications reduce adverse outcomes in stroke patients.

Chapter 7 describes a complex network meta-analysis model with covariate effects. The model allows the inclusion of studies with the joint distribution of covariates, those with marginal data for one or more covariates as well as those with no covariate information and provides an estimate of the covariate effects. Treatment effects are estimated using direct and indirect comparisons giving a more powerful comparison between different treatment types. This Chapter assesses the possible bias in treatment effect estimation when covariates are not accounted for in meta-analyses.

Chapter 8 applies the model described in Chapter 7 to determine whether any one system of inpatient care is most effective in improving patient outcomes while also estimating the effect of age and severity on patient outcome. Finally, Chapter 9 implements a cost-utility analysis to determine which pathway of inpatient care is most cost-effective.

The final Chapter discusses overall conclusions, alternative statistical and economical analyses that could be performed and the possibilities for future work.

Chapter 1

Introduction

1.1 Stroke

The World Health Organisation (WHO) defines stroke as 'rapidly developed clinical signs of focal or global disturbance of cerebral function lasting more than 24 hours or until death, with no apparent non-vascular cause' (WHO MONICA Project Principal Investigators, 1988). Essentially, stroke is the brain equivalent of a heart attack: a disruption of the blood supply to the brain through a blockage (ischaemic) or rupture (haemorrhagic) causing damage to the surrounding tissues. Symptoms and signs include sudden numbness and tingling; weakness in face, arm or leg especially on one side of the body; confusion; difficulty talking or understanding speech; trouble seeing in one or both eyes; trouble walking, dizziness, loss of balance or co-ordination; severe headaches without any known cause.

Stroke is the leading cause of long-term neurological disability in adults (Wolfe, 2000), the second leading cause of death worldwide (Murray and Lopez, 1997)

and the third most common cause of death in the Britain (Wolfe, 2000). Malmgren et al. (1989) have estimated that the total number of patients with first ever stroke would increase by approximately 30% between 1983 and 2023, and Morris et al. (2003) found that compared to the British mean in the British Regional Heart Study (3.54 first events per 1000 person-years), Scottish towns had a higher incidence of stroke (Ayr: 3.92; Dunfermline: 4.02; and Falkirk: 5.45)

Given these figures, the treatment of stroke is the focus of much clinical research both globally and nationally in order to reduce deaths and the dependency of survivors. Several systems of inpatient (stroke unit) care exist to treat patients following stroke, however, it is not clear which system is most effective or how the benefits of these systems of care are achieved. The aim of this thesis is to explore how inpatient (stroke unit) services improve patient outcomes and which inpatient service is most effective and most cost-effective.

1.2 Stroke services

One of the major problems in the treatment of stroke is that there is no powerfully effective drug intervention available for all stroke patients. Good inpatient and outpatient services are therefore required following stroke.

1.2.1 Inpatient services

Organised inpatient (stroke unit) care is the mainstay of stroke service. It is a complex organisational intervention comprising nurses, doctors and therapists who specialise in looking after stroke patients in hospital and work as a coordinated team to provide a complex package of care to stroke patients (Stroke Unit

Trialists' Collaboration, 2007). There are several forms of stroke unit care, some of which are subject to more specialised organisation than others. This leads to a hierarchy of service organisation which in descending order of complexity are:

1. Stroke ward: a multidisciplinary team including specialist nursing staff based in a discrete ward caring exclusively for stroke patients. This category can be subdivided further:

   a) acute stroke units which accept patients acutely but discharge early (usually within seven days). These can be broadly subcategorised:

      i) 'intensive' model of care with continuous monitoring, high nurse staffing levels and the potential for life support;

      ii) 'semi-intensive' model with continuous monitoring, high nurse staffing levels but no life support facilities; and

      iii) 'non-intensive' with none of the above.

   b) comprehensive stroke units (a combination of acute and rehabilitation) which accept patients acutely but also provide rehabilitation for several weeks if necessary.

   c) rehabilitation stroke units that accept patients after a delay, usually of seven days or more, and focus on rehabilitation;

2. Mixed rehabilitation ward: a multidisciplinary team including specialist nursing staff in a ward providing a generic rehabilitation service not exclusively for stroke patients.

3. Mobile stroke team: a multidisciplinary team (excluding specialist nursing staff) providing care in a variety of settings.

4. General medical wards: care in an acute medical or neurology ward without routine multidisciplinary input.

1.2.2 Discharge services

Following care in hospital, a range of different discharge services have been developed. The main aim of more organised discharge services compared to conventional systems is to accelerate the return home of stroke patients who are admitted to hospital. Early Supported Discharge Trialists (2008) describe a hierarchy of discharge service approaches, which in descending order of level of support are:

1. Early supported discharge (ESD) team co-ordination and delivery: a multidisciplinary team which co-ordinates discharge from hospital, post discharge care and provides rehabilitation and patient care at home. The multidisciplinary team meets on a regular basis to plan patient care.

2. ESD team coordination: a multidisciplinary team plans and supervises the discharge home and the immediate post-discharge care. However, care is subsequently handed over to existing community-based agencies who provide continuing rehabilitation and support at home. Community-based agencies do not usually provide co-ordinated multidisciplinary team care (that is, no input from a multidisciplinary team which would meet on a regular basis to plan patient care).

3. No ESD: patients have access to multidisciplinary team care in hospital but this ends at hospital discharge. Their subsequent care is provided by a

range of community stroke services which are not planned or provided by
a co-ordinated team or are provided by trained healthcare volunteers.

1.3   Stroke scales

Stroke is a complex condition which can affect many aspects of body function.
Measuring the impact of stroke may therefore be problematic. Stroke scales are
assessment tools designed to quantify different aspects of the effect of stroke, re-
covery and impairment following stroke. There does not exist a gold standard
scale which incorporates all criteria of an ideal stroke scale (Lyden and Hanston,
1998). Therefore, several separate scales are generally recorded for each patient,
all of which fall into one of four broad categories: pathology; impairment; dis-
ability (activity); or handicap (participation).

Pathology describes the diagnosis of stroke, therefore pathological scales are
usually measured using imaging or histology to categorise patients broadly. Im-
pairments are usually described as the symptoms and signs of stroke so impair-
ment scales numerically record specific findings of detailed neurological exami-
nation. This type of scale is particularly important in assessing patients almost
immediately after stroke onset. Disability, or activity, refers to the personally
meaningful functions or activities the patient can achieve, such as bathing and
dressing. These scales measure functional outcome and are used to assess stroke-
related disability and performance in occupational functions. Finally, handicap,
or participation, is the change in social position following stroke. These scales
are the most difficult to measure (Wolfe, 2000).

All stroke scales have their advantages and disadvantages as discussed in

Kasner (2006) and Lyden and Hanston (1998). Below follows a more detailed description of some of the more commonly used stroke scales.

1.3.1 Oxfordshire Community Stroke Project

The Oxfordshire Community Stroke Project (OCSP) classification is a pathological scale which attempts to classify patients by site and size of lesion. Presenting neurological symptoms and signs are used to distribute patients into one of four groups: lacunar stroke (LACS); total anterior circulation stroke (TACS); partial anterior circulation stroke (PACS); and posterior circulation stroke (POCS). The fourth letter at the end of each acronym may denote the type of stroke: syndrome (S) with indeterminate pathogenesis, prior to imaging; or infarct (I). Table 1.1 describes the definition of each group of the OCSP (Bamford et al., 1991).

The OCSP is useful in stratifying patients in clinical trials, especially in the acute phase since patients can be classified quickly and reliably so that appropriate treatments and therapies may be administered (Lindley et al., 1993). However, in the very acute stage classifications may change over time due to the

Table 1.1: The Oxfordshire Community Stroke Project classification.

Group	Description
LACS	Pure motor stroke, pure sensory stroke, sensori-motor stroke, or ataxic hemiparesis
TACS	Present with a triad of hemiparesis, dysphasia and homonymous hemianopia
PACS	Present with two of the features of TACS, or isolated dysphasia or parietal lobe signs
POCS	Brain stem or cerebellar signs, and/or isolated homonymous hemianopia

occurrence or resolution of neurological signs (Mead et al., 2000).

1.3.2 National Institutes of Health Stroke Scale

The National Institutes of Health Stroke Scale (NIHSS) was specifically designed to be used by non-neurologists. It is a 15 item neurological impairment scale with a maximum deficit of 42 points, originally devised by Brott et al. (1989). The NIHSS has been modified to give the version in current use (Lyden et al., 1999) which is described in Table 1.2. The scale measures key components of a standard neurological examination including eye movement, motor and sensory involvement and level of consciousness.

A simpler, modified version, the mNIHSS, has been proposed (Lyden et al., 2001) where the items consciousness, facial palsy, ataxia and dysarthia have been dropped. However, this has not been widely adopted.

A drawback of the NIHSS is that patients with identical scores could have quite distinct clinical diagnoses. However, the NIHSS is simple, valid, reliable and well established in the stroke community, with video training available to teach new carers how to apply it (Lyden et al., 2005).

1.3.3 Scandinavian Stroke Scale

The Scandinavian Stroke Scale (SSS) is an impairment scale designed as both an initial prognostic and long-term functional score for use by non-neurologists. The initial prognostic score includes consciousness, eye movement and motor power, while the long-term functional score incorporates items of functional significance to the patient such as speech and facial palsy (Scandinavian Stroke Study Group,

Table 1.2: Current form of the National Institutes of Health Stroke Scale.

Item	Name	Response	Item	Name	Response
1a	Level of consciousness	0 = Alert 1 = Not alert, arousable 2 = Not alert, obtunded 3 = Unresponsive	6a	Left motor leg	0 = Normal 1 = Drift before 5 seconds 2 = Falls before 5 seconds 3 = No effort against gravity 4 = No movement
1b	Questions	0 = Answer both correctly 1 = Answers one correctly 2 = Answers neither correctly	6b	Right motor leg	0 = Normal 1 = Drift before 5 seconds 2 = Falls before 5 seconds 3 = No effort against gravity 4 = No movement
1c	Commands	0 = Performs both tasks correctly 1 = Performs one task correctly 2 = Performs neither task			
			7	Ataxia	0 = Absent 1 = One limb 2 = Two limbs
2	Gaze	0 = Normal 1 = Partial gaze palsy 2 = Total gaze palsy			
			8	Sensory	0 = Normal 1 = Mild loss 2 = Severe loss
3	Visual fields	0 = No visual loss 1 = Partial hemianopsia 2 = Complete hemianopsia 3 = Bilateral hemianopsia	9	Language	0 = Normal 1 = Mild aphasia 2 = Severe aphasia 3 = Mute or global aphasia
4	Facial palsy	0 = No drift 1 = Minor paralysis 2 = Partial paralysis 3 = Complete paralysis	10	Dysarthria	0 = Normal 1 = Mild 2 = Severe
5a	Left motor arm	0 = Normal 1 = Drift before 10 seconds 2 = Falls before 10 seconds 3 = No effort against gravity 4 = No movement	11	Extinction/inattention	0 = Normal 1 = Mild 2 = Severe
5b	Right motor arm	0 = Normal 1 = Drift before 10 seconds 2 = Falls before 10 seconds 3 = No effort against gravity 4 = No movement	Total	42	

1985). All items of the SSS can also be used so that the impairment of a patient can be measured as numerical value between 0 and 58, where lower scores indicate greater impairment. Shortened versions of the SSS, often labelled prognostic and long-term scores, can also be used and have total scores of 22 and 48, respectively. The full, prognostic and long-term scores are described in Table 1.3.

Table 1.3: Full, initial prognostic and long-term functional scores of the Scandinavian Stroke Scale.

Function	Response	Prognostic Score	Long-term Score
Consciousness	6 = Fully conscious 4 = Somnolent, can be awakened to full consciousness 2 = Reacts to verbal command, but is not fully conscious	✓	
Eye movement	4 = No gaze palsy 2 = Gaze palsy present 0 = Conjugate eye deviation	✓	
Arm, motor power*	6 = Raises arm with normal strength 5 = Raises arm with reduced strength 4 = Raises arm with flexion in elbow 2 = Can move, but not against gravity 0 = Paralysis	✓	✓
Hand, motor power*	6 = Normal strength 4 = Reduced strength in full range 2 = Some movement, fingertips do not reach palm 0 = Paralysis		✓
Leg, motor power*	6 = Normal strength 5 = Raises straight leg with reduced strength 4 = Raises leg with flexion of knee 2 = Can move, but not against gravity 0 = Paralysis	✓	✓
Orientation	6 = Correct for time, place and person 4 = Two of these 2 = One of these 0 = Completely disoriented		✓
Speech	10 = No aphasia 6 = Limited vocabulary or incoherent speech 3 = More than yes/no, but not longer sentences 0 = Only yes/no or less		✓
Facial palsy	2 = None/dubious 0 = Present		✓
Gait	12 = Walks 5m without aids 9 = Walks with aids 6 = Walks with help of another person 3 = Sits without support 0 = Bedridden/wheelchair		✓
Total	58	22	48

* Motor power is assessed only on the affected side

Although the SSS is intended for use in conscious patients only, it has good interobserver agreement and the initial long-term scores correlate well with outcome at threes months (Scandinavian Stroke Study Group, 1987).

1.3.4 Barthel Index

The Barthel Index (BI) (Mahoney and Barthel, 1965), given in Table 1.4 is a measure of activities of daily living (ADL) competence. It is a 10 item disability, or activity, scale ranging from 0 to 100, where lower scores indicate greater dependency. Occasionally, this score may be re-scaled from 0 to 20, with each item weighting divided by five.

The main advantage of this scale is that patients can be evaluated by direct observation or by telephone interviews or reports of nurses and staff (Lyden and Hanston, 1998). One of the major disadvantages of the BI is that a "ceiling effect" can occur, where a patient with a maximal score may still be disabled in a such a way that they are unable to live independently (Dromerick et al., 2003).

1.3.5 Modified Rankin Scale

Originally, the Rankin Scale was argued to be a measure of handicap, or participation. However, the Rankin Scale is now generally considered to be a global or mixed assessment scale which groups patients into five categories based on their ability to perform previous activities and their requirements for assistance (Rankin, 1957). The modified Rankin Scale (mRS), described in Table 1.5, is a seven point scale that combines the five points of the Rankin Scale with additional levels for no dependence and death (van Swieten et al., 1988).

The mRS allows easy and rapid assessment of the effect of stroke, with a one-point shift in the scale often considered to be clinically important. However, the use of broad categories may reduce sensitivity to change and scoring is determined by the overall impression of the functional ability of a patient by the examiner

Table 1.4: The Barthel Index.

Item	Response
Bowels	0 = Incontinent (or need to be given enema)
	5 = Occasional accident (once per week)
	10 = Continent
Bladder	0 = Incontinent, or catheterised and unable to manage
	5 = Occasional accident (max once per 24 hours)
	10 = Continent (for more than 7 days)
Grooming	0 = Needs help with personal care
	5 = Independent face/hair/teeth/shaving
Toilet use	0 = Dependent
	5 = Needs some help, but can do something alone
	10 = Independent (on and off, dressing, wiping)
Feeding	0 = Unable
	5 = Needs help cutting, spreading butter, etc
	10 = Independent (food provided in reach)
Transfer	0 = Unable, no sitting balance
	5 = Major help (one or two people, physical), can sit
	10 = Minor help (verbal or physical)
	15 = Independent
Mobility	0 = Immobile
	5 = Wheelchair independent, including corners, etc
	10 = Walks with help of one person (verba or physical)
	15 = Independent (but may use any aid, e.g. stick)
Dressing	0 = Dependent
	5 = Needs help, but can do about half unaided
	10 = Independent (including buttons, zips, laces, etc)
Stairs	0 = Unable
	5 = Needs help (verbal, physical, carrying aid)
	10 = Independent up and down
Total	100

(Dromerick et al., 2003). More precise algorithms have been developed to improve the reliability of rating the Rankin Scale (Wilson et al., 2002).

Table 1.5: The modified Rankin Scale.

Level
0 = No symptoms
1 = No significant disability, despite symptoms Able to perform all usual duties and activities
2 = Slight disability Unable to perform all previous activities but able to look after own affairs without assistance
3 = Moderate disability Requires some help, but able to walk without assistance
4 = Moderately severe disability Unable to walk without assistance and unable to attend to own bodily needs without assistance
5 = Severe disability Bedridden, incontinent, and requires constant nursing care and attention
6 = Dead

1.4   Meta-analysis

In medical research, information on the efficacy of a treatment may be available from a number of clinical studies with similar treatment protocols. However, when the studies are considered separately they may be either too small or too limited in scope to come to a general conclusion about the effect of the treatment (DerSimonian and Laird, 1986). Meta-analysis, defined as the statistical analysis of a large collection of analytic results for the purpose of integrating the findings (Dickersin and Berlin, 1992), attempts to combine the findings across such studies in order to gain statistical power, strengthen the evidence about possible treatment effects and, in adequately powered studies, to find out more about subgroups and possible interactions.

'Systematic review' is a phrase that is often used to describe the whole process of collecting and analysing data. One of the first stages of a systematic review is the search for appropriate trials. However, since not all relevant trials may be published, the search must also include unpublished and ongoing trials. Therefore the search for appropriate studies to include in the analysis must be as thorough as possible.

Once appropriate trials have been found, the quality of the trial must also considered. Important aspects to consider are randomisation, allocation concealment, completeness of follow-up, presence of intention-to-treat analysis and blinded assessment. Trials that do not have adequate levels of any of these may still be included in the meta-analysis, but a sensitivity analysis excluding such trials should also be performed to identify possible bias in the results.

Meta-analysis using individual patient data (IPD) is widely regarded as the gold-standard approach to meta-analysis. If covariates are of interest the use of IPD typically has greater power than meta-regression incorporating study level covariates, which represent the average covariate response of each study. However, there is great effort involved to obtain IPD. Issues of ownership and access to data need to be addressed as well as the time and effort involved in the collection of data. It is often unclear whether benefits gained in using IPD outweigh the extra cost involved (Sutton and Higgins, 2008).

Meta-analysis has become increasingly popular not only in medical research but also statistical research. Several approaches to meta-analysis have been developed, each with its advantages and disadvantages. Methodology includes fixed-effects or random-effects models and frequentist or Bayesian approaches.

In a fixed-effects model the result from each study included in the analysis is assumed to estimate the same quantity with any deviations due to random sampling variability. Although this may sound reasonable it does not take into account the possible heterogeneity between trials such as slightly different types of patients, or slightly different treatment regimes. Therefore it may be unreasonable to assume that any differences between trials are due to random variation alone. A random-effects model on the other hand relaxes this assumption. Instead the model allows the treatment effects to differ from each other but assumes they are drawn from a common distribution of effect sizes. This distribution is usually assumed to be the Normal distribution, with a variance determined by the data. Accounting for between-study heterogeneity often leads to similar pooled-estimates to those obtained by a fixed-effects method, however, confidence intervals are wider and so the estimates are more conservative (Sutton et al., 2001).

There are two possible approaches for deciding between fixed- and random-effects models. Firstly, model choice could be informed by a test for heterogeneity, such as the Q or I^2 statistics (Higgins and Thompson, 2002). That is, if the test is non-significant then a fixed-effects model may be appropriate, whereas if the test is significant, a random-effects model should be used. However, these heterogeneity tests often have low power, implying that heterogeneity may exist when the test produces a non-significant result (Boissel et al., 1989). The other approach is always to use a random-effects model since the degree of variation between studies will determine the width of the confidence interval. Therefore, when between study variation is low, the inflation in the confidence interval may be negligible, producing a result similar to the fixed-effects method (Sutton et

al., 2001).

Frequentist approaches to meta-analysis generally calculate a weighted average of results obtained from individual studies. Bayesian methods, however, consider both the data and the model parameters to be random quantities about which there is uncertainty, instead of constant values which, in the case of parameters, are unknown. The likely values of these parameters are described through a probability distribution that quantifies this uncertainty (Schmid, 2001). Using this approach allows all parameter uncertainty to be automatically accounted for in the analysis. Also, in the meta-analysis context, individual trials can "borrow strength" from other trials, meaning that with every iteration the updated estimates take into account the results from all other trials in the analysis, giving a better estimate for the individual trial effects. Interpretation of Bayesian credible intervals is simpler than that of the frequentist confidence intervals since direct probability statements can be made regarding the model parameters, the quantities of interest. Additionally, prior knowledge of possible effects may be incorporated into the analysis. However, use of different prior distributions can change the conclusions and so a sensitivity analysis is always required. Further drawbacks of using a Bayesian approach include the possible computational complexity of such models. Also, there is no direct measure of statistical significance analogous to the p-value in frequentist inference (Sutton and Abrams, 2001).

The theoretical basis of commonly used fixed- and random-effects models using frequentist and Bayesian approaches is discussed below. Note that the models given are for the odds ratio measure of treatment effect but they can be modified to obtain other measures, for example risk difference or risk ratio. Models are illustrated for the outcome of death but translate directly to other

outcomes as required.

1.4.1 Peto meta-analysis

Yusuf et al. (1985) describe a frequentist fixed-effect approach to meta-analysis, the Peto method, similar to that proposed by Mantel and Haenszel (1959) for "2×2" tables.

Suppose there is a trial i with N_i patients, n_i of whom were treated and d_i of whom died. The observed number of patients who died in the treated group, O_i, are compared with the corresponding expected number E_i, where $E_i = n_i d_i / N_i$.

If there was no treatment effect then $O_i - E_i$ would only differ randomly from zero, whereas if treatment were beneficial then $O_i - E_i$ would tend to be negative, with variance V_i, where

$$V_i = \frac{E_i \left(1 - \dfrac{n_i}{N_i}\right)(N_i - d_i)}{(N_i - 1)},$$

and hence with standard error $\sqrt{V_i}$.

Now suppose there are m independent trials, the grand total (GT) would be given by

$$GT = \sum_{i=1}^{m} O_i - E_i. \tag{1.1}$$

If there was no treatment effect then this would also only differ randomly from zero, with variance equal to the sum of the individual variances (SIV)

$$SIV = \sum_{i=1}^{m} V_i.$$

Equation 1.1 assumes that all data from all randomised trials are available, without bias due to unavailability of unpromising results or to patient withdrawal. It does not assume patients across different trials can be compared directly, nor does it implicitly assume homogeneity of treatment effects in different trials. If, however, treatment effects are approximately similar then an estimate of the odds ratio (OR) for death among treated patients versus controls is given by

$$OR = \exp\left(\frac{GT}{SIV}\right),$$

with approximate 95% confidence interval (CI)

$$\exp\left(\frac{GT}{SIV} \pm \frac{1.96}{\sqrt{SIV}}\right).$$

1.4.2 DerSimonian and Laird meta-analysis

DerSimonian and Laird (1986) describe a frequentist random-effects method of meta-analysis based on the inverse-variance method where the weight given to each trial is chosen to be the inverse of the variance of the effect estimate.

As in the Peto method suppose there are m independent trials denoted $i = 1, ..., m$. Consider data from each trial consisting of the number of patients in the treatment ($n_{\mathrm{T}i}$) and control ($n_{\mathrm{C}i}$) groups, and the number of patients with some event in each of the groups, $r_{\mathrm{T}i}$ and $r_{\mathrm{C}i}$. Therefore, for a particular trial, i, the log odds ratio of the treatment effect estimate is given by

$$\delta_i = \ln\left(\frac{q_{\mathrm{T}i}(1 - q_{\mathrm{C}i})}{q_{\mathrm{C}i}(1 - q_{\mathrm{T}i})}\right),$$

where $q_{\mathrm{T}i} = r_{\mathrm{T}i}/n_{\mathrm{T}i}$ and $q_{\mathrm{C}i} = r_{\mathrm{C}i}/n_{\mathrm{C}i}$ are the proportions of events in treatment and control groups, respectively.

In the random-effects model, the individual trial treatment effect estimate, δ_i, can be defined as

$$\delta_i = \theta_i + \varepsilon_i,$$

where $\varepsilon_i \sim \mathrm{Normal}(0, \sigma_i^2)$ represents the sampling error of the ith study, and θ_i is the true treatment effect in the ith study given by

$$\theta_i = d + \upsilon_i.$$

It is assumed that $\upsilon_i \sim \mathrm{Normal}(0, \tau^2)$ is the deviation of the ith study's treatment effect from the mean log odds ratio of treatment, d, where τ^2 is a measure of the heterogeneity between studies. It is also assumed that ε_i and υ_i are independent. Therefore, δ_i is also normally distributed with mean d and variance $\upsilon_i + \varepsilon_i$.

The weighted estimator of the overall treatment effect for m independent trials is calculated as

$$\hat{d} = \frac{\sum_{i=1}^{m} w_i \delta_i}{\sum_{i=1}^{m} w_i},$$

where w_i is the inverse of the sampling variance of the ith trial.

Taking the exponent gives an estimate of the overall odds ratio of death among treated patients to that among controls, namely

$$OR = \exp(\hat{d}),$$

with approximate 95% confidence interval, as given in Sidik and Jonkman (2002),

$$\exp\left(\hat{d} \pm \frac{1.96}{\sqrt{\sum_{i=1}^{m} w_i}}\right).$$

1.4.3 Bayesian random-effect meta-analysis

Sutton et al. (2008) describe a model for Bayesian random-effects meta-analysis, combining individual patient data and aggregate data (grouped data) for individually allocated and cluster allocated studies. Part 3 of this model (individually allocated aggregate data studies) will be described in detail in this section.

As with the DerSimonian and Laird model described in Section 1.4.2 suppose there are m trials with data of the form n_{Ci}, n_{Ti}, r_{Ci} and r_{Ti}. It is assumed that

$$r_{Ci} \sim \text{Binomial}(p_{Ci}, n_{Ci}),$$
$$r_{Ti} \sim \text{Binomial}(p_{Ti}, n_{Ti}),$$

where p_{Ci} and p_{Ti} are the associated probabilities of event in the control and treatment groups, respectively.

The associated probabilities of event for the control group are modelled as

$$\text{logit}(p_{Ci}) = \mu_i, \tag{1.2}$$

where μ_i are the control group event rates for each study expressed on the logit scale. The control group event rates are each given vague prior distributions,

given here as a diffuse Normal distribution with mean 0 and variance 10^6,

$$\mu_i \sim \text{Normal}(0, 10^6).$$

The vague prior distributions may use different values to those quoted in this Chapter. However, all must have reasonably large values for the variance. For example, a variance of 10^3 is used in Chapters 7 and 8. There are no specific guidelines on which values can be considered reasonably diffuse.

For the treatment group the model for the associated probabilities of event is given by

$$\text{logit}(p_{\text{T}i}) = \mu_i + \delta_i, \qquad (1.3)$$

where δ_i are the treatment effects for each trial on the log-odds ratio scale and are assumed to be interchangeable across all studies. The treatment effects are Normally distributed with mean d and variance τ^2,

$$\delta_i \sim \text{Normal}(d, \tau^2).$$

The parameter d is given a vague prior distribution,

$$d \sim \text{Normal}(0, 10^6),$$

and a sensitivity analysis is performed on the prior required for τ^2. Based on previous research (Lambert et al., 2005), three different priors are considered. The Inverse-Gamma distribution is the most commonly used prior distribution for variance parameters. It has an approximately uniform distribution but with

a 'spike' of probability mass close to zero

$$\frac{1}{\tau^2} \sim \text{Gamma}(0.001, 0.001).$$

The Uniform prior distribution on the standard deviation is recommended by Spiegelhalter et al. (2004),

$$\tau \sim \text{Uniform}(0, 100).$$

Lastly, the Half-Normal prior distribution, truncated at zero, placed on the standard deviation has been used in previous meta-analysis applications

$$\tau \sim \text{Normal}(0, 100) \text{ for } \tau > 0.$$

The overall odds ratio for treatment effect is estimated by taking the exponent of the mean of the treatment effects

$$OR = \exp(d),$$

with 95% credible interval dictated by the posterior standard deviation of d.

1.5 Data sources and software

Several sets of data from different sources were obtained to answer the research questions. Four main sets of data were used, which will briefly be described here along with the software used for the analyses.

1.5.1 Glasgow Royal Infirmary cohorts

Data were obtained from two cohort studies. The first cohort (Barber et al., 2004) consisted of 873 consecutive patients admitted to the Glasgow Royal Infirmary over a two year period (2000-2002). In addition to baseline assessments of stroke characteristics and stroke severity, early neurological deterioration was assessed at day three and functional outcomes (death, dependency and requirement for institutional care) at 1 month following stroke.

The second (Sellars et al., 2007) was a prospective cohort of consecutive admissions to the Glasgow Royal Infirmary during a 17 month period from June 2004 to November 2005. Patients included were those with first or recurrent ischemic or haemorrhagic stroke within 7 days of admission to hospital. In total 412 patients were recruited into the study and were followed up at three months after stroke. As with the first cohort, in addition to baseline characteristics, the OCSP clinical classification, modified Rankin Scale and Barthel Activities of Daily Living Index were also recorded.

These datasets are used in Chapter 3 to investigate how much information is lost by categorising severity and whether the currently used categories are equivalent across different stroke scales.

1.5.2 Intravenous Magnesium Efficacy in Stroke trial

The Intravenous Magnesium Efficacy in Stroke (IMAGES) trial was an international, multicentre, double-blind, placebo-controlled, parallel group study which enrolled and randomised patients between October 1997 and April 2003 (Intravenous Magnesium Efficacy in Stroke (IMAGES) Study Investigators, 2004).

Randomised participants received either intravenous magnesium sulphate (1188 participants) or matching placebo (1198 participants). Outcomes included death and disability at 90 days (defined as Barthel score less than 95 or modified Rankin scale greater than 1). Baseline characteristics and OCSP classifications were also recorded during this trial.

This dataset is used in Chapter 2 to evaluate the internal consistency and validity of the IMAGES stroke scale.

1.5.3 Acute Stroke Unit of the Western Infirmary

The Acute Stroke Unit (ASU) of the Western Infirmary in Glasgow serves a catchment population of 220,000. Unselected patients who present within 72 hours of onset of acute neurological deficit with no known alternative to a vascular cause are admitted irrespective of age or severity of the neurological deficit. Approximately 800 patients are admitted each year. Patients have their stroke subtype categorised using the OCSP clinical classification, and brain imaging using either CT or MRI is performed promptly after admission (in most cases within 24 hours) on all patients with a clinical diagnosis of stroke. Clinical, radiological and biochemical data from each patient are reviewed and verified by a neurologist, radiologist and stroke physicians before being recorded prospectively on a computer database. Data extracted from this database included NIHSS assessments (including individual NIHSS items), database entries corresponding to the components of the IMAGES Stroke Scale, ISS, (described in more detail in Chapter 2), clinical classification (incorporating lateralisation, OCSP classification and final diagnosis), CT findings (to differentiate between ischaemic and haemorrhagic

stroke), age, and sex of each patient. Routine prospective recording of NIHSS assessment was available from 1998 until 2003. Hence NIHSS scores were available for a subset of ASU database patients.

This dataset is used in Chapter 2 to evaluate the internal consistency and validity of the IMAGES stroke scale.

1.5.4 Stroke Unit Trialists' Collaboration database

The Stroke Unit Trialists' Collaboration (SUTC) database consists of 31 prospective trials (Stroke Unit Trialists' Collaboration, 2007). Each used some form of random allocation of stroke patients to an organised system of inpatient (stroke unit) care or an alternative form of inpatient care, generally contemporary conventional care, but also including an alternative model of organised inpatient care. Trials were included if treatment allocation was carried out on a strictly random basis or with a quasi-random procedure such as bed availability or date of admission.

Eligibility and methodological quality of published trials were initially assessed by two review authors. Characteristics of unpublished trials were established through discussion with the trial co-ordinator. Descriptive information about the service characteristics of the treatment groups were obtained through a structured interview or correspondence with the trial co-ordinators. Trials were then allocated the service subgroups described in Section 1.2.1. Trialists were asked to provide information on the number of patients who were dead, requiring institutional care, dependent and missing at the end of scheduled follow up. The following subgroup information was also requested:

1. age: up to 75 years or greater than 75 years;

2. sex: male or female;

3. stroke severity: dependency at the time of randomisation (usually within 1 week of stroke):

   a) mild stroke: equivalent to a Barthel Index of between 10 and 20 during the first week;

   b) moderate stroke: equivalent to a Barthel Index from 3 to 9 during the first week;

   c) severe stroke: equivalent to a Barthel Index from 0 to 2 during the first week.

A more detailed description of the trials included in this database will be provided in Chapter 4.

This database is used in Chapters 4, 5, 6 and 8 to investigate which types of stroke unit are most effective and possible reasons why stroke units are effective.

Additionally, in Chapters 6 and 8, data from the Gateshead acute monitoring trial (Davis et al., 2000) was used alongside the SUTC database. The Gateshead trial randomised a total of 258 patients to receive "augmented care" (72 hours of continuous physiological monitoring) or "standard care" (routine 4 hourly observations) to test if intense monitoring could improve the detection of physiological complications which exacerbate neuronal injury thereby improving outcome. This trial is currently unpublished.

1.5.5 Software

Analyses were performed in Windows XP Professional using various statistical packages including: the SAS system 8.2 (SAS, Cary, NC); SPSS 15.0 for Windows (SPSS Inc, Chicago, IL); S-Plus 7.0 for Windows (Insightful Corporation, Seattle, WA); WinBUGS version 1.4.1 (Lunn et al., 2000); Review Manager (RevMan) 4.2 for Windows (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2003); and Minitab 15 (Minitab Incorporation, State College, PA).

Chapter 2

Consistency and validity of the IMAGES Stroke Scale

This Chapter introduces and assesses a novel stroke severity scale, the IMAGES Stroke Scale (ISS). Comparisons are made between this scale and current measures of stroke severity to determine whether the ISS offers a suitable alternative of measuring severity that requires no specific training and can be easily derived from routinely collected data. If ISS is found to be interchangeable with current stroke severity scales, this would allow patient severity to be measured using the same scale so that all patient severities are comparable across trials.

2.1 Background

Neurological impairment scales numerically grade deficits on neurological examination (Lyden and Lau, 1991) and have prognostic value in determining death

or dependency of acute stroke patients. Most widely used is the National Institutes of Health stroke scale (NIHSS), a 15-item scale measuring key components of a standard neurological examination (Brott et al., 1989) as described in Section 1.3.2. However, the NIHSS is recognised to be complex and to contain items with poor inter-observer reliability. Simpler modifications have been proposed (Lyden et al., 2001) but not widely adopted. Retrospective estimation of NIHSS from documentation of neurological deficits is possible but requires detailed medical records to be available (Kasner, 2006). Alternatively, the Oxfordshire Community Stroke Project classification (OCSP) assigns patients one of four clinical subtypes according to presenting neurological symptoms and signs (Bamford et al., 1991) and is also prognostically valuable, although not developed for this purpose.

The Intravenous Magnesium Efficacy in Stroke (IMAGES) trial was an international, multi-centre, double-blind, placebo-controlled, parallel group study investigating whether intravenous magnesium sulphate, given within 12 hours of stroke onset, reduced death or disability at 90 days (Intravenous Magnesium Efficacy in Stroke (IMAGES) Study Investigators, 2004). A secondary analysis formulated the IMAGES Stroke Scale (ISS), which incorporates elements used to derive OCSP classification (Bamford et al., 1991), and the Glasgow Coma Scale (GCS) (Jennett and Teasdale, 1977).

In this Chapter the novel IMAGES Stroke Scale is investigated. The internal consistency of the ISS and NIHSS and the construct validity of the ISS in predicting NIHSS are assessed. The ability of the ISS to predict death or dependency 90 days after stroke is investigated along with whether it provides additional prognostic information to a simple clinical classification such as the OCSP and how

its predictive accuracy compares to that of the NIHSS.

2.2   Methods

The ISS is created by combining components of the OCSP, as described in Section 1.3.1, and the Glasgow Coma Scale (Jennett and Teasdale, 1977). The Glasgow Coma Scale (Table 2.1) is an impairment scale which measures impairment after head trauma, for example, head injuries or non-traumatic conditions such as stroke. The scale assesses eye movement, motor and verbal response where high and low numbers represent normal and impaired responses, respectively. The total score, the EMV Score, is the sum of the individual responses giving a maximum value of 15 (normal) and minimum value of 3 (maximum impairment).

Table 2.1: The Glasgow Coma (or Responsiveness) Scale.

Observation	Response
Eye opening (E)	4 = Spontaneous
	3 = To speech
	2 = To pain
	1 = Nil
Best motor response (M)	6 = Obeys
	5 = Localises
	4 = Withdraws
	3 = Abnormal flexion
	2 = Extends
	1 = Nil
Verbal response (V)	5 = Orientated
	4 = Confused conversation
	3 = Inappropriate words
	2 = Incomprehensible sounds
	1 = Nil

Table 2.2: The IMAGES Stroke Scale.

Item	Score
Unilateral weakness of:	
face	1 point
arm	2 points
leg	2 points
Sensory loss	1 point
Dysphasia	2 points
Neglect	1 point
Hemianopia	3 points
Brainstem signs	1 point
Any abnormality in the GSC components of:	
eye	1 point
motor	1 point
verbal	1 point
Total	16 points

The ISS (Table 2.2) consists of the presence or absence of: unilateral face weakness (1 point), unilateral arm weakness (2 points), unilateral leg weakness (2 points), unilateral sensory loss (1 point), dysphasia (2 points), neglect (1 point), hemianopia (3 points), brainstem signs (1 point), any abnormality in the GCS eye component (1 point), motor component (1 point) and verbal component (1 point), yielding a maximum deficit of 16 points. The score assigned to each item was loosely based on its weighting in the NIHSS stroke scale (Table 1.2).

To evaluate the internal consistency and construct validity of the ISS and whether the ISS and NIHSS were interchangeable measures of impairment, data from an Acute Stroke Unit (ASU), described in Section 1.5.3, were used. The ASU admits patients within 72 hours of stroke onset irrespective of age or severity.

Prospectively documented brain imaging findings and OCSP classification were available for all patients, while NIHSS score was available for a subset of patients and the ISS score was derived retrospectively from the database. Since NIHSS was only available for a subset of patients, the dependence of recording of NIHSS on patient characteristics of age and severity (measured using ISS) was examined.

The internal consistency of the ISS and NIHSS was measured using Cronbach's alpha correlation coefficient (Armitage and Colton, 1998), which tests whether the items within the scale measure a single construct. Subjective guidelines for interpretation of Cronbach's alpha are: < 0.60 unacceptable, $0.61 - 0.65$ undesirable, $0.66 - 0.70$ minimally acceptable, $0.71 - 0.80$ respectable, $0.81 - 0.90$ very good, and > 0.90 some items possibly redundant (Armitage and Colton, 1998). Construct validity, which refers to the extent to which ISS can be related to other constructs, was measured by analysing the ISS relationship with NIHSS using the Spearman correlation coefficient. Bland-Altman plots (Bland and Altman, 1999) were created to assess whether the ISS and NIHSS are interchangeable methods of assessing stroke. ISS and NIHSS were scaled as a percentage of the maximum deficits since the ranges of the scales are different. This limits interpretation of systematic bias. However, the variability in differences in percentages will give some information about whether the scales are interchangeable.

Internal consistency, construct validity and Bland-Altman plots were analysed for all data available in the ASU, as well as within the OCSP clinical subgroups: total anterior, partial anterior and posterior circulation syndromes (TACS, PACS and POCS), and lacunar syndrome (LACS). Additionally, internal consistency of NIHSS was measured when its items were dichotomised as present or absent to determine if the dichotomisation of scale components leads to poorer internal

consistency.

The prognostic value of the ISS was analysed using the 1198 patients ran-
domised to the control group in the original IMAGES trial described in Sec-
tion 1.5.2. The primary outcome was the combined outcome of death or depen-
dency defined as Barthel Index (BI) less than 95 and the secondary outcome of
death or dependency defined as modified Rankin Score greater than or equal to
2. Outcomes were analysed using logistic regression and effect sizes expressed
in terms of odds ratios. Although this contrasts with Chapters 3 and 4 where
dependency is defined as mRS greater than 2, cut-offs are subjective and may be
defined differently depending on the context. Significance of effects was tested us-
ing generalised likelihood ratio tests. ROC curves (Armitage and Colton, 1998)
were used to assess the usefulness of ISS in predicting outcome of stroke and
compared to previous ROC curves calculated for NIHSS (Muir et al., 1996).

2.3   Results

2.3.1   Internal consistency and construct validity

The 8284 patients admitted between 1990 and 2005 in the ASU database of the
Western Infirmary in Glasgow had ISS recorded retrospectively. Of these, 2111
(25.5%) had both ISS and NIHSS recorded, 6851 (82.7%) had both ISS and OCSP
classification recorded, and 1777 (21.5%) patients had all three stroke scales (ISS,
NIHSS and OCSP) recorded. Figure 2.1 shows that the recording of NIHSS did
not depend on the patient characteristics of age and severity.

The standardised Cronbach's alpha for ISS, based on 8284 patients, was 0.48,

Figure 2.1: Boxplots showing distribution of (a) age and (b) ISS total for those patients with and without NIHSS recorded.

indicating 'unacceptable' internal consistency. This is compared with the standardised Cronbach's alpha for NIHSS, based on 2111 patients, of 0.85, indicating 'very good' internal consistency. Internal consistency for ISS based on the NIHSS subset was similar (0.49). Table 2.3 gives the standardised internal consistencies

Table 2.3: Internal consistency and Spearman rank correlation of ISS and NIHSS by OCSP classification.

OCSP	ISS observations (%)	Cronbach's Alpha ISS	NIHSS observations (%)	Cronbach's Alpha NIHSS	Spearman rank correlation (95% CI)
LACS	2223 (32%)	0.41	627 (35%)	0.76	0.38 (0.31, 0.45)
PACS	2377 (35%)	0.46	600 (34%)	0.78	0.50 (0.44, 0.56)
POCS	735 (11%)	0.46	179 (10%)	0.88	0.40 (0.27, 0.51)
TACS	1516 (22%)	0.24	371 (21%)	0.67	0.21 (0.11, 0.31)
Total	6851	0.49	1777	0.85	0.63 (0.60, 0.65)*

*Total correlation calculated using 1777 patients who had ISS, NIHSS and OCSP classifications recorded. Total correlation for patients with only ISS and NIHSS recorded is given as 0.61 (0.59, 0.64).

Table 2.4: Internal consistency NIHSS by OCSP classification when the NIHSS items are dichotomised.

OCSP	Observations (%)	Cronbach's Alpha
LACS	627 (35.3)	0.67
PACS	600 (34.8)	0.74
POCS	179 (10.1)	0.80
TACS	371 (20.9)	0.59
Total	1777	0.80*

*Cronbach's alpha based on 1777 patients who had ISS, NIHSS and OCSP classifications recorded. Cronbach's alpha for all patient with NIHSS recorded is 0.80

within each OCSP category which were 'unacceptable' for ISS and at least 'minimally acceptable' for NIHSS. The standardised internal consistency of NIHSS appears to reduce when the items are dichotomised, as shown in Table 2.4

The Spearman correlation between ISS and NIHSS, based on 2111 patients, showed a modest and significantly positive correlation (0.61; 95% CI: 0.59, 0.64) between ISS and NIHSS (Figure 2.2). Table 2.3 shows significant but weaker positive correlations between ISS and NIHSS within each OCSP subgroup. Figure 2.2 shows for a given value of ISS, there is large variability in the equivalent value of NIHSS.

The Bland-Altman plot for the complete data, Figure 2.3(a), shows that the difference between the scaled measures for small averaged measures (0-20) increases as the average increases. The same pattern emerges within the OCSP classifications of PACS, POCS and LACS patients. This non-uniform proportional difference in the Bland-Altman plot could be due to a flooring effect of
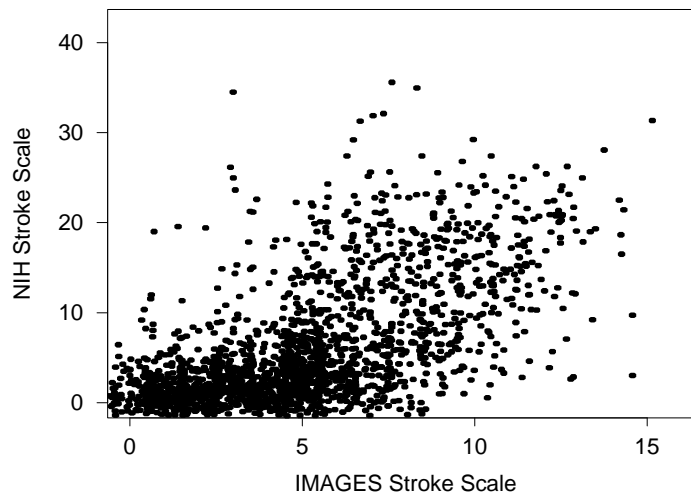
Figure 2.2: Scatterplot of ISS and NIHSS.

the scales. The majority of patients have small averaged scores but patients cannot score less than zero, meaning that the difference between scaled measures can take a wider range of values as the averaged score increases. This makes interpretation difficult as the underlying assumptions for interpretation are uniform bias and variability about the mean. However, for TACS patients, whose strokes are generally more severe (and therefore the measures obtained for ISS and NIHSS are larger), Figure 2.3(b) shows that the points are randomly scattered within the plot, making underlying assumptions valid.

2.3.2   Prediction of functional outcome

Of the 1198 patients randomised into the control group, 787 (66%) were recorded as having a Barthel Index of less than 95 at 90 days following stroke and 858 (72%) had a modified Rankin Scale greater than or equal to 2. Table 2.5 shows
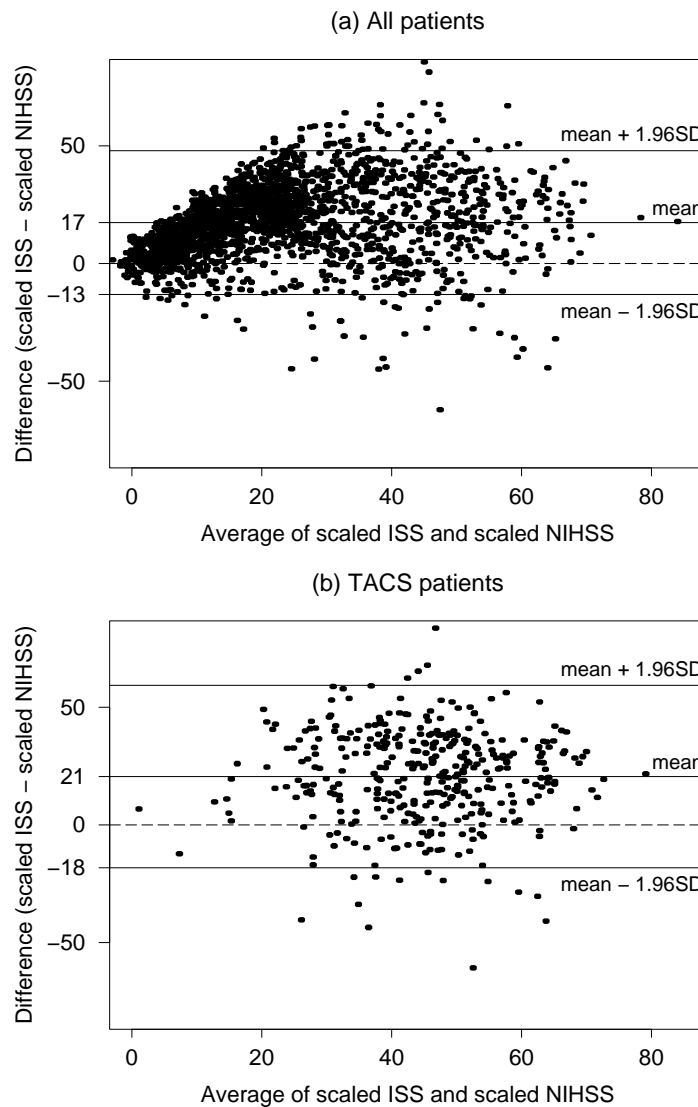
Figure 2.3: Bland-Altman plot of ISS and NIHSS for (a) all patients and (b) TACS patients. ISS and NIHSS are scaled as a percentage of maximum due to different ranges of each scales. Re-scaling allows comparability of the scales on the Bland-Altman plot.

the proportions of patients with good (BI $\geq$ 95 or mRS $<$ 2) and poor outcome (BI $<$ 95 or mRS $\geq$ 2) in each of the OCSP classifications.

Figure 2.4 shows that those with poor outcomes have slightly larger ISS scores

Table 2.5: Percentage of patients with good (BI $\geq$ 95 or mRS $<$ 2) and poor (BI $<$ 95 or mRS $\geq$ 2) outcome in each OCSP classification.

OCSP	Barthel Index		mRankin Scale		Total
	BI $\geq$ 95	BI $<$ 95	mRS $<$ 2	mRS $\geq$ 2	
LACS	178 (47%)	204 (53%)	153 (40%)	229 (60%)	382
PACS	175 (40%)	266 (60%)	141 (32%)	300 (68%)	441
POCS	11 (42%)	15 (58%)	8 (31%)	18 (69%)	26
TACS	47 (13%)	302 (87%)	38 (11%)	311 (89%)	349
Total	411 (34%)	787 (66%)	340 (28%)	858 (72%)	1198

than those with good outcomes. Median ISS scores for TACS patients with poor outcomes appears to be greater than the median scores for other levels of OCSP, with LACS patients having the lowest ISS score.

Table 2.6 describes the models fitted for each outcome. OCSP and ISS were each independently and significantly associated with both outcomes (model 1: $p < 0.001$; and model 2: $p < 0.001$). When ISS alone is included in the model, for each point increment of ISS the odds of BI < 95 increased by a factor of 1.31 (95% CI 1.25, 1.37) and the odds of mRS ≥ 2 by 1.31 (95% CI 1.25, 1.38). ISS added significant prognostic information when added to OCSP for both outcomes (model 3: $p < 0.001$). However, if ISS (which is constructed from OCSP inter alia) is known, OCSP does not add any significant prognostic information (model 4: BI: $p = 0.20$; mRS: $p = 0.57$).

For both functional outcomes, the area under the ROC curve for ISS (Figure 2.5) was 0.71 (95% CI: 0.67, 0.74). Patients scoring 8 or greater could be predicted as being dead or dependent at 90 days since this cut-point best maximises sensitivity and specificity. The sensitivity and specificity at the cut-point

(a) Barthel Index

OCSP

(b) modified Rankin Scale
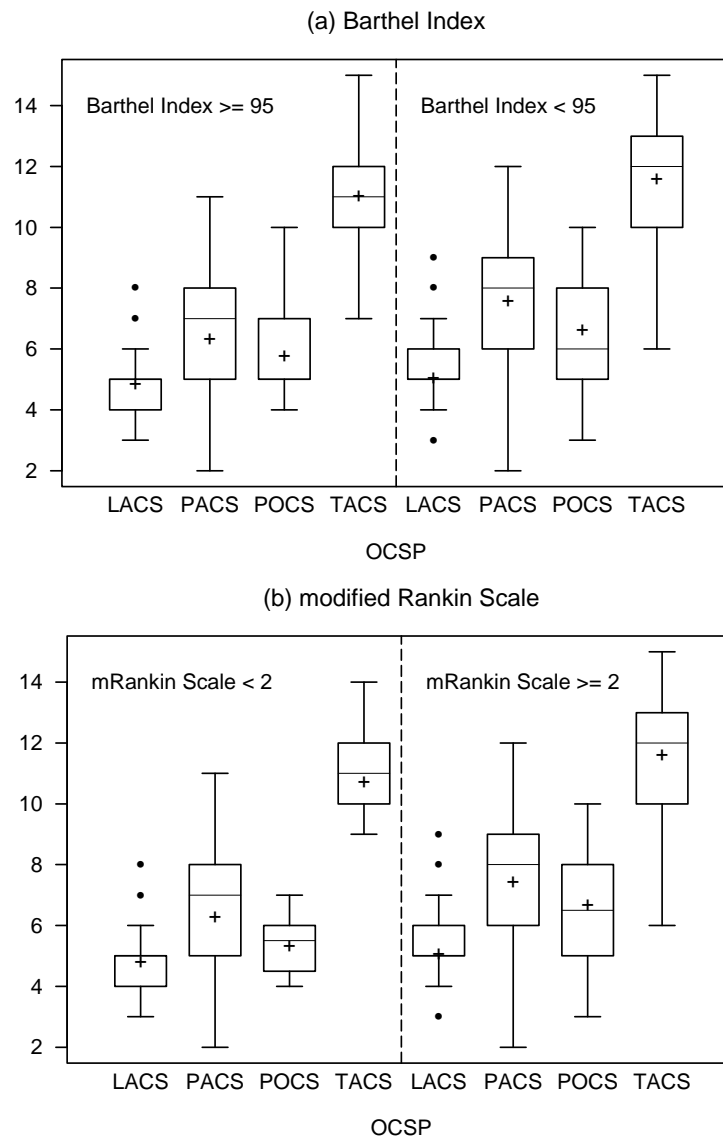
OCSP

Figure 2.4: Boxplot of IMAGES Stroke Scale by OCSP for outcomes (a) Barthel Index < 95 and (b) modified Rankin Scale ≥ 2. Solid circles represent possible outliers, while $+$'s represent means.

of 8 for BI < 95 were 0.58 (0.54, 0.61) and 0.75 (0.71, 0.79), respectively, indicating that approximately 58% of events (poor outcomes) were classified correctly and 75% of non-events (good outcomes) were classified correctly. For mRS ≥ 2

Table 2.6: Summary of models fitted for each outcome and p-values for the effects of each term.

Model	BI< 95 p-value	mRS≥ 2 p-value
1. OCSP	<0.001	<0.001
2. ISS	<0.001	<0.001
3. ISS (given OCSP)	<0.001	<0.001
4. OCSP (given ISS)	0.20	0.57

the sensitivity and specificity at the cut-point of 8 were 0.56 (0.52, 0.59) and 0.77 (0.72, 0.81), respectively. In a previous analysis of ROC curves for NIHSS (Muir et al., 1996), which subjectively used an NIHSS of 7 as the best possible cut-point for maximising sensitivity and specificity, the sensitivity and specificity were approximately 0.85 and 0.70, respectively.
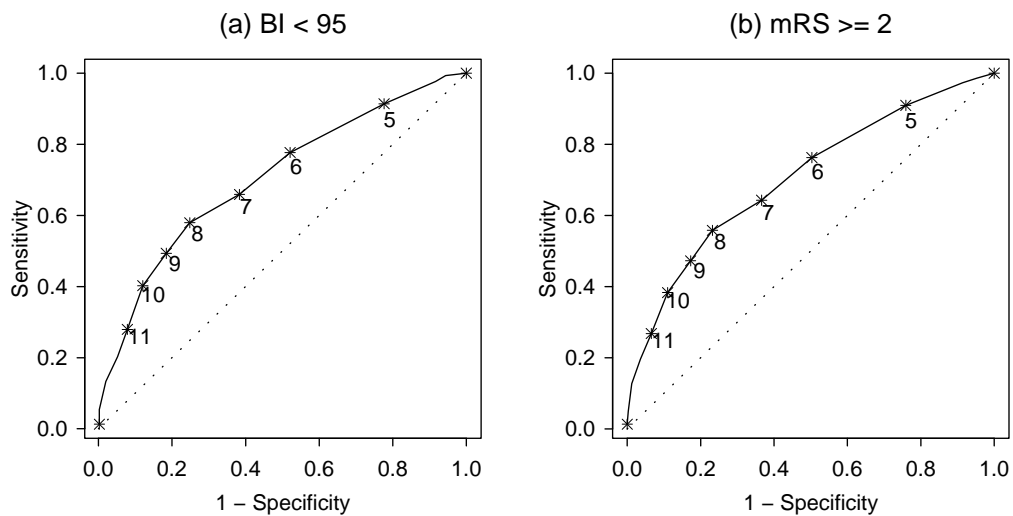
Figure 2.5: ROC curves for (a) Barthel Index < 95 and (b) modified Rankin Scale ≥ 2.

2.4   Discussion

The IMAGES stroke scale can be derived from data documented routinely in most stroke patients. In contrast, retrospective estimation of NIHSS requires detailed neurological examination findings to be recorded (Kasner, 2006). However, according to established guidelines for the interpretation of internal consistency (Armitage and Colton, 1998), the performance of ISS was found to be 'unacceptable' while that of NIHSS was 'very good'. The binary coding of the ISS components (present or absent), compared with the more finely graded NIHSS items, may partially explain the low internal consistency of ISS. The internal consistency of NIHSS was reduced when the items of NIHSS were dichotomised, but were still not as low as the ISS score. Additionally, the ISS measures a range of aspects of stroke severity, clinical symptoms and signs, and measures of consciousness giving a further explanation for the low internal consistency and also an explanation for the modest correlation between ISS and NIHSS, shown in Figure 2.2.

Reliability (inter-observer, intraobserver, or test-retest agreement) was not assessable based on the data available. Criterion validity was not assessed as there is no 'gold-standard' scale to measure dependence after stroke (Kasner, 2006). Additionally, content validity was not assessed as this is based on expert opinion and is not a quantitative measure.

The large variability in the equivalent value of NIHSS for a given value of ISS (Figure 2.2) shows that the scales may not be interchangeable, which is also reflected in the Bland-Altman plots. The large variability in the proportional difference between ISS and NIHSS in the Bland-Altman plots suggests that the

scales may not be interchangeable since, for a given scaled value of NIHSS, the corresponding scaled ISS value could proportionally score between 18% lower and up to 59% higher than the scaled NIHSS. This lack of interchangeability may be explained by ISS and NIHSS measuring different constructs.

Both OCSP and ISS have significant association with the outcome of stroke (BI < 95 and mRS ≥ 2). If ISS is known, OCSP does not add any significant additional prognostic information, but if OCSP is known then ISS adds significant additional prognostic information. Finally, by analysing the ROC curves and comparing this to previous work (Muir et al., 1996) it is shown that when identifying the cut-point on each of ISS and NIHSS that maximises sensitivity and specificity, the NIHSS allows a better prediction of outcome than ISS.

A large data set from a broad case mix of patients was used to analyse the consistency and validity of the ISS. This has an advantage over highly selected trial patients as conclusions are not restricted to a specific subset of stroke patients. However, not all patients had OCSP and NIHSS recorded in addition to ISS, reducing the sample size available for analysis. Although this limitation could be a possible source of bias, an examination of the data showed that the recording of NIHSS appeared to be independent of patient characteristics.

The original IMAGES trial data were rigorously collected meaning that there were few missing data, allowing the calculation of the prognostic value of ISS compared to the OCSP. The most obvious limitation to the study of the prognostic value of ISS and NIHSS is that there is not a direct comparison of functional disability outcome for both of these scales. ROC curves from previously published work were used to estimate the sensitivity and specificity of NIHSS.

Its wide familiarity and standardised video training make the NIHSS the

preferred acute stroke scale. However, in large simple trials, especially those conducted in countries where there are language barriers and with limited access to training material, the ISS offers an alternative since it is easily derived from routinely available data without the need for specific training. However, further development to improve its internal consistency and prognostic value would be required to make it useful in practice.

Chapter 3

Categorising stroke severity using different stroke scales

In the previous Chapter, it was shown that the IMAGES Stroke Scale was not interchangeable with NIHSS and the NIHSS allows a better prediction of outcome compared to ISS. Therefore, since ISS cannot be used as a general method of measuring severity in stroke patients, it would beneficial to examine whether current methods of measuring stroke severity are equivalent in predicting outcomes of patients. Additionally, the categorisation of stroke severity is common in meta-analysis (see Chapters 6, 7 and 8) and it is of further interest to examine how much information is retained when stroke scales are categorised, and if these categorisations are also equivalent.

3.1   Background

Stroke is a very heterogeneous condition where patients can present with a range of severities, from those that resolve within a few hours, to those which are rapidly fatal. When measuring severity using scales such as those described in Section 1.3, further variation is introduced by the use of different scales, or the redundancy of items within a standard scale which masks the true severity of an individual patient.

Meta-analysis often tries to compare the efficacy of an intervention between stroke patients with different severities. However, the use of different methods of measuring stroke severity may complicate analyses. Furthermore, although individual patient data is considered the gold standard for meta-analyses, the collection of individual patient data in meta-analysis is difficult, time-consuming and sometimes not feasible. Therefore, severity data is often stratified into a small number of categories, making the collection of aggregate data over the small number of strata easier and meta-analyses simpler.

There is uncertainty as to whether the categorisations used for different stroke scales are equivalent. While discrepancies between severity classifications in the different stroke scales do not influence the treatment effect, they may influence the effect of interactions between severity and treatment. Therefore, the amount of information retained by categorising severity, and whether the currently used cut-offs are equivalent across different stroke scales, was investigated.

3.2 Methods

3.2.1 Data

In order to investigate severity and its categorisations, data from the Glasgow Royal Infirmary cohorts, as described in Section 1.5.1, were used. The first cohort (Barber et al., 2004) contains measurements of the Scandinavian stroke scale (SSS), modified Rankin Scale (mRS) and the Barthel Index (BI) as described in Sections 1.3.3, 1.3.5 and 1.3.4, respectively. In addition to the modified Rankin Scale and Barthel Index, the second cohort (Sellars et al., 2007) also recorded measurements of the National Institutes of Health stroke scale (NIHSS) described in Section 1.3.2. Table 3.1 shows how each of these stroke scales are used to categorise patients into mild, moderate and severe. For both cohorts the outcomes of death, dependency and institutional care were recorded at the end of follow up (one month in Barber et al. (2004) and three months in Sellars et al. (2007)), with dependency defined as mRS > 2.

Table 3.1: Mild, moderate and severe categories using baseline levels for four established stroke scales.

Scale	Mild	Moderate	Severe
Barthel Index	10 - 20	3 - 9	0 - 2
modified Rankin Scale	0 - 3	4	5
Scandinavian Stroke Scale	43 - 58	26 - 42	0 - 25
NIH Stroke Scale	0 - 5	6 - 14	15 - 42

Barthel Index, Scandinavian Stroke Scale and modified Rankin Scale recorded on day 3, National Institutes of Health Stroke Scale recorded on day 1-5.

3.2.2 Statistical methods

Several methods have been developed in order to choose the optimal cut-points and number of groupings for continuous variables. Taylor and Yu (2002) summarise the work of a number of authors in this area. Cox (1957) discusses methods of grouping variates so that the intervals retain as much information as possible. Connor (1972) considered the choice of optimal cut-points for a continuous variable and found relatively little loss in efficacy with as few as three or four optimal intervals. Altman et al. (1994) describe a data-dependent approach to defining "low" and "high" risk groups using a minimum p-value approach. Cut-points are varied systematically when categorising a continuous variable and a p-value is computed for each cut-point. The cut-points are then chosen to correspond to the most significant relationship with outcome. The authors warn that naive use of this approach is associated with considerable inflation of false-positive error rates due to multiple testing. While a correction of the minimal p-value allows for multiple testing in large sample sizes, they also warn that the effect of the covariate may still be considerably overestimated. The authors recommend constructing three or more prespecified groups, and basing the choice of cut-point on simplicity, biological reasoning and knowledge of measurement techniques.

Stroke severity has generally been stratified into three categories: mild, moderate and severe. The categories have been chosen in a similar, though less formal, fashion to the minimum p-value approach described by Altman et al. (1994). Three categories were chosen for several reasons. Suppose a small trial was conducted which stratified its randomisation by severity plus another variable, for example, age. This would lead to at least six strata in total which, in a small

trial, may lead to strata containing few patients and hence randomised blocks
that are incomplete. This may lead to an imbalance of randomised treatments
across stratification variables, undermining the very aim of stratification. There-
fore, it would not be feasible to stratify severity by more than three categories.
However, dichotomising severity may lead to too few categories and possibly to
the loss of too much information.

Receiver operating characteristic (ROC) curves were used to assess the useful-
ness of each stroke scale in predicting outcome of stroke. In addition to assessing
the area under the ROC curve (AUC) of each scale, the AUC for each full scale
was compared to that of the categorised scale to determine how much infor-
mation is retained after categorising severity as mild, moderate or severe. The
differences in AUC were calculated using an algorithm developed by SAS Insti-
tute Inc. (2007). Subjective guidelines for interpretation of the AUC, based on
suggestions by Swets (1988) and Greiner et al. (2000) are: 0.50 non-informative;
$0.51-0.70$ low accuracy; $0.71-0.90$ moderate accuracy; $0.91-0.99$ high accuracy;
and 1.00 perfect accuracy.

The agreement between the categorisations of different scales was assessed.
In ordinal measures such as severity, certain disagreements are more serious than
others. For example, a severe patient classified as mild is a more serious disagree-
ment than a moderate patient classified as mild. Therefore, a weighted kappa
analysis (Cohen, 1968) was performed to account for the ordering of the cate-
gories using the Cicchetti-Allison kappa coefficient (Cicchetti and Allison, 1971)
where exact agreement is given the maximum weight of 1 and disagreements
in adjacent and disparate categories weighted as 0.5 and 0, respectively. Sub-
jective guidelines for the interpretation of the kappa statistic are: < 0.00 poor

agreement; $0.0 - 0.20$ slight agreement; $0.21 - 0.40$ fair agreement; $0.41 - 0.60$ moderate agreement; $0.61 - 0.80$ substantial agreement; and $0.81 - 1.00$ almost perfect agreement (Landis and Koch, 1977).

3.3 Results

Of the 733 acute stroke patients in the Barber et al. (2004) Glasgow Royal Infirmary cohort, 665 (91%) had modified Rankin Score, Barthel Index and Scandinavian Stroke scale recorded within three days of admission. The outcomes of death and death or dependency were available for all patients and death or institutional care was available for 572 patients (78%). Outcomes were recorded at one month following stroke.

In the Sellars et al. (2007) cohort, of the 412 acute stroke patients, 405 (98%) had modified Rankin Score, Barthel Index and NIH stroke scale recorded within three days of admission. Again the outcomes of death and death or dependency

Table 3.2: Percentage of patients in mild, moderate and severe categories for each stroke scale in each cohort.

Study/scale	Mild	Moderate	Severe	Total
Barber et al. (2004)				
SSS	409 (56%)	170 (23%)	147 (20%)	726
mRS	357 (54%)	193 (29%)	115 (17%)	665
BI	416 (57%)	157 (22%)	151 (21%)	724
Sellars et al. (2007)				
NIHSS	221 (54%)	125 (31%)	63 (15%)	409
mRS	222 (54%)	109 (26%)	81 (20%)	412
BI	232 (57%)	93 (23%)	83 (20%)	408

SSS represents Scandinavian Stroke Scale; mRS - modified Rankin Scale; NIHSS represents National Institutes of Health Stroke Scale; and BI - Barthel Index.

were available for all patients while death or institutional care was available for 406 patients (99%). Outcomes were recorded at three months following stroke.

Table 3.2 shows, for the Barber et al. (2004) and Sellars et al. (2007) cohorts, the number and percentage of patients in the mild, moderate and severe categories, as defined in Table 3.1. Table 3.2 shows that the percentages are approximately the same across each scale within cohorts as well as between cohorts.

3.3.1 Comparison of full and categorised stroke scales

The accuracy of predicting outcome can be determined by examining the area under the ROC curve. Tables 3.3 and 3.4 show the estimated areas with 95% confidence limits for the full scales, categorised scales and the differences between these for the Barber et al. (2004) and Sellars et al. (2007) cohorts, respectively. P-values are also given for the test of statistical significance of the differences.

All areas under the ROC curves for the Barber et al. (2004) cohort (Table 3.3) show moderate or high accuracy in predicting outcome for both the full and categorised scales. Although the p-values show statistically significant differences between the AUCs of the full and categorised scales (differences which appear to be larger for the death or dependency and death or institutional care outcomes), the difference estimates are small with narrow confidence intervals. The comparison of the full and categorised scales can be seen more clearly in the ROC curves in Figure 3.1, which demonstrate little difference between the full and categorised scales for the outcome of death while there is slightly lower predictive accuracy of the categorised scales for the outcomes of death or dependency and death or institutional care. Note also that the predictive accuracy for the full scales is also

Table 3.3: Estimates and 95% confidence limits for the area under the ROC curve of each stroke scale for the full scale, categorised scale and the difference between them (full−categorised) for the Barber et al. (2004) cohort.

Outcome	SSS	mRS	BI
Death			
Full	0.93 (0.90, 0.96)	0.92 (0.89, 0.95)	0.92 (0.90, 0.95)
Category	0.88 (0.85, 0.92)	0.92 (0.89, 0.95)	0.90 (0.88, 0.93)
Difference	0.044 (0.028, 0.059)	0.0004 (-0.001, 0.001)	0.019 (0.008, 0.029)
	$p < 0.001$	$p = 0.357$	$p < 0.001$
Death or dependency			
Full	0.86 (0.84, 0.89)	0.88 (0.86, 0.91)	0.89 (0.87, 0.91)
Category	0.79 (0.76, 0.82)	0.83 (0.81, 0.86)	0.81 (0.79, 0.84)
Difference	0.075 (0.059, 0.092)	0.049 (0.034, 0.065)	0.081 (0.063, 0.098)
	$p < 0.001$	$p < 0.001$	$p < 0.001$
Death or institutional care			
Full	0.92 (0.90, 0.94)	0.92 (0.90, 0.94)	0.93 (0.91, 0.95)
Category	0.86 (0.83, 0.89)	0.88 (0.85, 0.90)	0.86 (0.83, 0.89)
Difference	0.062 (0.045, 0.079)	0.042 (0.028, 0.057)	0.067 (0.049, 0.086)
	$p = 0.002$	$p < 0.001$	$p < 0.001$

Note: positive values indicate the area under the ROC curve is greater for the full scale, while negative values indicate the area under the ROC curve is greater for the categorised scale. SSS represents Scandinavian Stroke Scale; mRS - modified Rankin Scale; and BI - Barthel Index.

lower for these outcomes.

The areas under the curve for the Sellars et al. (2007) cohort (Table 3.4) also show moderate to high accuracy in predicting outcome. Again the *p*-values show statistically significant differences between the full and categorised scales although, as before, the estimates of the differences are small with narrow confidence intervals. Figure 3.2 shows the comparisons for the Sellars et al. (2007) cohort.

Notice also that the area under the ROC curves for mRS and BI in the Barber et al. (2004) cohort, where outcomes were recorded at one month, are larger than

Death
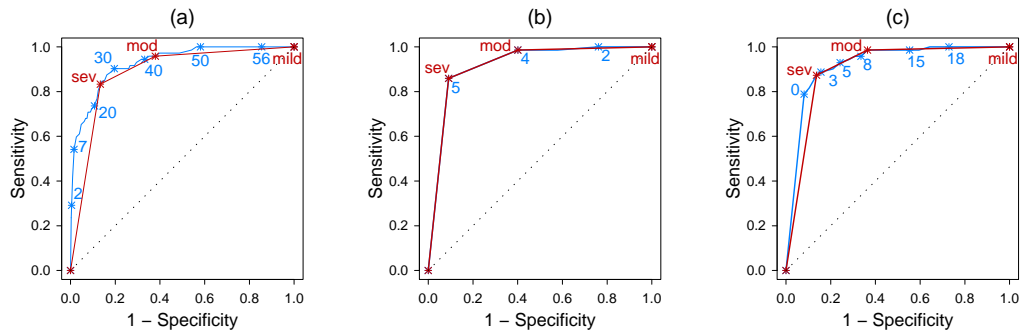
Death or dependency
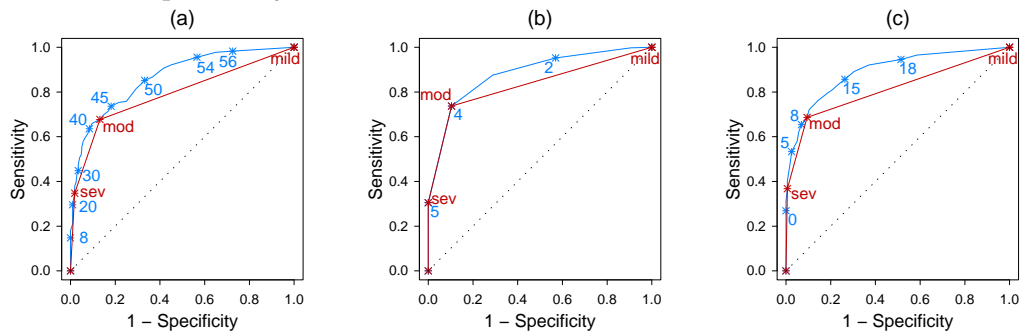
Death or institutional care
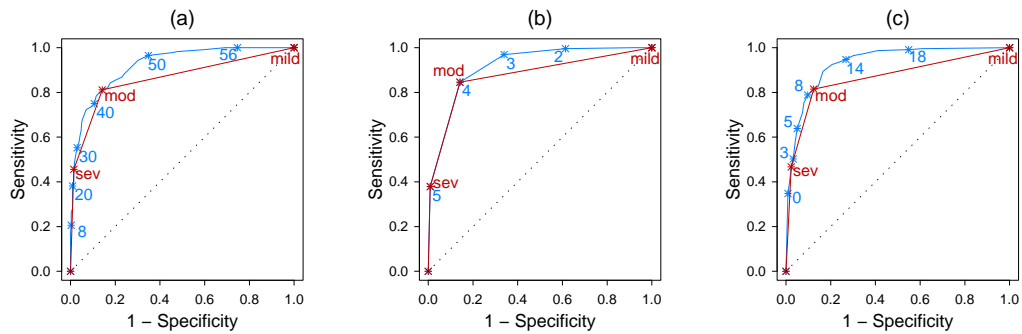
Figure 3.1: ROC curves of (a) Scandinavian Stroke Scale (SSS), (b) modified Rankin Scale (mRS) and (c) Barthel Index (BI) in the Barber et al. (2004) cohort for the outcomes of death, death or dependency and death or institutional care, where the blue line represents the full scale and the red line represents the categorised scale. Sensitivity and specificity are calculated at each level: less than or equal to the level for SSS and BI, and greater than or equal to the level for mRS. The dotted line represents the line of 'no information'.

Table 3.4: Estimates and 95% confidence limits for the area under the ROC curve of each stroke scale for the full scale, categorised scale and the difference between them (full−categorised) for the Sellars et al. (2007) cohort.

Outcome	NIHSS	mRS	BI
Death			
Full	0.84 (0.78, 0.89)	0.84 (0.79, 0.88)	0.87 (0.83, 0.92)
Category	0.81 (0.75, 0.86)	0.82 (0.77, 0.88)	0.84 (0.79, 0.89)
Difference	0.026 (0.006, 0.046)	0.015 (0.001, 0.029)	0.032 (0.014, 0.051)
	$p = 0.011$	$p = 0.031$	$p < 0.001$
Death or dependency			
Full	0.77 (0.73, 0.82)	0.80 (0.76, 0.84)	0.80 (0.75, 0.84)
Category	0.73 (0.68, 0.77)	0.75 (0.70, 0.79)	0.73 (0.69, 0.77)
Difference	0.044 (0.021, 0.067)	0.057 (0.042, 0.092)	0.067 (0.036, 0.078)
	$p < 0.001$	$p < 0.001$	$p < 0.001$
Death or institutional care			
Full	0.84 (0.80, 0.89)	0.87 (0.83, 0.90)	0.90 (0.86, 0.93)
Category	0.81 (0.76, 0.85)	0.84 (0.80, 0.89)	0.86 (0.82, 0.90)
Difference	0.034 (0.015, 0.052)	0.022 (0.008, 0.036)	0.034 (0.017, 0.052)
	$p = 0.002$	$p < 0.001$	$p < 0.001$

Note: positive values indicate the area under the ROC curve is greater for the full scale, while negative values indicate the area under the ROC curve is greater for the categorised scale. NIHSS represents National Institutes of Health Stroke Scale; mRS - modified Rankin Scale; and BI - Barthel Index.

the areas under the curves in the Sellars et al. (2007) cohort. This is to be expected as in the latter cohort, outcome is predicted over a longer time period (3 months versus 1 month) during which events unrelated to initial stroke severity are more likely to occur and influence outcome.

3.3.2 Comparison of stroke scales

The estimates, 95% confidence limits and p-values for the differences in area under the ROC curve between each scale for the Barber et al. (2004) and Sellars et al. (2007) cohort are given in Tables 3.5 and 3.6, respectively. Both tables give the

Death
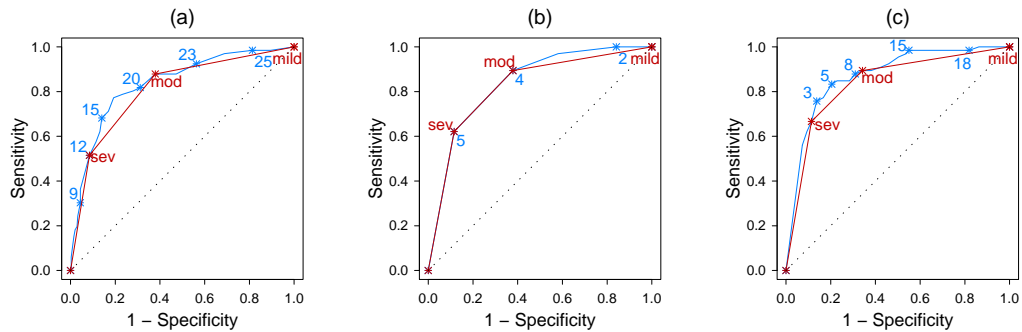
Death or dependency
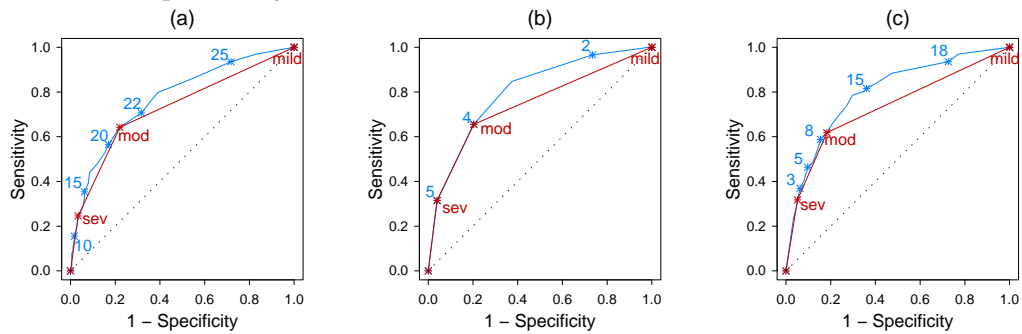
Death or institutional care
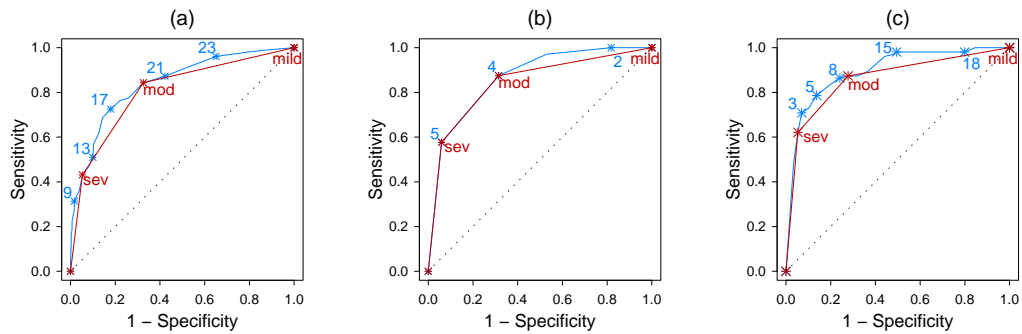
Figure 3.2: ROC curves of (a) National Institutes of Health Stroke Scale (NIHSS), (b) modified Rankin Scale (mRS) and (c) Barthel Index (BI) in the Sellars et al. (2007) cohort for the outcomes of death, death or dependency and death or institutional care, where the blue line represents the full scale and the red line represents the categorised scale. Sensitivity and specificity are calculated at each level: less than or equal to the level for BI and greater than or equal to the level for NIHSS and mRS. The dotted line represents the line of 'no information'.

Table 3.5: Estimates and 95% confidence limits for the differences in areas under the ROC curves between each stroke scale for the full and categorised versions in the Barber et al. (2004) cohort.

Outcome	SSS - mRS	SSS - BI	mRS - BI
Death			
Full	0.007 (-0.014, 0.028)	0.006 (-0.009, 0.021)	-0.001 (-0.015, 0.013)
	$p = 0.512$	$p = 0.423$	$p = 0.908$
Category	-0.036 (-0.062, -0.011)	-0.019 (-0.040, 0.002)	0.017 (0.003, 0.032)
	$p = 0.006$	$p = 0.079$	$p = 0.020$
Death or dependency			
Full	-0.017 (-0.030, -0.003)	-0.027 (-0.041, -0.013)	-0.010 (-0.021, 0.0002)
	$p = 0.014$	$p < 0.001$	$p = 0.055$
Category	-0.043 (-0.062, -0.024)	-0.022 (-0.041, -0.003)	0.021 (0.007, 0.035)
	$p < 0.001$	$p = 0.023$	$p = 0.003$
Death or institutional care			
Full	-0.005 (-0.009, -0.020)	-0.009 (-0.023, -0.004)	-0.015 (-0.026, -0.003)
	$p = 0.470$	$p = 0.181$	$p = 0.011$
Category	-0.014 (-0.038, 0.010)	-0.004 (-0.026, 0.019)	0.010 (-0.006, 0.027)
	$p = 0.250$	$p = 0.752$	$p = 0.217$

Note: positive values indicate the area under the ROC curve is greater for the first scale, while negative values indicate the area under the ROC curve is greater for the second scale. SSS represents Scandinavian Stroke Scale; mRS - modified Rankin Scale; and BI - Barthel Index.

difference in area between scales for the full and categorised versions.

For the Barber et al. (2004) cohort (Table 3.5) the majority of the significant differences in areas lie within the death or dependency outcome. For both the full and categorised scales, mRS has a larger area under the curve than SSS and BI, while BI has a larger area than SSS, suggesting mRS has the best predictive accuracy for one month following stroke. However, as before, even though the differences are statistically significant, their estimates are small with narrow confidence intervals.

In the Sellars et al. (2007) cohort (Table 3.6) any differences in predictive accuracy between stroke scales are less apparent three months after stroke. There

Table 3.6: Estimates and 95% confidence limits for the differences in areas under the ROC curves between each stroke scale for the full and categorised versions in the Sellars et al. (2007) cohort.

| Outcome	NIHSS - mRS	NIHSS - BI	mRS - BI
Death			
Full	-0.002 (-0.039, 0.035)	-0.035 (-0.068, -0.002)	-0.033 (-0.057, -0.008)
	$p = 0.900$	$p = 0.037$	$p = 0.009$
Category	-0.013 (-0.053, 0.026)	-0.029 (-0.064, 0.006)	-0.015 (-0.044, 0.013)
	$p = 0.511$	$p = 0.109$	$p = 0.291$
Death or dependency			
Full	-0.031 (-0.063, 0.002)	-0.026 (-0.063, 0.011)	0.004 (-0.022, 0.031)
	$p = 0.063$	$p = 0.164$	$p = 0.752$
Category	-0.018 (-0.051, 0.015)	-0.004 (-0.040, 0.032)	0.014 (-0.015, 0.044)
	$p = 0.291$	$p = 0.845$	$p = 0.341$
Death or institutional care			
Full	-0.027 (-0.057, 0.003)	-0.055 (-0.083, -0.028)	-0.029 (-0.052, -0.005)
	$p = 0.082$	$p < 0.001$	$p = 0.018$
Category	-0.038 (-0.072, -0.005)	-0.054 (-0.086, -0.023)	-0.016 (-0.046, 0.013)
	$p = 0.025$	$p < 0.001$	$p = 0.284$

Note: positive values indicate the area under the ROC curve is greater for the first scale, while negative values indicate the area under the ROC curve is greater for the second scale. NIHSS represents the National Institutes of Health Stroke Scale; mRS - modified Rankin Scale; and BI - Barthel Index.

are no significant differences in the death or dependency outcome. For the death or institutional care outcome, categorised mRS and BI have significantly larger areas under the ROC curves than categorised NIHSS, but as for the Barber et al. (2004) cohort, the confidence intervals are narrow and the estimated differences are small.

Weighted kappa analysis between categorised scales within each cohort was performed to test the agreement between scales (Table 3.7). The lower triangular entries (values in blue) show the results obtained in the Barber et al. (2004) cohort, while the upper triangular entries (values in red) show the results for the Sellars et al. (2007) cohort. Subjective guidelines, given in Section 3.2, suggest

Table 3.7: Weighted kappa estimates and 95% confidence intervals between categorised stroke scales for both cohorts.

	SSS	mRS	BI
NIHSS	-	0.66 (0.60, 0.71)	0.66 (0.60, 0.71)
mRS	0.77 (0.74, 0.81)	-	0.77 (0.72, 0.82)
BI	0.80 (0.76, 0.83)	0.85 (0.82, 0.89)	-

Note: values in blue in the lower triangular represent the Barber et al. (2004) cohort, while values in red in the upper triangular represent the Sellars et al. (2007) cohort. Values represent the agreement in categorising stroke patients into mild, moderate and severe categories. SSS represents Scandinavian Stroke Scale; NIHSS - National Institutes of Health Stroke Scale; mRS - modified Rankin Scale; and BI - Barthel Index.

BI, mRS and SSS all have excellent agreement with each other when categorised into mild, moderate and severe, while NIHSS has substantial agreement with mRS and BI. Lastly, notice that the agreement measures between mRS and BI for the two cohorts are quite different.

3.4 Discussion

This study shows that the Scandinavian Stroke Scale, the modified Rankin Scale, the Barthel Index and the National Institutes of Health Stroke Scale all have moderate to high predictive accuracy. This is particularly true for the outcome of death and only slightly less so for the composite outcomes of death or dependency and death or institutional care. When the scales are categorised into mild, moderate and severe the reduction in the area under the ROC curve, although statistically significant, is small and may not be clinically important. This suggests that there is little predictive information lost when these scales are categorised. A secondary finding was that the predictive accuracy of the full and categorised

scales was slightly greater in the Barber et al. (2004) cohort compared to the Sellars et al. (2007) cohort, most likely due to the longer follow-up times in the Sellars et al. (2007) cohort.

When comparing the area under the ROC curves between stroke scales it was found that at one month follow-up (Barber et al. (2004) cohort) the mRS predicted death or dependency better than both the SSS and the BI. Dependency was measured using the mRS, therefore it would be expected that baseline mRS would be better at predicting one month mRS. At three months follow-up (Sellars et al. (2007) cohort) the mRS and BI predicted death or institutional care better than NIHSS. Although these comparisons were statistically significant, estimates for the difference between scales either when full or categorised were small, suggesting they may not be clinically important. Lastly, the weighted kappa analyses for the categorised scales showed that there was substantial to almost perfect agreement amongst all scales.

These analyses used data from two different cohorts, both of which contained several hundred patients from unselected hospital admissions with a broad range of case mix. Within each cohort there were few missing data, limiting the possibility of a bias in the collection of data possibly due to the recording of severity being dependent on patient characteristics. Each cohort had several measurements of severity allowing four different stroke scales to be tested for equivalence between the full and categorised versions. The scales used in this analysis are recorded in the vast majority of stroke trials. Additionally, the follow-up times for the Barber et al. (2004) and Sellars et al. (2007) cohorts were different (one month and three months, respectively) showing that the results obtained do not simply apply to one stroke scale or to a specific length of follow-up.

One obvious limitation to these analyses is that data were not available to directly compare SSS with NIHSS, or with any of the other stroke scales currently used in practice. Also, data were not available to allow a direct comparison of predictive ability at different follow-up lengths via longitudinal measurements of outcome within a cohort.

Future research should focus on comparing the stroke scales used here with other scales used in practice such as the Glasgow Coma Scale (Jennett and Teasdale, 1977), the Canadian Neurological Scale (Côté et al., 1989) or the Edinburgh Stroke Predictor (Weir et al., 2003). Follow-up lengths in meta-analyses are often longer than those for which data were available, for example, six or twelve months. Therefore, since predictive accuracy is lower for longer follow-up times, it may be of interest to perform comparisons of AUC for stroke scales at later follow-up times to assess if the equivalence of full and categorised scales is maintained.

This study shows that the categorisation of stroke scales into mild, moderate and severe does not substantially reduce the predictive ability of the scale. Stroke scales stratified in this way appear to be equivalent to each other and although the prognostic accuracy for longer follow-up is reduced, it is not further reduced by categorisation of the stroke scales. These findings emphasise that stratification of randomisation in acute stroke clinical trials according to severity classified as mild, moderate or severe is a pragmatic approach which retains much of the prognostic information contained in the corresponding full assessment scale.

Chapter 4

Organised inpatient (stroke unit) care for stroke

In this Chapter frequentist meta-analysis approaches are used in order to address the question of whether organised inpatient (stroke unit) care improves the survival and independence of stroke patients and their ability to return home. Several comparisons of "more" organised with "less" organised care are conducted allowing more specific service comparisons to be viewed. This analysis includes outcome data from recent trials and there is interest in whether the conclusions reached differ from the previous version of this review.

4.1 Background

During the acute phase of stroke, patients are frequently admitted to hospital where they can receive care in a variety of different ways or settings. Previously, the care of patients with stroke was provided within general medical or neurology

wards alongside a range of other patient groups where they would be managed by non-specialist staff without routine multidisciplinary input. Organised inpatient (stroke unit) care is a term used to describe the focusing of care for patients with stroke in hospital under a multidisciplinary team who specialise in stroke management (Stroke Unit Trialists' Collaboration, 2007).

There is uncertainty whether the perceived effort and cost in focusing the care of stroke patients in hospital within specially organised units would improve the survival and recovery of patients receiving that care (Stroke Unit Trialists' Collaboration, 2007). In this study, the questions addressed are whether improving the organisation of inpatient stroke services can bring about improvements in important patient outcomes and whether the conclusions reached following the inclusion of outcome data from recent trials are altered from the previous version of this review (Stroke Unit Trialists' Collaboration, 2006).

4.2 Methods

4.2.1 Trial identification

The Stroke Unit Trialists' Collaboration searched the Cochrane Stroke Group Trials Register (last searched April 2006) to identify appropriate trials for inclusion (full details of search strategy given in Sandercock et al. (2008)). In order to identify additional trials, reference lists of relevant articles were studied, colleagues and researchers were contacted and preliminary findings were publicised at stroke conferences in the UK, Scandinavia, Germany, Netherlands, Switzerland, Spain, Canada, South America, Australia, Belgium, USA and Hong Kong.

The search was not restricted by date, language or any other criteria.

The study included all prospective clinical trials that used some form of random allocation of stroke patients to an organised system of inpatient (stroke unit) care or an alternative form of inpatient care. The definition of stroke unit care used was very broad (defined as a multidisciplinary team specialising in stroke care) and included services based in a discrete ward or provided by a mobile stroke team. Contemporary conventional care was defined as being provided in a general medical ward or less organised form of stroke care. Section 1.2.1 provides a detailed description of the types of intervention considered in this study. A clinical definition of stroke (focal neurological deficit due to cerebrovascular disease, excluding subarachnoid haemorrhage and subdural haematoma) was used and any patients admitted to hospital following stroke were eligible.

Published trials were scrutinised by two review authors who assessed their eligibility and methodological quality. Characteristics of unpublished trials were established through discussion with the trial co-ordinator. Trials were included if treatment allocation was carried out in a strictly random basis or with a quasi-random procedure, such as bed availability or date of admission. A formal scoring system for methodological quality was not used but method of allocation concealment, completeness of follow-up, presence of intention to treat analysis and the presence of blinded assessment to follow-up were recorded.

4.2.2 Data extraction

Descriptive information about the service characteristics of the organised inpatient (stroke unit) care and conventional care settings were obtained through a

structured interview or correspondence with the trial co-ordinators. Trials were then allocated to six service subgroups (acute semi-intensive stroke unit; comprehensive stroke unit; rehabilitation stroke unit; mixed rehabilitation ward; mobile stroke team; and general medical ward) as defined in Section 1.2.1.

The primary outcomes were death, and the combined outcomes of death or dependency and death or the requirement for institutional care at the end of scheduled follow-up of the original trial. Dependency was defined as requiring physical assistance for one or more of the following criteria: transfers, mobility, dressing, feeding or toileting. These criteria were approximately equivalent to modified Rankin Scale > 2, Barthel Index < 19, or an Activity Index (AI) > 83 as described by Hamrin (1982). The requirement for institutional care was taken to be care in a residential home, nursing home or hospital at the end of scheduled follow-up. Trial co-ordinators were asked to provide the number of patients who were dead, dependent, requiring institutional care and missing at the end of scheduled follow-up. Outcome data from published sources were confirmed and supplemented with unpublished information provided by the trial co-ordinators. A secondary outcome was duration of stay in hospital, institution or both. Patient quality of life and patient and carer satisfaction were also assessed as secondary outcomes but are not analysed in this review update.

The following subgroup information was sought primarily for the outcome of death or requiring institutional care:

1. age: up to 75 years or greater than 75 years;

2. sex: male or female;

3. stroke severity: mild, moderate or severe, as defined by the Barthel Index

in Table 3.1.

Unpublished aggregate data were obtained for the majority of trials but insufficient amounts of individual patient data were available to allow a comprehensive IPD analysis.

4.2.3 Description of studies

A total of 48 trials were identified. Of these, 13 were excluded, two are awaiting assessment and two are ongoing. The remaining 31 trials contained outcome information on a total of 6936 participants (Ronning and Guldvog, 1998; Vemmos et al., 2001; Ma et al., 2004; Peacock et al., 1972; Patel, 2000; Stevens et al., 1984; Garraway et al., 1980; Svensson et al., 1993; Fagerberg et al., 2000; Sulter et al., 2003; Kaste et al., 1995; Gordon and Kohn, 1966; Cabral et al., 2003; Sivenius et al., 1985; Dey et al., 2005; Wood-Dauphinee et al., 1984; Feldman et al., 1962; Aitken et al., 1993; Juby et al., 1996; Kalra et al., 1993; Kalra and Eade, 1995; Kalra et al., 2000; Yagura et al., 2005; Cavallini et al., 2003; Hankey et al., 1997; von Arbin et al., 1980; Laursen et al., 1995; Ilmavirta, 1994; Indredavik et al., 1991; Strand et al., 1985; Hamrin, 1982).

The principal review author conducted a structured interview with the trial co-ordinators for 23 trials to determine the service characteristics. For four trials, access to detailed unpublished information was available, while for the remaining four there was access to published information only. Service comparisons within these 31 trials are summarised in Figure 4.1. The most common comparison was between comprehensive stroke ward (service 2) and general medical ward (service 6) with eleven trials looking at this comparison.
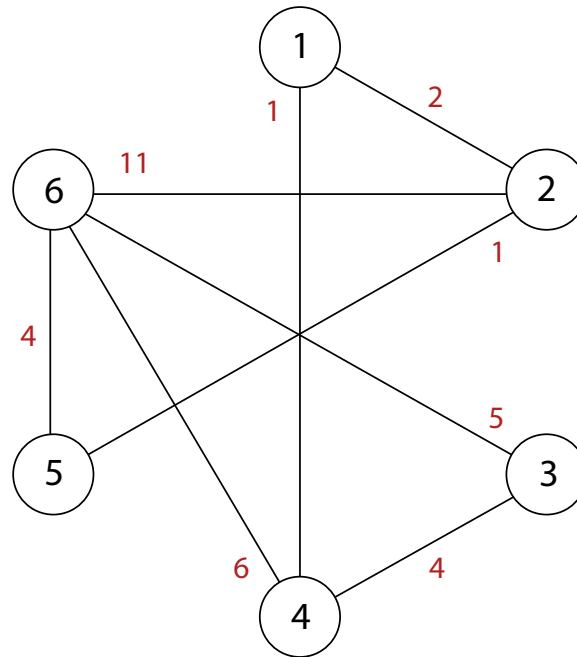
Figure 4.1: Summary of number of service comparisons within included trials (denoted by red numbers). Numbers within nodes denote treatments: 1 is acute stroke ward; 2 is comprehensive stroke ward; 3 is rehabilitation stroke ward; 4 is mixed rehabilitation ward; 5 is mobile stroke team; and 6 is general medical ward. Treatments 1, 2 and 3 are subcategories of stroke unit. Since three of the trials each had three treatment arms, the total number of comparisons is greater than the number of trials.

Of the 31 trials where data were available, 16 used a formal randomisation procedure such as random numbers, sequentially numbered sealed envelopes or central randomisation. Eight trials used an unclear method of randomisation and the remaining seven used an informal randomisation based on bed availability (Yagura et al., 2005; Cavallini et al., 2003; von Arbin et al., 1980; Strand et al., 1985), strict admission rota (Patel, 2000; Hamrin, 1982) or patient date of birth (Ronning and Guldvog, 1998). Sensitivity analysis was performed removing these seven trials from the analysis to exclude possible bias in the conclusions. Only

ten trials used a blinded assessment of outcome for all patients.

For trials with missing outcome data it was assumed these patients were alive and living at home, which may have introduced a small bias in favour of the control group in the scenario where patient outcomes are better for those admitted to a stroke unit.

A more detailed description of the trials included in this study can be found in Stroke Unit Trialists' Collaboration (2007).

4.2.4 Statistical methods

Data were analysed using the Cochrane Collaborations statistical software, Revman (Cochrane Collaboration, 2003), where primary outcomes (death, death or dependency, death or institutional care) were calculated as the odds ratios with 95% confidence intervals of an adverse outcome in more organised versus contemporary conventional care. Unless there was significant heterogeneity, a fixed-effect approach (Peto method, Section 1.4.1) was used. However, in the case of heterogeneity, a random-effects approach (DerSimonian and Laird method, Section 1.4.2) was used instead.

The Q-statistic is used to measure heterogeneity and is calculated as:

$$Q = \sum_{i=1}^{k} w_i (\theta_i - \hat{\theta})^2,$$

where $\hat{\theta}$ is the estimated pooled treatment effect, θ_i are the individual trial treatment effect estimates and $w_i = 1/var(\theta_i)$. This is compared to a χ^2 distribution with $k-1$ degrees of freedom, where k is the number of trials, to obtain a p-value indicating the extent of between-study variability (Higgins and Thompson, 2002).

Length of stay in a hospital or institution was analysed using standardised mean difference (SMD) with random-effects (Cochrane Collaboration, 2005). The SMD for each trial, i, including adjustment for small sample bias, is given by:

$$\text{SMD}_i = \left(\frac{m_{1i} - m_{2i}}{s_i} \right) \left(1 - \frac{3}{4N_i - 9} \right),$$

where m_{1i} and m_{2i} are the means in each treatment arm and

$$s_i = \sqrt{\frac{(n_{1i} - 1)\tau_{1i}^2 + (n_{2i} - 1)\tau_{2i}^2}{N_i - 2}},$$

are the pooled standard deviation of the two groups. N_i is the total number of subjects in trial i, n_{1i} and n_{2i} are the total number of subjects in each treatment arm of trial i, and τ_{1i} and τ_{2i} are the standard deviations in each treatment arm of trial i. The overall pooled SMD is given by:

$$\text{SMD} = \frac{\sum_{i=1}^{m} w_i \text{SMD}_i}{\sum_{i=1}^{m} w_i},$$

where $w_i = 1/var(\text{SMD}_i)$.

4.3 Results

Due to the different service comparisons within this study, the results are presented in three sections: Section 4.3.1 compares organised inpatient (stroke unit) care with an alternative service; Section 4.3.2 describes the most common comparison of organised inpatient (stroke unit) care against general medical ward; and finally Section 4.3.3 compares different forms of organised inpatient (stroke

unit) care.

4.3.1 Organised inpatient (stroke unit) care vs. alternative service

This section examines the impact of increased levels of organisation of stroke care on patient outcomes. If both services satisfy the definition of stroke unit then the less organised system of care was taken as the comparator service. Figure 4.2 summarises the comparison between organised (stroke unit) care with an alternative (less organised) form of care. Comparisons are: (a) stroke ward (including rehabilitation ward, comprehensive ward and acute ward) versus general medical ward; (b) mixed rehabilitation ward versus general medical ward; (c) mobile stroke team versus general medical ward; (d) stroke ward versus mixed rehabilitation ward; (e) stroke ward versus mobile stroke team; and (f) more organised stroke ward versus less organised stroke ward.

These analyses were not complicated by significant heterogeneity between trials (death: $p=0.28$; death or institutional care: $p=0.16$; death or dependency: $p=0.06$), therefore, the Peto meta-analysis method was used in the analyses. Conclusions remained unchanged when trials with short or variable follow-up were excluded, or when trials with informal randomisation procedures were excluded. Additionally, for the death or dependency outcome, when analysis was restricted to trials with unequivocal blinded final assessment for all patients, the conclusions also remained unchanged.

The odds of death were significantly lower in stroke ward compared to general medical ward (odds ratio 0.83, 95% confidence interval 0.71 to 0.96; $p = 0.01$)
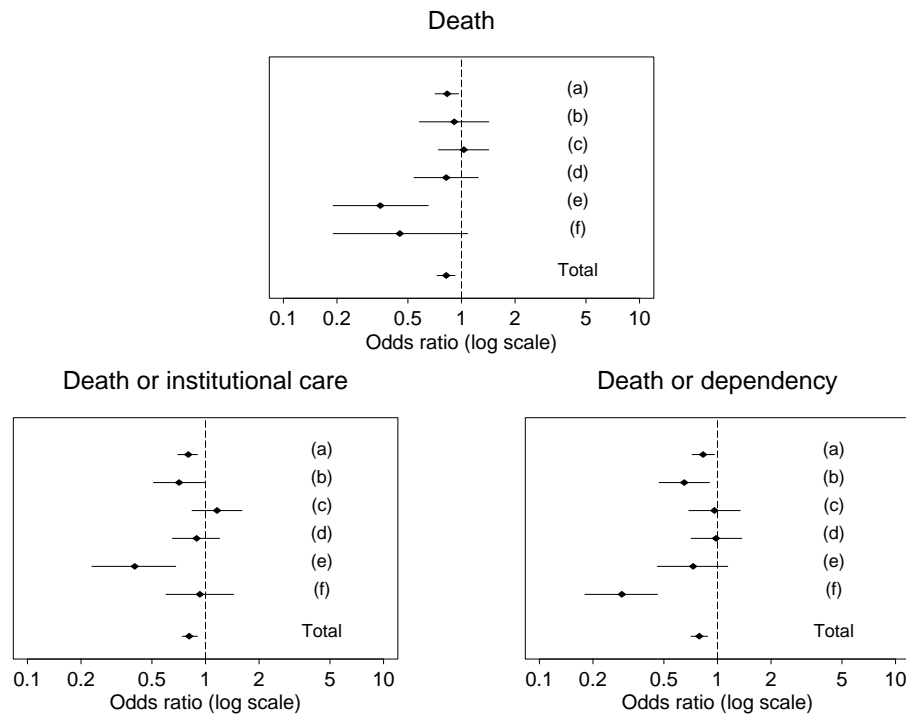
Figure 4.2: Comparing organised stroke unit care with alternative (less or-
ganised) service for the outcomes of death, death or institutional
care and death or dependency. Comparisons are: (a) stroke ward
vs general medical ward; (b) mixed rehabilitation ward vs general
medical ward; (c) mobile stroke team vs general medical ward;
(d) stroke ward vs mixed rehabilitation ward; (e) stroke ward vs
mobile stroke team; and (f) more organised stroke ward versus
less organised stroke ward. The total is the overall pooled esti-
mate from the six comparisons. Odds ratios, obtained using the
Peto meta-analysis method, are represented by the shaded dia-
mond with corresponding 95% confidence intervals represented
by the line. Odds ratios are plotted on the natural log-scale.

and mobile stroke team (0.35, 0.19 to 0.65; $p < 0.001$). The overall estimate

showed a significant reduction in odds of death for more organised care (0.82,

0.73 to 0.92; $p = 0.001$).

For death or requirement for institutional care there was also a reduction in

the odds of poor outcome in stroke ward compared to general medical ward (0.80,

0.70 to 0.90; $p < 0.001$) and mobile stroke team (0.40, 0.23 to 0.68; $p < 0.001$), also with a significant reduction in overall odds of poor outcome in the more organised care group (0.81, 0.74 to 0.90; $p < 0.001$).

For death or dependency there was a significant reduction in odds of poor outcome in stroke ward (0.83, 0.72 to 0.96; $p = 0.01$) and mixed rehabilitation ward (0.65, 0.47 to 0.90; $p = 0.01$) compared to general medical ward, and a significant reduction in more organised stroke ward compared to less organised stroke ward (0.29, 0.18 to 0.46; $p < 0.001$). There was a significant reduction in overall odds of death or dependency in the more organised care group (0.79, 0.71 to 0.88; $p < 0.001$).

Three trials (Vemmos et al., 2001; Juby et al., 1996; Indredavik et al., 1991) extended follow up of patients for a further five years post stroke and two trials for a further ten years (Juby et al., 1996; Indredavik et al., 1991). The summary of results for these trials are given in Figure 4.3. The odds of death and death or institutional care continue to be reduced in stroke unit care five (death: 0.74, 0.59 to 0.94; death or institutional care: 0.62, 0.43 to 0.89) and ten years post stroke (death: 0.53, 0.36 to 0.80; death or institutional care: 0.57, 0.37 to 0.88), while the reduction in odds of death or dependency is still significant after five years (0.59, 0.38 to 0.92) but not after ten years (0.77, 0.45 to 1.31).

For the length of stay outcome, a random-effects model on the standardised mean difference (SMD) showed a modest reduction in length of stay in comprehensive stroke ward compared to general medical ward (SMD -0.19, 95% CI -0.31 to -0.06), which is approximately equivalent to 2.5 days (0.5 to 4.5 days). However, these summary estimates are complicated by significant heterogeneity and there are major methodological limitations to consider. Trials included in this
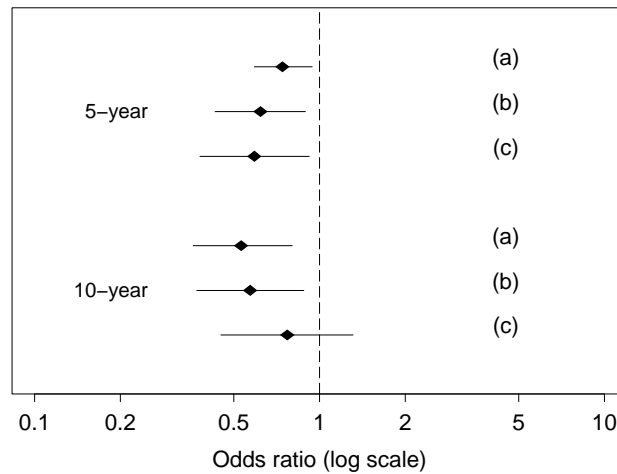
Figure 4.3: Comparing organised stroke unit care with alternative service for the outcomes of (a) death, (b) death or institutional care and (c) death or dependency at 5- and 10-year follow-up. Odds ratios presented as in Figure 4.2.

study recorded length of stay differently (mean or median) and in some cases standard deviations had to be inferred from p-values or from results of similar trials. Trials also calculated length of stay in different ways, such as acute stay or total stay in hospital or institution. These limitations prevent more general conclusions to be formed from this analysis.

4.3.2 Organised inpatient (stroke unit care) care vs. general medical ward

The results described in this section are for the most common comparison of organised stroke unit care versus general medical ward. Figure 4.4 gives a summary
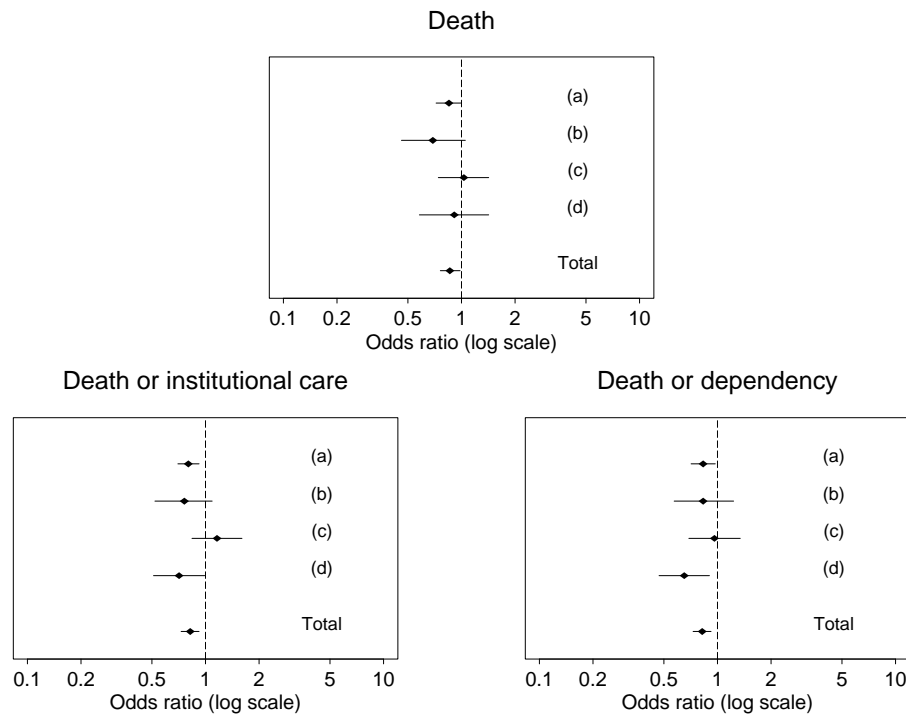
Figure 4.4: Comparing organised stroke unit care with general medical wards for the outcomes of death, death or institutional care and death or dependency. Comparisons are: (a) comprehensive stroke ward; (b) rehabilitation stroke ward; (c) mobile stroke team; and (d) mixed rehabilitation ward. The total is the overall pooled estimate from the four comparisons. Odds ratios presented as in Figure 4.2.

of the comparisons between organised care and general medical wards for the outcomes of death, death or institutional care and death or dependency. Comparison (a) is comprehensive stroke ward; (b) rehabilitation stroke ward; (c) mobile stroke team; and (d) mixed rehabilitation ward.

As in Section 4.3.1, there was no significant heterogeneity between trials (death: $p=0.81$; death or institutional care: $p=0.29$; death or dependency: $p=0.55$) enabling the use of the Peto meta-analysis method.

Table 4.1: Odds ratios and confidence intervals for the comparisons of comprehensive stroke wards and mixed rehabilitations wards with general medical wards for the outcomes of death or institutional care and death or dependency.

Outcome	Odds ratio (95% CI)	*P*-value
Comprehensive stroke ward		
Death or institutional care	0.80 (0.70 to 0.92)	0.002
Death or dependency	0.83 (0.71 to 0.97)	0.02
Mixed rehabilitation ward		
Death or institutional care	0.71 (0.51 to 0.99)	0.04
Death or dependency	0.65 (0.47 to 0.90)	0.01

Compared to general medical wards, the odds of death are significantly reduced in comprehensive stroke wards (0.85, 0.72 to 0.99; $p = 0.03$) and non-significantly reduced in rehabilitation wards (0.69, 0.46 to 1.05; $p = 0.08$) with an overall reduction in odds of death (0.86, 0.76 to 0.98; $p = 0.02$). For the outcomes of death or institutional care and death or dependency there were significant reductions in odds of poor outcome for comprehensive stroke wards and mixed rehabilitation wards, given in Table 4.1, with identical overall reductions in odds of poor outcome in stroke wards (0.82, 0.73 to 0.92, $p < 0.001$).

4.3.3 Comparisons of different forms of organised inpatient (stroke unit) care

This section determines whether the benefits of stroke unit care are only achieved through specialised stroke wards or if the benefits can also be achieved by mobile stroke team care or a generic disability service. Therefore, the results are presented for direct comparisons of different forms of organised care. Figure 4.5 summarises the comparisons of (a) acute stroke ward versus comprehensive stroke
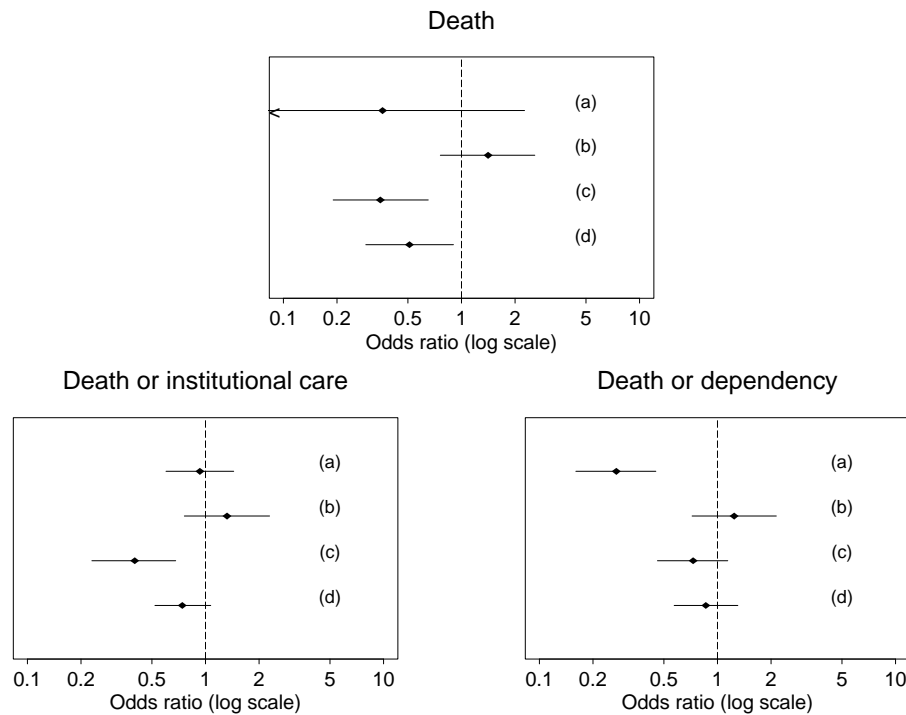
Figure 4.5: Comparing different systems of organised care for the outcomes of death, death or institutional care and death or dependency. Comparisons are: (a) acute stroke ward vs comprehensive stroke ward; (b) acute stroke ward vs mixed rehabilitation ward; (c) comprehensive stroke ward vs mobile stroke team; and (d) rehabilitation stroke ward vs mixed rehabilitation ward. Odds ratios presented as in Figure 4.2.

ward; (b) acute stroke ward vs mixed rehabilitation ward; (c) comprehensive stroke ward vs mobile stroke team; and (d) rehabilitation stroke ward vs mixed rehabilitation ward for the outcomes of death, death or institutional care and death or dependency.

The Peto meta-analysis method was used to analyse the outcomes of death and death or institutional care since there was no significant heterogeneity between trials (Table 4.2). However, there was significant heterogeneity observed for the outcome of death or dependency when comparing acute stroke ward and

Table 4.2: *P*-values for test for heterogeneity in the comparisons of different forms of stroke unit care.

Outcome	*P*-value
Acute stroke ward*	
Death	0.06
Death or institutional care	0.25
Death or dependency	< 0.01
Rehabilitation stroke ward*	
Death	0.95
Death or institutional care	0.92
Death or dependency	0.72

*Versus alternative service.

alternative service. Therefore, for this comparison, the DerSimonian and Laird meta-analysis method was used. For the comparison of comprehensive stroke ward and mobile stroke team, the test for heterogeneity was not applicable as only one trial analysed this comparison.

For the outcomes of death and death or institutional care there tended to be reductions in odds of poor outcome in comprehensive stroke wards compared to mobile stroke teams (death: 0.35, 0.19 to 0.65; death or institutional care: 0.40, 0.23 to 0.68) and stroke rehabilitation wards compared with mixed rehabilitation wards (death: 0.51, 0.29 to 0.90; death or institutional care: 0.74, 0.52 to 1.07). However, there were few trials available for these comparisons and the numbers within trials were small, preventing any formal conclusions from being drawn.

There was a significant reduction in odds of death or dependency in acute stroke wards compared to comprehensive stroke wards (0.27, 0.16 to 0.45), however, there were also significant heterogeneity in the results. After applying a random-effects model, there was no significant difference between services.

4.4   Discussion

This study addresses the question of whether more organised inpatient care could improve patient outcomes in comparison to less organised care and allows more specific service comparisons to be viewed. Section 4.3.1 confirms the conclusions described in previous versions of this review (Stroke Unit Trialists' Collaboration, 2006): patients receiving organised (stroke unit) care are more likely to survive, return home and regain independence than those receiving a less organised service, with a suggestion that the benefit may last for five or even ten years post stroke. Organised inpatient care is typically provided within a discrete ward which offers a substantial period of rehabilitation if required.

There were substantial amounts of new data available for this updated review. A total of eight new trials (for one of which individual patient data were available) were included in addition to the 23 trials in the previous version of the review. Therefore, the main strength of this study is the large number of trials available for analysis. The large amount of new information collected allowed the comparison between several different types of stroke unit care and the use of combined outcomes, such as death or dependency and death or institutional care, are particularly important when event numbers within trials are small. While this information allows the comparison between different types of stroke unit, for some comparisons there are few or no trials available, particularly for comparison of different forms of stroke unit care. This illustrates the need for an indirect comparisons approach to estimate the differences between service types.

Dependency is an important measure of patient outcome but can be affected by observer bias if final assessment was not carried out in a blinded fashion.

The sensitivity analysis performed in the study, where only trials that used un-equivocal blinded assessment were included, suggested that the results were not influenced by this bias. The requirement for institutional care is a good substitute for dependency. However, it may be affected by national health care systems or cultural factors.

The results obtained from the length of stay analysis in this study may be affected by the different methods of recording and calculating the length of stay in each trial, possibly causing the significant heterogeneity. It seems most reasonable to conclude that there was very little, if any, difference in length of stay between more and less organised care.

Finally, one last limitation of this study is that although the majority of trials are recent, some of the trials included are relatively old and standards of care may have changed in recent years. However, these trials are also randomised and so should have reasonable internal validity and not have a confounding effect on the final conclusions.

This study does not explain why stroke unit care improves patient outcomes. Some suggestions are the prevention of complications or more intense monitoring of acute patients. Future research should focus on which components of stroke unit care provide benefit to stroke patients, as examined in Chapters 5 and 6. Direct comparisons of different models of stroke unit care could also be examined; however, since this may not be achievable, Chapter 8 examines possible differences using an indirect comparisons method. Collaboration between stroke trialists should be pre-planned to ensure similar variables and outcomes are measured within trials in order to lessen the problems that occur in retrospective systematic reviews.

Chapter 5

Prevention of complications in organised inpatient (stroke unit) care

In Chapter 4 it was shown that patients receiving more organised inpatient (stroke unit) care are more likely to survive regain independence and return home than those receiving a less organised service. However, the analysis did not explain how these benefits were achieved. Using simple Bayesian random-effects models, this Chapter explores whether the use of interventions to prevent complications explains the benefit of organised inpatient (stroke unit) care.

5.1   Background

It has been known for many years that organised inpatient (stroke unit) care reduces the risk of death after stroke (Stroke Unit Trialists' Collaboration, 1997a),

but it is not clear how this benefit is achieved. The Stroke Unit Trialists Collaboration carried out an analysis 10 years ago which suggested that Stroke Units may reduce deaths through preventing complications (Stroke Unit Trialists' Collaboration, 1997*b*). However, this analysis had limited statistical power and its conclusions were speculative.

In the most recent update of the stroke unit systematic review (Stroke Unit Trialists' Collaboration, 2007), data were available from a larger number of controlled clinical trials. This allowed the question "does the prevention of complications explain the survival benefit of stroke unit care?" to be revisited. If this is the case then one would expect the following observations to be associated with stroke unit care:

a) the more frequent use of interventions designed to prevent complications,

b) a smaller number of recorded serious complications, and

c) fewer deaths attributed to complications.

This Chapter describes a further analysis of the stroke unit review which addresses these questions.

5.2 Methods

5.2.1 Data

This is a further analysis of a collaborative systematic review carried out by the Stroke Unit Trialists Collaboration (Stroke Unit Trialists' Collaboration, 2007) as described in Section 1.5.4 and Chapter 4. In summary, this database includes

31 prospective trials using some form of random allocation of stroke patients to an organised system of inpatient (stroke unit) care or an alternative form of inpatient care. In addition to the main outcomes (death, dependency and need for institutional care) and subgroup data (age and severity) already available, information was sought on the following outcomes:

1. specific interventions directed at reducing complications,

2. complications recorded during early hospital care (first 4 weeks), and

3. certified cause of death during follow-up.

The exact criteria used were those defined in the individual trials.

The specific interventions directed at reducing complications included antibiotic therapy, measures to prevent aspiration (systematic assessment of swallowing and modification of dietary intake), intravenous fluids, insulin, oxygen, paracetamol, tube feeding and urinary catheterisation.

Complications were classified into four categories to reflect previous epidemiological work linking complications to cause of death (Bamford et al., 1990):

a) neurological (cerebral oedema, stroke recurrence, stroke progression, seizures, anxiety, depression),

b) cardiovascular complications (myocardial infarction, arrhythmia, congestive cardiac failure),

c) complications of immobility (chest infection, urinary tract infection, other infections, dehydration, venous thromboembolism, falls, pressure sores, pain), and

d) other complications (for example cancer, gastro-intestinal haemorrhage, suicide).

In addition, common 'physiological complications', defined as physiological abnormalities which did not fulfil a conventional medical diagnosis, were recorded. These included hypertension, hyperglycaemia, hypoxia, hypotension and pyrexia. The specific definitions of these complications were as reported within the original trials.

Cause of death was recorded at the end of scheduled follow up for the majority of trials with this information. However, there were three trials which recorded at discharge (Gordon and Kohn, 1966; Yagura et al., 2005; Cavallini et al., 2003), three trials which recorded at an earlier fixed time point (Vemmos et al., 2001; Kalra et al., 2000; Hamrin, 1982) and one trial with incomplete data (Aitken et al., 1993). The median time for recorded cause of death was six months with an inter-quartile range of three to twelve months.

5.2.2 Statistical methods

Data were analysed using hierarchical Bayesian models in WinBUGS. A direct random-effects model was used to calculate odds ratios and 95% credible intervals (CrI). The model used is described in Section 1.4.3 with an Inverse-Gamma distribution for the treatment effect variance. The WinBUGS code for this model is given in Appendix A.

Absolute risk differences for cause of death were also calculated using hierarchical models in WinBUGS. This model is described by Warn et al. (2002). As with the log-odds ratio model described in Section 1.4.3, suppose, for trial i, n_{Ci}

and $n_{\mathrm{T}i}$ are the number of patients in the control and treatment groups, respectively, and $r_{\mathrm{C}i}$ and $r_{\mathrm{T}i}$ are the number of events in the control and treatment groups, respectively, where

$$
\begin{aligned}
r_{\mathrm{C}i} &\sim \mathrm{Binomial}(p_{\mathrm{C}i}, n_{\mathrm{C}i}), \\
r_{\mathrm{T}i} &\sim \mathrm{Binomial}(p_{\mathrm{T}i}, n_{\mathrm{T}i}).
\end{aligned}
$$

In Equations 1.2 and 1.3, the logits of the control and treatment group risks ($p_{\mathrm{C}i}$ and $p_{\mathrm{T}i}$) are replaced by the risks themselves, giving

$$
\begin{aligned}
p_{\mathrm{C}i} &= \mu_i, \\
p_{\mathrm{T}i} &= \mu_i + \delta_i,
\end{aligned}
$$

where δ_i is now the absolute risk difference. However, since $p_{\mathrm{C}i}$ and $p_{\mathrm{T}i}$ are probabilities, they must be constrained to lie between 0 and 1. Therefore $\mu_i \sim$ Uniform$(0, 1)$ and δ_i is constrained to the interval $[-p_{\mathrm{C}i}, 1-p_{\mathrm{C}i}]$ so that the model is now

$$
\begin{aligned}
p_{\mathrm{C}i} &= \mu_i, \\
p_{\mathrm{T}i} &= \mu_i + \min(\max(\delta_i, -p_{\mathrm{C}i}), 1 - p_{\mathrm{C}i}).
\end{aligned}
$$

As with the log-odds ratio model, $\delta_i \sim$ Normal(d, τ^2) where $d \sim$ Normal$(0, 10^6)$ and $1/\tau^2 \sim$ Gamma$(0.001, 0.001)$.

The WinBUGS code for this model is given in Appendix B.

Model fit was checked using residual deviance (Spiegelhalter et al., 1998)

calculated as

$$\bar{D}_{res} = \sum_{i,j} 2\left(r_{ij}\log\left(\frac{r_{ij}}{\hat{r}_{ij}}\right) + (n_{ij} - r_{ij})\log\left(\frac{n_{ij} - r_{ij}}{n_{ij} - \hat{r}_{ij}}\right)\right),$$

where n_{ij} is the total number of subjects in trial i treatment j, r_{ij} is the observed number of events and $\hat{r}_{ij}$ is the expected number of events from the model, calculated as $\hat{r}_{ij} = \hat{p}_{ij}n_{ij}$.

Deviance measures the fit of the model to the datapoints using the likelihood function. To calculate residual deviance, the deviance of the saturated model is subtracted from the fitted model and, under the null hypothesis of an adequate fit, one would expect the residual deviance to have a mean equal to the number of datapoints (Spiegelhalter et al., 2002).

5.3 Results

A subset of trials within the updated systematic review (Stroke Unit Trialists' Collaboration, 2007) were able to provide much more detailed data for these additional analyses of complications as outlined below. Further details of the included trials are summarised in Chapter 4. For all outcomes the fit of the model was adequate, as given by residual deviance. Sensitivity analysis was carried out since Bayesian analyses can be sensitive to the choice of priors and initial values. The conclusions were unaffected by choice of prior distribution and initial values.

5.3.1 Interventions to prevent complications

Up to seven trials (1652 patients) recorded the number of patients receiving one or more intervention to prevent complications (Ronning and Guldvog, 1998; Patel, 2000; Sulter et al., 2003; Wood-Dauphinee et al., 1984; Kalra et al., 1993, 2000; Indredavik et al., 1991). The results of this analysis are shown in Table 5.1. The use of the following interventions was associated with stroke unit care: measures to prevent aspiration (odds ratio 2.42, 95% credible interval 1.36 to 4.36); oxygen therapy (2.39, 1.36 to 4.66); paracetamol (2.80, 1.14 to 4.83); and possibly reduced use of urinary catheterisation (0.58, 0.27 to 1.11). The credible interval for tube feeding in incredibly wide indicating that there is large uncertainty for the effect of tube feeding. This could be due to heterogeneity or the credible interval being driven by the use of diffuse priors due to lack of data.

Table 5.1: Frequency and comparison of intervention use in stroke units versus conventional care.

| Intervention | Number of events | | Odds ratio |
	Stroke unit (%)	Control (%)	(95% CrI)
Antibiotics	109 (36.6)	75 (24.5)	1.76 (0.25, 18.77)
Aspiration prevention	44 (28.9)	22 (14.5)	2.42 (1.36, 4.36)
Intravenous fluids	473 (76.5)	319 (48.4)	2.59 (1.20, 33.02)
Insulin	46 (8.6)	34 (6.3)	1.42 (0.72, 3.03)
Oxygen	185 (52.3)	120 (33.9)	2.39 (1.36, 4.66)
Paracetamol	212 (37.9)	106 (18.7)	2.80 (1.14, 4.83)
Tube feeding	58 (26.7)	27 (12.4)	2.36 (0.02, 100>)
Urinary catheter	72 (21.1)	99 (29.3)	0.58 (0.27, 1.11)

Results are presented as median (posterior) odds ratios with 95% credible intervals of intervention use in stroke units versus conventional care.

5.3.2   Complications during acute hospital stay

Up to eight trials (1824 patients) recorded the number of patients having one or more complication (Ma et al., 2004; Sulter et al., 2003; Wood-Dauphinee et al., 1984; Kalra et al., 1993, 2000; Cavallini et al., 2003; Ilmavirta, 1994; Indredavik et al., 1991). The main findings are summarised in Table 5.2. Statistically significant reductions in stroke progression or recurrence, chest infection, other infections, falls and pressure sores were seen in stroke units. There were no significant differences in odds of physiological complications.

5.3.3   Certified cause of death

Seventeen trials (3327 participants) had information on certified cause of death available (Vemmos et al., 2001; Stevens et al., 1984; Garraway et al., 1980; Sulter et al., 2003; Gordon and Kohn, 1966; Sivenius et al., 1985; Dey et al., 2005; Wood-Dauphinee et al., 1984; Feldman et al., 1962; Aitken et al., 1993; Kalra et al., 2000; Yagura et al., 2005; Cavallini et al., 2003; Hankey et al., 1997; Ilmavirta, 1994; Indredavik et al., 1991; Hamrin, 1982). Within this group of trials organised (stroke unit) care resulted in reduced all-cause case fatality (0.75, 0.59 to 0.92). The results for certified cause of death are summarised in Table 5.3 and indicate that significant reductions in deaths were observed for complications of immobility (0.59, 0.41 to 0.86), but not for any other categories. When these are analysed as absolute risk difference notice that there is a reduction in deaths attributed to complications of immobility of approximately 3 deaths per 100 stroke patients.

Table 5.2: Frequency and comparison of occurrence of complications in stroke units versus conventional care.

| Complication | Number of events | | Odds ratio |
	Stroke unit (%)	Control (%)	(95% CrI)
Neurological			
Anxiety or depression	112 (16.7)	132 (19.7)	0.74 (0.27, 1.97)
Seizures	15 (2.7)	17 (3.1)	0.86 (0.37, 1.95)
Stroke progression/recurrence*	85 (9.4)	121 (13.5)	0.66 (0.46, 0.95)
Cardiovascular			
Cardiovascular†	83 (14.2)	66 (11.0)	1.52 (0.58, 4.54)
Complications of immobility			
Chest infection	87 (12.0)	134 (18.6)	0.60 (0.42, 0.87)
Dehydration	21 (5.1)	43 (10.1)	0.81 (0.31, 2.53)
Venous thromboembolism‡	30 (4.4)	35 (5.0)	0.85 (0.49, 1.49)
Falls	28 (18.4)	43 (28.3)	0.57 (0.33, 0.97)
Other infections§	122 (13.5)	201 (21.9)	0.56 (0.40, 0.84)
Pain	70 (12.1)	71 (12.3)	0.73 (0.14, 2.60)
Pressure sores	21 (4.7)	43 (9.6)	0.44 (0.22, 0.85)
Other complications			
Other	22 (2.9)	24 (3.1)	0.95 (0.46, 2.10)
Physiological complications			
High blood pressure	21 (13.0)	9 (5.6)	1.95 (<0.01, 100>)
Hyperglycaemia	55 (14.3)	71 (17.8)	0.78 (0.43, 1.49)
Hypoxia	33 (10.5)	28 (8.9)	2.33 (0.04, 100>)
Low blood pressure	60 (22.1)	68 (25.1)	1.85 (0.04, 100>)
Pyrexia	102 (19.6)	112 (20.9)	0.88 (0.39, 1.86)

Results are presented as median odds ratios with 95% credible intervals of occurrence of complications in stroke units versus conventional care.

*Stroke progression and early recurrence were often not distinguished in the original trials.

†Individual cardiovascular complications (for example ischemic heart disease, arrythmia) were usually grouped together.

‡Includes deep vein thrombosis and pulmonary embolism

§Predominately urinary tract infection

5.4 Discussion

It has been recognised over the last decade that patients who are managed in an organised inpatient (stroke unit) setting are more likely to survive, return

Table 5.3: Comparison of certified cause of death in stroke unit versus conventional care.

Cause of death	Stroke unit (%)	Control (%)	Absolute risk difference (95% CrI)	Odds ratio (95% CrI)
Neurological	8.1	7.8	0.20 (-2.66, 2.90)	1.07 (0.81, 1.44)
Cardiovascular	3.6	4.9	-1.20 (-4.27, 1.24)	0.71 (0.45, 1.10)
Immobility-related	4.4	7.0	-2.84 (-5.81, -0.10)	0.59 (0.41, 0.86)
Other	2.4	3.2	-1.42 (-4.16, 0.90)	0.74 (0.47, 1.17)
Overall	18.5	23.0	-4.77 (-8.81, -0.83)	0.75 (0.59, 0.92)

Percentages (%) are the percentage of total patients in each group that died from a particular cause of death. Absolute risk difference is given as the number of deaths per 100 patients with stroke with corresponding 95% credible interval. Odds ratios are median with 95% credible intervals

home, and regain independence, than those managed in conventional care settings (Stroke Unit Trialists' Collaboration, 1997*a*). However, there has been considerable uncertainty as to why this benefit may occur and how stroke unit care could influence outcomes. In a previous analysis from Stroke Unit Trialists' Collaboration (1997*b*), it was suggested that some of the survival benefit of stroke unit care may be explained by a reduction in complications. However, there was limited statistical power to carry out this analysis. In the current update, considerably larger amounts of data were available, which indicated that stroke unit care appeared to reduce complications of immobility (in particular, infections), although there were also reductions in stroke recurrence or progression. The current analysis suggests that some of these reductions could be explained by a more comprehensive implementation of measures to prevent complications: in particular, measures to prevent aspiration, oxygen treatment, and treatment for pyrexia.

Although this analysis has a number of strengths, in particular using a much larger dataset than previously available, a number of limitations must also be

acknowledged. Firstly, although a pooled analysis of a number of trials was carried out there is still limited information around some complications (particularly physiological complications) and credible intervals are correspondingly broad. Secondly, the recording of some complications in the included trials were often not done in a blinded fashion and varying definitions of complications may have been used. For example, it was often difficult to distinguish between the complications of very early stroke recurrence and progression of the original stroke symptoms. Therefore, the current analysis may have been subject to observer bias. Similarly, the information on certified cause of death is frequently not confirmed by post-mortem examination and so could also be subject to bias. Thirdly, the analysis of complications may be difficult to interpret. In theory, careful monitoring could identify and treat more problems than those identified in a less careful model of care. Fourthly, early mobilisation and training was reported as an objective of care in most of the included trials. However, no standard definition of measuring mobilisation was used, meaning this potentially important aspect of care could not be analysed. Likewise, other components of stroke unit care (for example, prompt use of antithrombotic drugs, improved monitoring) could not individually be analysed. Finally, the analysis demonstrates an association between stroke unit care and reduction in certain complications, but does not explain how this effect was achieved. Individual patient data analysis would enable a more direct analysis of association between the increase in use of interventions and the decrease of recorded complications. That is, determining whether those patients who received interventions are also the ones who avoided complications.

Although this analysis suggests that stroke unit care may have helped prevent

complications, the picture is likely to be complex and there are other possibilities. It is plausible that early stroke unit care could have resulted in patients having less disabling symptoms and hence were less prone to suffer complications. It is also plausible that if patients in stroke units are less likely to die through other (identified) mechanisms, they would also be less likely to suffer complications associated with the last stages of life. The analysis cannot conclusively discriminate between these competing possibilities.

Despite these remaining uncertainties, the findings emphasise the potential importance of complications as a treatable factor in stroke outcome. Future research should explore the best ways of preventing and managing specific complications, particularly those that seem to carry a high risk of causing harm.

Chapter 6

Routine automated physiological monitoring in acute stroke

Similarly to Chapter 5, this Chapter examines how the benefits of organised inpatient (stroke unit) care are achieved. By expanding the aggregated data obtained from trials to form individual patient data, covariate effects are introduced to the Bayesian random-effects models used in Chapter 5 to examine whether routine automated monitoring for, and treatment of physiological complications reduce adverse outcomes in stroke patients.

6.1 Background

Although organised inpatient (stroke unit) care is well established in acute stroke management there remain considerable uncertainties about the value of individual components of stroke unit care (Langhorne and Dennis, 2004). One topical and controversial area concerns the lengths to which clinical staff should go to identify

89

and treat common physiological abnormalities (such as pyrexia, hypoxia, blood pressure abnormalities, fast or slow heart rate and hyperglycaemia). Although a small number of clinical trials have addressed this question, the findings have been inconclusive.

A collaborative systematic review and meta-analysis was conducted comparing a conventional stroke unit approach with conventional stroke unit care plus continuous monitoring in the acute phase, and addresses whether routine automated monitoring for and treatment of physiological abnormalities (compared with conventional intermittent monitoring) will:

1. result in more physiological abnormalities being detected and treated,

2. reduce the risk of stroke progression,

3. reduce the risk of long term dependency or death, and

4. be most effective in "high risk" patients (those who are older with co-morbidity and more severe stroke).

6.2 Methods

6.2.1 Data

A search was conducted for trials which compared routine automated monitoring in a stroke unit versus conventional stroke unit care as source data for meta-analysis and meta-regression. In addition to data from the Stroke Unit Trialists Collaboration systematic review (Stroke Unit Trialists' Collaboration, 2007) (where the Cochrane Stroke Group Trials Register was last searched for trials in

April 2006, full details of search strategy given in Sandercock et al. (2008)), reference lists of related articles were scanned, and colleagues and researchers were contacted in an effort to identify published, unpublished and ongoing trials. In addition to existing grouped data, individual patient data was requested. However, if these could not be obtained, the most detailed published and unpublished data available were used.

The search identified four completed clinical trials with a total of 785 patients (Davis et al., 2000; Sulter et al., 2003; Cavallini et al., 2003; Ilmavirta, 1994), as summarised in Table 6.1, which compared routine automated monitoring in a stroke unit versus conventional stroke unit care. Three of the trials (517 patients) were randomised (Davis et al., 2000; Sulter et al., 2003; Ilmavirta, 1994). Individual patient data were obtained from two trials (Davis et al., 2000; Sulter et al., 2003) and two provided published and unpublished grouped data (Cavallini et al., 2003; Ilmavirta, 1994).

Physiological complications recorded by trials included increased or decreased glucose, blood and heart rate abnormalities, pyrexia and hypoxia. The main outcomes of interest were stroke progression (SP), death, death or dependency, death or need for institutional care, and length of stay. Subgroups were defined

Table 6.1: Summary of trials included in analysis.

Trial	Type[*]	N	Status	Data[†]
Davis et al. (2000)	RCT	258	Unpublished	IPD
Sulter et al. (2003)	RCT	54	Published 2003	IPD
Cavallini et al. (2003)	CCT	268	Published 2003	AD
Ilmavirta (1994)	RCT	213	Published 1993	AD

[*]Type: RCT = randomised control trial; CCT = controlled clinical trial
[†]Data: IPD = individual patient data; AD = aggregate (grouped) data

by severity, age, risk group and comorbidities (atrial fibrillation, diabetes, and heart failure).

For two studies (Davis et al., 2000; Sulter et al., 2003), SP was defined as any deterioration in power of arm, leg, speech, consciousness level, or death between admission and day three. Deterioration was defined as a rise in National Institutes of Health Stroke Scale (NIHSS) by at least 1 point, or at least a 2 point fall in Scandinavian Stroke Scale (SSS). One study (Cavallini et al., 2003) used a local definition while in another (Ilmavirta, 1994), data were limited and SP was defined as becoming unconscious between admission and the worst point of the first week of hospital stay.

Death, dependency and institutional care were recorded at three months after stroke for three of the trials (Davis et al., 2000; Sulter et al., 2003; Ilmavirta, 1994) while outcomes were recorded at discharge in another trial (Cavallini et al., 2003). The criteria for independence were equivalent to a modified Rankin score of 0 - 2, or a Barthel Index of 19 - 20. Those with these scores missing who were in institutional care were assumed to be dependent.

In the four trials, age was classified as those under 70 versus those aged 70 or more. Table 6.2 shows how severity was defined in each of the trials. Chapter 3

Table 6.2: Definition of severity in each included trial.

Severity	Mild	Moderate	Severe
Sulter et al. (2003) (NIHSS)	0 - 5	6 - 13	>13
Davis et al. (2000) (SSS)	>42	25 - 42	<25
Cavallini et al. (2003) (Barthel)	10 - 20	3 - 9	0 - 2
Ilmavirta (1994) (patient functionality)	Mobile	Not mobile/ conscious	Reduced consciousness

Table 6.3: Definition of risk groups using age and OCSP classification.

	LACS/POCS	TACS/PACS
Age < 70	low	medium
Age ≥ 70	medium	high

shows the equivalence of several methods of categorising severity, therefore, it is assumed the severity categorisations used in this analysis were also equivalent. In the current analysis, severity was dichotomised by combining the moderate group with the severe group and comparing this to the mild group.

Risk group is based on a combination of the Oxfordshire Community Stroke Project (OCSP) classification and age (Table 6.3). Older patients also tend to have more severe strokes and Bamford et al. (1991) show that patients with an OCSP classification of partial anterior circulation syndrome (PACS) or total anterior circulation syndrome (TACS) generally have more severe strokes than those with lacunar syndrome (LACS) or posterior circulation syndrome (POCS). Risk group was therefore calculated for two of the trials (Davis et al., 2000; Sulter et al., 2003) and in the current analysis risk group was dichotomised by combining the medium and high risk groups and comparing this to the low risk group.

6.2.2 Statistical methods

Complications and main outcomes data were analysed using hierarchical Bayesian models in WinBUGS. For the complications data, absolute risk difference was analysed using the same model as described in Section 5.2.2, whereas the main outcomes were analysed as both odds ratios (using the model described in Section 1.4.3) and absolute risk differences, each with 95% credible intervals (CrI).

Both models used an Inverse-Gamma distribution for the treatment effect variance. In addition, length of hospital stay was compared between conventional acute stroke unit care and routine monitoring using the weighted mean difference calculated by the Cochrane Collaboration statistical software, Revman (Cochrane Collaboration, 2003).

In order to analyse subgroups, the aggregated data can be expanded to form individual patient data. For example, if the number of subjects who died within each severity subgroup of each treatment is known then this information could be used to extrapolate to the individual patient data. The data are stacked with an extra variable denoting the treatment of each patient allowing the data to be modelled more easily. Therefore, a slightly different modelling approach, as described by Sutton et al. (2008), is adopted.

Let Y_{ij} denote the outcome of subject j in trial i, which is modelled directly as

$$Y_{ij} \sim \text{Bernoulli}(p_{ij}),$$

where p_{ij} is the probability of an event, which is then modelled as

$$\text{logit}(p_{ij}) = \mu_i + \delta_i t_{ij} + \beta_j^0 x_{ij} + \beta x_{ij} t_{ij}. \tag{6.1}$$

As with previous models, let the prior for the baseline event rates be $\mu_i \sim$ Normal$(0, 10^6)$, and let $\delta_i \sim$ Normal(d, τ^2) be the random treatment effects in trial i where $d \sim$ Normal$(0, 10^6)$ is the prior for the mean treatment effects and the prior for the between-trial treatment variance is

$$\frac{1}{\tau^2} \sim \text{Gamma}(0.001, 0.001).$$

The additional terms in equation (6.1) represent the inclusion of the subgroup effects. The first term, $\beta_j^0 x_{ij}$, denotes the main (random) effect of the covariate x_{ij} with coefficient $\beta_j^0 \sim \text{Normal}(b, \tau_b^2)$, where $b \sim \text{Normal}(0, 10^6)$ gives the prior for the mean covariate effect on the outcome, and the variance is distributed as an Inverse-Gamma

$$\frac{1}{\tau_b^2} \sim \text{Gamma}(0.001, 0.001).$$

Finally, the second additional term, $\beta x_{ij} t_{ij}$, denotes the interaction between the covariate x_{ij} and treatment t_{ij}. The coefficient of this (fixed-effect) term, with prior $\beta \sim \text{Normal}(0, 10^6)$, determines how the covariate changes the effect of treatment.

The WinBUGS code for this model is given Appendix C.

6.3 Results

All trials recorded physiological complications of interest and were able to provide information on the main outcomes. Three studies (Davis et al., 2000; Sulter et al., 2003; Ilmavirta, 1994) provided data on age of patient for the outcomes of death and death or institutional care and two of these (Davis et al., 2000; Sulter et al., 2003) provided additional data on age of patient for outcomes of death or dependency and SP. All four trials provided data on severity for the outcomes of death, death or dependency and death or institutional care and two (Davis et al., 2000; Sulter et al., 2003) also provided data on SP. Risk group data were obtained from two studies (Davis et al., 2000; Sulter et al., 2003) which also provided data on atrial fibrillation, diabetes and heart failure for all main outcomes.

6.3.1 Identified complications

Data were available on recorded physiological complications in the monitoring and conventional care groups, given as number and percentage of complications recorded in each group. Table 6.4 shows there were no significant differences in absolute risk of physiological complications recorded between monitoring and conventional care groups.

Table 6.4: Frequency and comparison of recorded physiological complications in monitoring units versus conventional care.

Complication	Number of events		Absolute risk
	Monitoring (%)	Control (%)	difference (95% CrI)
Pyrexia	89 (22.7)	87 (21.8)	0.02 (-0.17, 0.21)
Hypoxia	19 (4.9)	13 (3.3)	0.02 (-0.52, 0.56)
Glucose*	36 (9.2)	48 (12.0)	-0.02 (-0.11, 0.09)
Blood pressure*	81 (20.7)	51 (12.8)	0.12 (-0.20, 0.49)
Heart rate*	84 (21.5)	70 (17.5)	0.05 (-0.11, 0.23)

Results are presented as median posterior absolute risk difference with 95% credible interval of complications in monitoring units versus conventional care.

*Either increased or decreased.

6.3.2 Main outcomes

The results for main outcomes are summarised in Figure 6.1. There were no significant differences in odds of death (odds ratio 0.85; 95% credible interval 0.27 to 1.67), death or dependency (0.61; 0.17 to 1.98), or death or institutional care (0.92; 0.61 to 1.33) between conventional stroke units and the monitoring units. There was a non-significant tendency towards a reduction in SP in the monitoring groups (0.70; 0.43 to 1.16) compared with conventional care. However, there were no significant absolute differences between the two treatment groups for
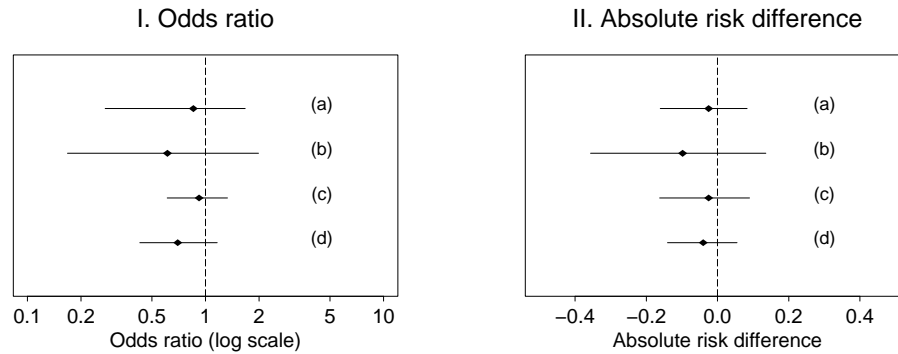
Figure 6.1: Comparison of main outcomes in acute monitoring units versus conventional care. Results are presented in odds ratio (I.) and absolute risk difference (II.). Main outcomes are: (a) death; (b) death or dependency; (c) death or institutional care; (d) stroke progression. Results are presented as median ORs or ARDs (shaded diamond) in stroke units versus conventional care, with corresponding 95% credible intervals represented by the line. Odds ratios are plotted on the natural log-scale.

these outcomes. There was a small reduction in length of stay in the monitoring group (weighted mean difference: 4.5 days; CI: 0.08 to 8.93), however, there was heterogeneity between trials ($p = 0.02$).

6.3.3 Subgroups: age, stroke severity and risk group

There were no significant differences in the odds of outcome between monitoring and conventional care groups for each subgroup of patients. However, Figure 6.2 shows there was a significantly greater reduction in odds of SP in the monitoring group for those who had mild strokes compared to those who had moderate or severe strokes (mild: 0.12, 0.01 to 1.46; moderate or severe: 0.90, 0.10 to 8.65; ratio: 7.29, 1.46 to 62.05). There was also a tendency towards a greater reduction in odds of death or dependency in the monitoring group for those

Age
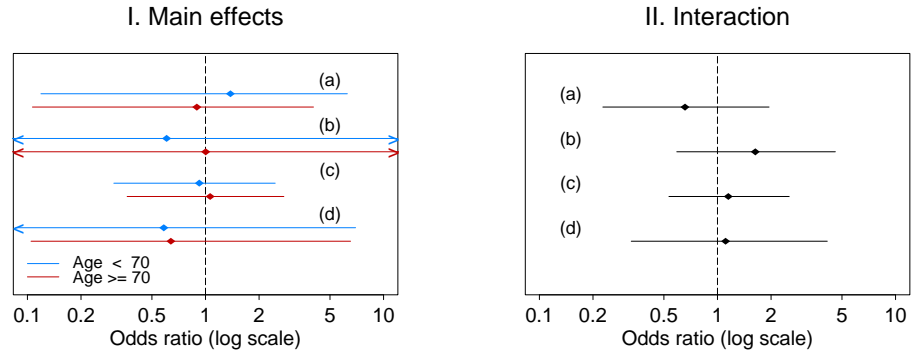
Severity
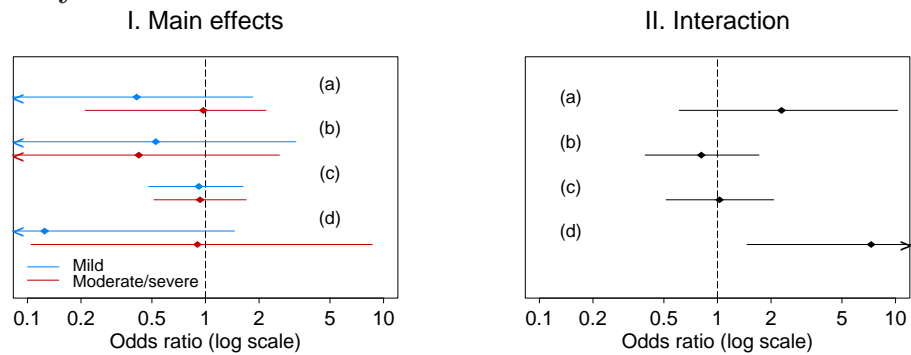
Risk group
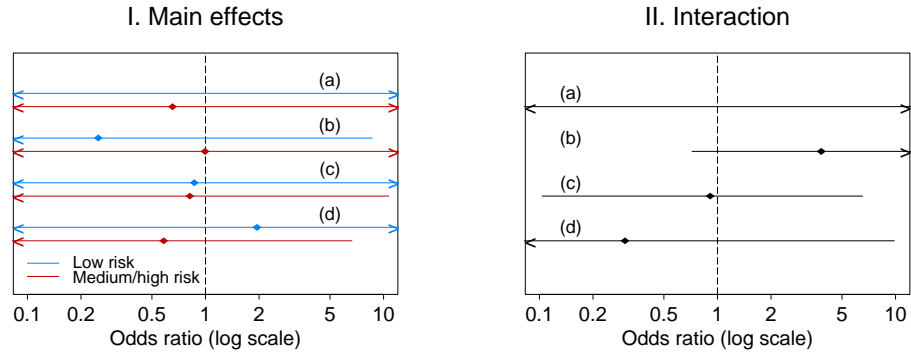
Figure 6.2: Comparison of main outcomes in monitoring units versus conventional care for each covariate subgroup (I.) and interaction between severity and treatment group (II.) where comparisons are: age < 70 vs ≥ 70; severity moderate or severe vs mild; risk group medium or high vs. low. Main outcomes are: (a) death; (b) death or dependency; (c) death or institutional care; and (d) stroke progression. Odds ratios presented as in Figure 6.1.

classed as low risk compared to those classed as medium or high risk (low: 0.25, 0.002 to 8.64; medium or high: 0.99, 0.01 to 29.46; ratio: 3.82, 0.72 to 24.7). The credible intervals are extremely wide due to lack of data, resulting in considerable imprecision in the results.

6.3.4 Subgroups: comorbidities

Figure 6.3 shows there were no significant differences between monitoring and conventional care groups in each comorbidity subgroup. However, there was a significantly greater reduction in the odds of death in the monitoring group for those who did not suffer heart failure compared to those who did suffer heart failure (ratio - heart failure/no heart failure: 9.38, 1.68 to 66.69).

6.4 Discussion

This review analysed the added impact of routine automated monitoring on a background of organised stroke unit care. There were no significant increases in the number of recorded physiological complications in the monitoring unit. However, routine monitoring may reduce the risk of stroke progression and there was also a slight reduction in length of hospital stay in the monitoring group. From these analyses, it was also found that: the reduction in risk of stroke progression in the monitoring group was greater in mild stroke patients than in moderate or severe patients; the reduction in death or dependency in the monitoring group was greatest in those patients classed as low risk; and the reduction in death was greatest in those who did not suffer heart failure. However, caution is needed when interpreting these conclusions since there was little data

Atrial fibrillation
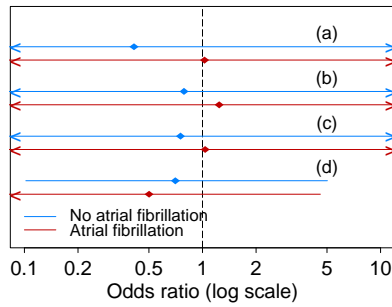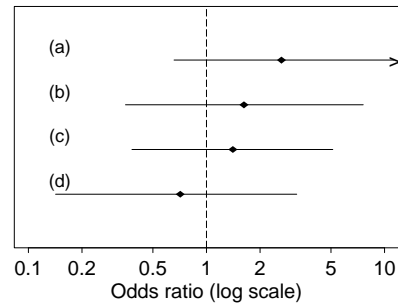
Diabetes
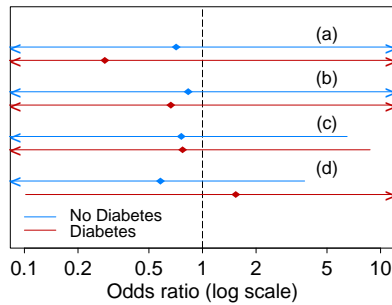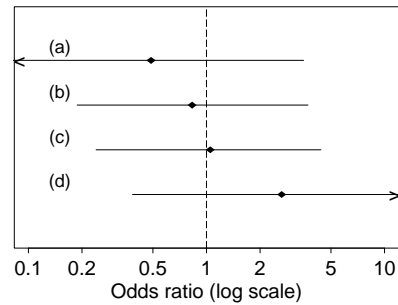
Heart failure
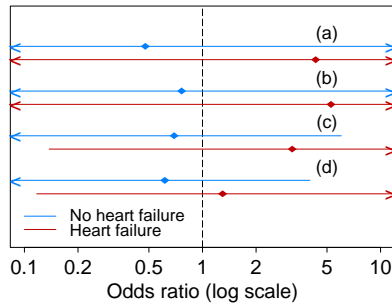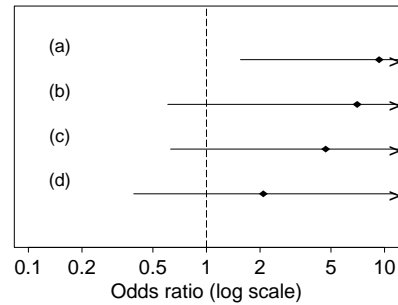
Figure 6.3: Comparison of main outcomes in monitoring units versus conventional care for each covariate subgroup (I.) and interaction between severity and treatment group (II.) where comparisons were comorbidity vs. no comorbidity. Main outcomes are: (a) death; (b) death or dependency; (c) death or institutional care; and (d) stroke progression. Odds ratios presented as in Figure 6.1.

resulting in wide credible intervals.

All trials included in this study were randomised controlled trials or controlled clinical trials. Individual patient data was obtained for two of the trials and detailed grouped data, published and unpublished, were available for the others allowing the estimation of possible subgroup effects.

Given that there are a small number of trials used in this analysis and a small number of events within each trial, especially when analysing subgroups, the type of model chosen to analyse this data has several benefits. Outcomes were modelled directly as the Bernoulli distribution, which is advantageous here since pooling individual trial effects sizes as odds ratios can often require the assumption of normality that may not always be valid. Additionally, the direct model allows zero events to be modelled without the need for continuity correction. Lastly, a random-effects model allows the trial specific effects to be different from each other but assumes that they are from a common distribution, in this case the Normal distribution. In other words, it assumes that all trials are similar but not identical. In contrast, a fixed-effects model would assume that all trials are estimating exactly the same treatment effect.

However, given these strengths a number of weaknesses must also be acknowledged. Firstly, there are only four trials in total in these analyses, and for some analyses, data were only available from two trials. This means that there may not be enough data to estimate accurately treatment or subgroup effects. This is reflected in the relatively large credible intervals and large heterogeneity between trials. Additionally, individual patient data could not be obtained for all studies, thus limiting the analysis to two trials when analysing co-morbidities. Secondly, not all trials used the same definitions of outcome, in particular, stroke

progression. Thirdly, the findings did not reflect all the initial beliefs as it would appear from this analysis that patients with mild strokes appear to benefit more from routine monitoring than moderate or severe patients. However, a possible explanation may be that, since mild patients are likely to have a better prognosis than moderate or severe patients, this advantage may increase in a monitoring setting since any complications they suffer may be reduced, and may therefore also reduce the risk of stroke progression. Finally, there was significant heterogeneity in length of stay among trials and one trial (Cavallini et al., 2003) was methodologically weaker than the other trials. However, the lack of data for some outcomes also meant the lack of potential for sensitivity analysis and so the trial with methodological limitations (Cavallini et al., 2003) was not excluded from any analyses.

Despite the uncertainties in this study, the findings highlight the potential importance of routine automated monitoring in reducing risk of stroke progression and reducing length of hospital stay. Since the magnitude of benefits of routine automated physiological monitoring may be small, a large multi-centred trial may be required to clarify the uncertainties found in this meta-analysis and identify if the added costs and efforts of such a policy are justified by better patient recovery. The trial should be sufficiently large in order to examine appropriately which subgroups of patients and subgroups by co-morbidity may benefit most from routine automated monitoring.

Chapter 7

Collapsed and overlapping covariate categories

This Chapter introduces a network meta-analysis model with covariate effects, allowing the inclusion of all trials either with or without covariate data. This model allows direct and indirect information to be used to give a more powerful estimate of differences between treatment services while also accounting for covariates. This analysis assesses the possible bias in treatment effect estimation (aggregation, or ecological bias) when covariates are not accounted for in the model. Given the results from Chapter 3, this Chapter assumes that the categorisation of severity from different measures of stroke severity are equivalent and that most of the prognostic information from the full scales is retained.

7.1   Background

Meta-analysis, defined as the statistical analysis of a large collection of analytic results for the purpose of integrating the findings (Dickersin and Berlin, 1992), attempts to combine results across studies in order to gain statistical power and to strengthen the evidence about possible treatment effects, and in adequately powered studies to find out more about subgroups and possible interactions. Several meta-analysis techniques have been designed to incorporate most, if not all, of the information available for analysis. This includes accommodating more than one active treatment of interest, for example indirect comparisons (Lu and Ades, 2006), or combining individual patient data (IPD), covariate data, and aggregated data (Sutton et al., 2008).

Often covariate data are also available, which can be taken into account. Generally, covariate data which are most likely to be available in a meta-analysis context will be in categorical form, for instance sex, or old age versus young. Some trials may provide information on treatment effects within each cell of the covariate classification, while in other trials only marginal data (that is, information on one covariate aggregated over the another) for one or more of the covariates may be available, or no covariate breakdown data at all. Previously, this covariate data has often been ignored or analysed in such a way that does not permit estimation of how these covariates are related to outcome, for example in a stratified analysis where each stratum is treated as a separate trial.

Meta-analysis of randomised trials based on aggregated data is, however, vulnerable to forms of aggregation, or ecological bias (Rothman and Greenland,

1998). Rothman and Greenland (1998) define ecological bias as 'the failure of expected ecological effect estimates to reflect the biological effect at the individual level'. The conditions for a covariate to be a confounder are different in individual and aggregated data. For a covariate to be a confounder at the individual level, the outcome and covariate must be associated. Covariates which influence the outcome can give rise to ecological bias even though, due to the randomisation, they are not associated with the treatment. This type of bias will occur in only non-linear models (Rothman and Greenland, 1998). However, non-linear models, for example logistic regression models, are by far the most commonly used in meta-analysis of binary outcome data. This ecological bias can be shown using simple calculations.

It is assumed that the treatment effects in two covariate strata are equal and the effect of a covariate is equal in both treatment and control groups. It is also assumed that the treatment and covariate effects are additive on the logit scale. Figure 7.1 shows that the bias in the treatment effect increases as the absolute covariate effect size increases. The size of the bias is more dramatic as the treatment effect size in each stratum increases. Therefore, the bias will increase for larger covariate effects and larger within-stratum treatment effects, suggesting that it may be important to account for the effect of covariates when performing meta-analyses, particularly when both treatment effects and covariate effects are strong. However, while this can be easily done when treatment effects are reported within covariate categories, it raises questions about trials where treatment effects are pooled over covariate levels.

The objective of this Chapter is to develop and illustrate methods for controlling of ecological bias where trials fail to report treatment effects within covariate
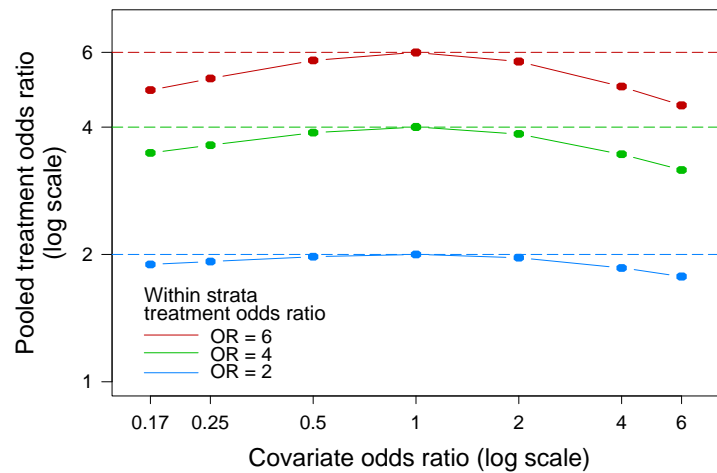
Figure 7.1: Scatterplot showing the effect of covariate odds ratio and within strata treatment odds ratio on the pooled treatment odds ratio for within stratum odds ratios of 2, 4 and 6. Note that for this plot the proportion of survivors in the treatment group of stratum one was chosen as 0.8.

categories, but instead "collapse" over categories. Treatment effects will be estimated controlling for ecological bias by using the data available within the trials to impute the distribution of covariates in trials where this is not reported. These methods are applied in a mixed treatment comparison evidence synthesis (Lu and Ades, 2006; Higgins and Whitehead, 1996). Indirect treatment comparisons emerge when treatments of interest are not directly compared to each other, but are compared with some other common treatment. For example, suppose trials T_{AB} compare treatment A versus B, while trials T_{BC} compare treatments B versus C. It may be of interest to compare treatments A and C, but due to ethical issues this may not be directly possible. However, the comparison of A versus B and B versus C can provide information indirectly on the comparison of A versus

C. Where more than 3 treatments are being compared, these data structures have been termed mixed treatment comparisons (MTC) (Lu and Ades, 2006).

The model controlling for ecological bias is compared to two simpler approaches: in the first the analysis is stratified as finely as the available data allows, treating each stratum as a separate trial; in the second only fully aggregated data are used, ignoring all covariate information. In Section 7.2 a mixed treatment comparison model is described that includes two covariate effects, as it would be if the treatment effect data were available within each covariate category. Also, described are variant models for random covariate effects and interactions between covariates, and methods for model choice. Section 7.3 describes the extensions required for the model in Section 7.2.1 to obtain the collapsed categories models (stratified and aggregated), while Section 7.4 gives an example of these models applied to trials comparing various systems of acute stroke care. Section 7.5 concludes with a discussion of the potential impact of the proposed models.

In common with many of the recent developments in evidence synthesis for complex data structures (Ades and Sutton, 2006), a Bayesian Markov chain Monte Carlo framework with vague priors was adopted, and computed using WinBUGS (Lunn et al., 2000).

7.2 Mixed treatment comparisons model with covariate effects

7.2.1 Model description

Covariate information is most likely to be available in categorical form. For clarity in the following notation, two covariates, each having up to three levels, are allowed. This is likely to be sufficient in most contexts, but the model may easily be extended to allow a greater number of levels per covariate.

In the basic covariate model, fixed covariate effects and random treatment effects are assumed. The fixed effect assumes the covariate effect from each study included in the analysis estimates the same quantity with any deviations due to random sampling variability, whereas the random effect allows the treatment effects to differ across studies but assumes they are drawn from a common distribution of effect sizes.

The probability of an event in cell k of Table 7.1 for the treatment in arm j

Table 7.1: Summary of format of covariate data

		Covariate 2			Covariate 1
		1	2	3	marginal
	1	[1]	[2]	[3]	[A]
Covariate 1	2	[4]	[5]	[6]	[B]
	3	[7]	[8]	[9]	[C]
Covariate 2 marginal		[T]	[U]	[V]	[Z]

$(j = 2, ..., J)$ of trial i is given by Λ_{ijk}, such that

$$\text{logit}(\Lambda_{ijk}) = \lambda_{ik} + \delta_{ij}, \tag{7.1}$$

where δ_{ij} are the estimated treatment effects in arm J, described in more detail below, and

$$
\begin{aligned}
\lambda_{i1} &= \mu_i, \\
\lambda_{i2} &= \mu_i + \alpha_1, & (7.2) \\
\lambda_{i3} &= \mu_i + \alpha_2, & (7.3) \\
\lambda_{i4} &= \mu_i + \beta_1, & (7.4) \\
\lambda_{i5} &= \mu_i + \alpha_1 + \beta_1, & (7.5) \\
\lambda_{i6} &= \mu_i + \alpha_2 + \beta_1, & (7.6) \\
\lambda_{i7} &= \mu_i + \beta_2, & (7.7) \\
\lambda_{i8} &= \mu_i + \alpha_1 + \beta_2, & (7.8) \\
\lambda_{i9} &= \mu_i + \alpha_2 + \beta_2. & (7.9)
\end{aligned}
$$

The α_1 and α_2 are the additional effects of levels 2 and 3, respectively, of covariate 1 compared to the baseline category of level 1, while β_1 and β_2 are the equivalent for covariate 2. All covariate effects are given vague priors,

$$\alpha_1, \alpha_2, \beta_1, \beta_2 \sim \text{Normal}(0, 10^3). \tag{7.10}$$

The $\mu_i \sim \text{Normal}(0, 10^3)$ are the baseline event rates. This refers to observations in level 1 of both covariates for treatment 1. This formulation only allows the

modelling of the relative effect of the covariates on the baseline, not the absolute effects.

Therefore, given a trial with full information of both covariates (that is, those trials with information on cells [1] to [9] of Table 7.1), the numbers of events are

$$r_{ijk} \sim \text{Binomial}(\Lambda_{ijk}, n_{ijk}).$$

In the mixed treatment effect model it is easiest to begin by considering a situation where every trial might include all 6 treatments. Then the true effects of treatments 2 to 6 relative to treatment 1 are distributed as

$$\begin{bmatrix} \delta_{i2(1)} \\ \delta_{i3(1)} \\ \delta_{i4(1)} \\ \delta_{i5(1)} \\ \delta_{i6(1)} \end{bmatrix} \sim \text{Normal} \left(\begin{bmatrix} d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \end{bmatrix}, \begin{bmatrix} \tau^2 & 0.5\tau^2 & 0.5\tau^2 & 0.5\tau^2 & 0.5\tau^2 \\ 0.5\tau^2 & \tau^2 & 0.5\tau^2 & 0.5\tau^2 & 0.5\tau^2 \\ 0.5\tau^2 & 0.5\tau^2 & \tau^2 & 0.5\tau^2 & 0.5\tau^2 \\ 0.5\tau^2 & 0.5\tau^2 & 0.5\tau^2 & \tau^2 & 0.5\tau^2 \\ 0.5\tau^2 & 0.5\tau^2 & 0.5\tau^2 & 0.5\tau^2 & \tau^2 \end{bmatrix} \right),$$

(7.11)

The priors for the means of the individual treatment effects, d_t, are given by

$$d_1 = 0,$$

$$d_2, ..., d_T \sim \text{Normal}(0, 10^3),$$

where T is the total number of treatments and the prior for the variance parameter is given as

$$\frac{1}{\tau^2} \sim \text{Gamma}(0.01, 0.01).$$

It is then assumed that the selection of which treatments are actually included in each trial is independent of the results.

7.2.2   Random covariates and covariate interactions

The need for fixed or random effects on the covariate, as well as the possibility of an interaction existing between the covariates, may be investigated. The following four models were compared:

1. fixed covariate effects (as described in Section 7.2.1),

2. random covariate effects,

3. fixed covariate effects, with additional interaction term between covariates,

4. random covariate effects, with additional interaction term between covariates.

For the random covariate model (model 2), equations (7.2)-(7.9) of the fixed covariate model described in Section 7.2.1 are changed to give

$$
\begin{aligned}
\lambda_{i2} &= \mu_i + \alpha_{i1}, \\
\lambda_{i3} &= \mu_i + \alpha_{i2}, \\
\lambda_{i4} &= \mu_i + \beta_{i1}, \\
\lambda_{i5} &= \mu_i + \alpha_{i1} + \beta_{i1}, \\
\lambda_{i6} &= \mu_i + \alpha_{i2} + \beta_{i1}, \\
\lambda_{i7} &= \mu_i + \beta_{i2}, \\
\lambda_{i8} &= \mu_i + \alpha_{i1} + \beta_{i2},
\end{aligned}
$$

$$\lambda_{i9} \;=\; \mu_i + \alpha_{i2} + \beta_{i2},$$

with priors for the covariate effects given in equation (7.10) becoming

$$\alpha_{i1} \sim \text{Normal}(\alpha_1, \sigma^2),$$

$$\alpha_{i2} \sim \text{Normal}(\alpha_2, \sigma^2),$$

$$\beta_{i1} \sim \text{Normal}(\beta_1, \sigma^2),$$

$$\beta_{i2} \sim \text{Normal}(\beta_2, \sigma^2),$$

with vague priors on the hyper-parameters

$$\alpha_1, \alpha_2, \beta_1, \beta_2 \sim \text{Normal}(0, 10^3),$$

and

$$\frac{1}{\sigma^2} \sim \text{Gamma}(0.01, 0.01).$$

The interaction effects (model 3) were included by adjusting equations (7.5), (7.6), (7.8) and (7.9) to include the additional interaction terms

$$\lambda_{i5} \;=\; \mu_i + \alpha_1 + \beta_1 + \gamma_1,$$

$$\lambda_{i6} \;=\; \mu_i + \alpha_2 + \beta_1 + \gamma_2,$$

$$\lambda_{i8} \;=\; \mu_i + \alpha_1 + \beta_2 + \gamma_3,$$

$$\lambda_{i9} \;=\; \mu_i + \alpha_2 + \beta_2 + \gamma_4,$$

where

$$\gamma_1, \gamma_2, \gamma_3, \gamma_4 \sim \text{Normal}(0, 10^3).$$

Finally, model 4 includes the random effects of the covariates as in model 2 and also random interaction effects where equations (7.5), (7.6), (7.8) and (7.9) are now

$$\lambda_{i5} = \mu_i + \alpha_{i1} + \beta_{i1} + \gamma_{i1},$$

$$\lambda_{i6} = \mu_i + \alpha_{i2} + \beta_{i1} + \gamma_{i2},$$

$$\lambda_{i8} = \mu_i + \alpha_{i1} + \beta_{i2} + \gamma_{i3},$$

$$\lambda_{i9} = \mu_i + \alpha_{i2} + \beta_{i2} + \gamma_{i4},$$

with interaction priors given as

$$\gamma_{i1} \sim \text{Normal}(\gamma_1, \sigma^2),$$

$$\gamma_{i2} \sim \text{Normal}(\gamma_2, \sigma^2),$$

$$\gamma_{i3} \sim \text{Normal}(\gamma_3, \sigma^2),$$

$$\gamma_{i4} \sim \text{Normal}(\gamma_4, \sigma^2),$$

where

$$\gamma_1, \gamma_2, \gamma_3, \gamma_4 \sim \text{Normal}(0, 10^3).$$

Model choice was determined by the Deviance Information Criterion (DIC) (Spiegelhalter et al., 1998) and whether the posterior credible interval for inter-action effects contained 0. The DIC is the sum of the 'fit' (that is, the posterior expectation of the deviance denoted by $\bar{D}$) and the complexity (the effective number of parameters, p_D) of a model. The terms making up the DIC therefore represent a goodness-of-fit term and a term representing a penalty for increasing

model complexity.

7.3 Collapsed categories model

Up to this point, treatment effects and covariate effects were modelled in a standard logistic regression framework, in which numerators and denominators for treatment and control were available for each set of covariate values. However, some trials may only have marginal data for one or both of the covariates and often trials have no covariate breakdown at all. Those trials with collapsed and overlapping categories can also be included into the covariate model and their structure can be expressed in terms of the parameters introduced thus far.

The approach estimates the proportions of the trial population, π_{ik}, for trial i in each cell k of Table 7.1. These proportions are assumed to be samples from a common Dirichlet distribution,

$$\pi_{i,1}, ..., \pi_{i,9} \sim \text{Dirichlet}(\kappa_1, ..., \kappa_9),$$

and the hyper-parameters, $\kappa_1, ..., \kappa_9$, are given vague priors and estimated from the data. To ensure positive values greater than one, which would represent a uniform prior, the $\log(\kappa_1, ..., \kappa_9)$ were assigned Half Normal priors truncated at zero:

$$\log(\kappa_j) \sim \text{Normal}(0, 10^2) \text{ for } \log(\kappa_j) > 0.$$

7.3.1   Incorporating data on collapsed and overlapping categories

7.3.1.1   Full covariate data

The full covariate model is described in Sections 7.2.1 and 7.2.2. The cell sizes when full data are available are represented by

$$n_{ij(1:9)} \sim \text{Multinomial}(\pi_{i(1:9)}, N_{ij}),$$

where $N_{ij} = \sum_{k=1}^{9} n_{ijk}$, is the total number of patients in all k cells in treatment arm j of trial i.

7.3.1.2   Data collapsed over either covariate

Cells [A] to [C] of Table 7.1 correspond to those trials with marginal data for covariate 1 and these give the following probabilities of event

$$
\begin{aligned}
\theta_{ijA} &= \frac{(\pi_{i1}\Lambda_{ij1} + \pi_{i2}\Lambda_{ij2} + \pi_{i3}\Lambda_{ij3})}{\pi_{i(\alpha_0)}}, \\
\theta_{ijB} &= \frac{(\pi_{i4}\Lambda_{ij4} + \pi_{i5}\Lambda_{ij5} + \pi_{i6}\Lambda_{ij6})}{\pi_{i(\alpha_1)}}, \\
\theta_{ijC} &= \frac{(\pi_{i7}\Lambda_{ij7} + \pi_{i8}\Lambda_{ij8} + \pi_{i9}\Lambda_{ij9})}{\pi_{i(\alpha_2)}},
\end{aligned}
$$

where

$$
\begin{aligned}
\pi_{i(\alpha_0)} &= \pi_{i1} + \pi_{i2} + \pi_{i3}, \\
\pi_{i(\alpha_1)} &= \pi_{i4} + \pi_{i5} + \pi_{i6}, \\
\pi_{i(\alpha_2)} &= \pi_{i7} + \pi_{i8} + \pi_{i9}.
\end{aligned}
$$

The number of events in each level of the covariate are therefore

$$r_{ijk} \sim \text{Binomial}(\theta_{ijk}, n_{ijk}),$$

where $k = A, B, C$ and the number of patients in each cell has a multinomial distribution

$$(n_{ijA}, n_{ijB}, n_{ijC}) \sim \text{Multinomial}((\pi_{i(\alpha_0)}, \pi_{i(\alpha_1)}, \pi_{i(\alpha_2)}), N_{ij}),$$

Similarly, cells [T] to [V] of Table 7.1 correspond to those trials with marginal data for covariate 2 with the following probabilities of event

$$\begin{aligned}
\theta_{ijT} &= \frac{(\pi_{i1}\Lambda_{ij1} + \pi_{i4}\Lambda_{ij4} + \pi_{i7}\Lambda_{ij7})}{\pi_{i(\beta_0)}}, \\
\theta_{ijU} &= \frac{(\pi_{i2}\Lambda_{ij2} + \pi_{i5}\Lambda_{ij5} + \pi_{i8}\Lambda_{ij8})}{\pi_{i(\beta_1)}}, \\
\theta_{ijV} &= \frac{(\pi_{i3}\Lambda_{ij3} + \pi_{i6}\Lambda_{ij6} + \pi_{i9}\Lambda_{ij9})}{\pi_{i(\beta_2)}},
\end{aligned}$$

where

$$\begin{aligned}
\pi_{i(\beta_0)} &= \pi_{i1} + \pi_{i4} + \pi_{i7}, \\
\pi_{i(\beta_1)} &= \pi_{i2} + \pi_{i5} + \pi_{i8}, \\
\pi_{i(\beta_2)} &= \pi_{i3} + \pi_{i6} + \pi_{i9}.
\end{aligned}$$

such that the number of events in each level are given as

$$r_{ijk} \sim \text{Binomial}(\theta_{ijk}, n_{ijk}),$$

where $k = T, U, V$ and the number of patients in each cell has a multinomial distribution

$$(n_{ijT}, n_{ijU}, n_{ijV}) \sim \text{Multinomial}((\pi_{i(\beta_0)}, \pi_{i(\beta_1)}, \pi_{i(\beta_2)}), N_{ij}).$$

7.3.1.3 Collapsed and overlapping data for covariates 1 and 2

Some trials may have the marginal data for covariate 1 and the marginal data for covariate 2. In considering the information on treatment effects it is important to avoid using information more than once. For example, if binomial information on treatment effects was available for covariate 1 (marginals [A], [B], and [C] in Table 7.1) as well as covariate 2 (marginals [U] and [V]), then this is enough information to determine the marginal numerator and denominator for cell [T]. Thus, conditional on the information in the other marginals, [T] provides no further information on treatment effects. The likelihood is therefore specified as

$$r_{ijk} \sim \text{Binomial}(\theta_{ijk}, n_{ijk}),$$

in each cell where $k = A, B, C, U, V$.

The likelihood for the cell frequency information is, however, the product of the likelihoods for the two marginal multinomials given in Section 7.3.1.2. This is because the distribution of the marginal totals for covariate 1 ([A], [B] and [C] in Table 7.1) does not put any constraints on the distribution of the marginals totals of covariate 2 ([T], [U] and [V]).

7.3.1.4   Collapse of data over both covariates

In Table 7.1, cell [Z] represents those trials where there were no data available for either covariate. In this case the number of events are given as

$$r_{ijZ} \sim \text{Binomial}(\theta_{ijZ}, N_{ij}),$$

where the probability of event is given as

$$\theta_{ijZ} = \sum_{k=1}^{9} \pi_{ik}\Lambda_{ijk}.$$

7.3.1.5   Trials on subsets of the patient population

In some trials there may be an absence of data in one or more levels of a covariate. It is assumed here that these zeros in the marginal totals are structural. That is, instead of there simply being no cases within a covariate level, the zeros are due to the inclusion criteria of trials where data on one or more levels of a covariate were not collected. In the case where a trial has structural zeros for a covariate, this trial is not allowed to inform the distribution of the covariate where data has not been collected (that is the π_{ik}) but it can inform the distribution for the other covariate.

7.3.2   Stratified and fully aggregated models

The results from the model described in Sections 7.2 and 7.3.1, which makes maximal use of covariate information, is contrasted with two alternative approaches. The stratified model (where each stratum can be considered a 'separate' trial) and

aggregated model (where no covariate data is used) are technically the same, with the difference between these models lying in the data format used. In the stratified model, any data available for cells [1] to [9] (in Table 7.1) are marginalised, giving data in the cells [A] to [C], [T] to [V], or [Z] if no covariate breakdown were available. Where a trial has marginal data for both covariates, only one set of marginal data is chosen. The choice is determined by the amount of information available for both covariates. The covariate with the least data available across all trials is chosen in order to increase the amount of information for this covariate. This subgroup data is then incorporated by treating each subgroup stratum as if it were a different trial.

For the aggregated model it is assumed there are no subgroup data, and all data are aggregated into one level for each treatment in each trial, equivalent to cell [Z].

In order to analyse these collapsed data structures, equation (7.1) is changed. The numbers of events are assumed to be Binomially distributed with probability p_{ij} such that

$$\text{logit}(p_{ij}) = \mu_i + \delta_{ij},$$

where $\mu_i \sim \text{Normal}(0, 10^3)$ and δ_{ij} is described in equation (7.11).

7.3.3   Parameter estimation

All analyses were carried out in WinBUGS 1.4.1 (Lunn et al., 2000). The multivariate normal distribution for the $\delta_{it(1)}$ was realised as a series of conditional

univariate distributions. The common Dirichlet distribution for the π_{ik} was realised as a series of conditional beta distributions (Gelman et al., 2004). Convergence was assessed using the Brooks-Gelman-Rubin diagnostic tool (Brooks and Gelman, 1998) and considered adequate after 20,000 iterations in all models. For each model, a further sample of 50,000 iterations was then run on which the results are based.

7.4   Example: organised inpatient (stroke unit) care

7.4.1   Data

Data from a collaborative systematic review carried out by the Stroke Unit Trialists Collaboration (Stroke Unit Trialists' Collaboration, 2007) are used. This database is described in Chapter 4. In summary, the review examined whether improving the organisation of inpatient stroke services can bring about improvements in important patient outcomes. A very broad definition of stroke unit care was used and the review included any trial which compared organised (stroke unit) care (defined as a multi-disciplinary team specialising in stroke care) versus the contemporary conventional care such as a general medical ward or less organised form of stroke care. Stroke unit care could include services based in a discrete ward or provided by a mobile stroke team. Outcome data on the number of deaths within each treatment group of each trial was recorded. In addition to this, the number of deaths within subgroups of age and severity were sought. Age was categorised into those less than or equal to 75 versus those greater than

Table 7.2: Number of trials providing subgroup data.

Trials providing data on distribution of severity and age	
Joint age and severity	2
Marginal age and severity	9
Marginal severity only	15
Marginal age only	1
No breakdown	2

Trials providing data on a subset of severity and age	No age recorded	Old only	All ages
No severity recorded	2	-	1
Moderate only	4	-	-
Mild and moderate	1	-	-
Moderate and severe	1	-	-
All severities	9	1	10

75 years of age, while baseline severity was categorised into mild, moderate and severe using the Barthel Index (BI) of activities of daily living (Mahoney and Barthel, 1965), where mild is equivalent to BI 10-20, moderate BI 3-9 and severe BI 0-2.

Altogether, the SUTC provided data for 31 trials. Service comparisons within these 31 trials are summarised in Figure 4.1. The commonest comparison was between comprehensive stroke ward (service 2) and general medical ward (service 6) with eleven trials looking at this comparison. Davis et al. (2000) also provided their data comparing an acute stroke ward system of care with a comprehensive stroke ward system of care. Therefore, data were available from a total of 32 trials. Three trials were excluded as the number of deaths recorded was zero. Table 7.2 summarises the remaining 29 trials. Note there were only 2 trials that provided individual patient data, therefore providing data on the combination of age and severity. Table 7.2 also summarises the number of trials that provided

Table 7.3: Summary of format of data available

		Age		Severity
		Young	Old	marginal
Severity	Mild	[1]	[2]	[A]
	Moderate	[3]	[4]	[B]
	Severe	[5]	[6]	[C]
Age marginal		[T]	[U]	[Z]

data from a subset of the six severity and age cells as shown in Table 7.3. Table 7.3 shows the combination of the covariates where cells [1] to [6] correspond to data for both age and severity, cells [A] to [C] represent severity marginal data only, cells [T] and [U] represent age marginal data and finally cell [Z] represents no breakdown data available for age or severity.

7.4.2 Covariate model choice

Firstly, the covariate models were fitted to the data. The DIC, summarised in Table 7.4, and the presence of substantial interaction effects determined the model choice. The Dbar and pD results for the cell frequency data was very much the same for all models and is not shown in Table 7.4. The DIC in the random covariate model (model 2) is lower than those for the fixed covariate model and the fixed interaction model, indicating that random covariate effects are needed. The posterior credible intervals for the interaction terms in model 4 were not substantial (old*moderate: OR 0.27, 95% CrI 0.002 to 9.09; old*severe: 0.40, 0.003 to 8.17). Therefore, the random covariate model was chosen for further investigation of the effect of collapsing covariate categories. The WinBUGS code for this model is given in Appendix D.

Table 7.4: DIC for the four covariate models

Model		Dbar	Dhat	pD	DIC
1. Fixed covariate effect	r	698.4	639.7	58.7	757.2
	Total	1386.5	1279.7	106.8	1493.3
2. Random covariate effect	r	669.8	600.3	69.5	739.4
	Total	1357.0	1240.6	116.4	1473.5
3. Fixed covariate effect	r	698.1	638.5	59.6	757.7
with interaction	Total	1386.5	1278.4	108.0	1494.5
4. Random covariate effect	r	669.3	599.9	69.4	738.6
with interaction	Total	1356.6	1240.7	115.9	1472.4

7.4.3 Collapsed categories results

The estimated treatment effects obtained from the three modelling strategies (aggregated, stratified and random covariate) are summarised in Figure 7.2. The estimates of treatment effect appear to be consistent in the three models, with the estimates in the aggregated model slightly closer to 1. Although the treatment effect estimates for the covariate model are slightly more variable than the other models, this model also provides additional estimates for the effects of severity and age on outcome. As expected, the odds of death increases for older patients (OR 7.69, 95% CrI 3.41 to 20.03) and also increases as severity increases (moderate vs. mild: 6.83, 3.47 to 14.63; severe vs. mild: 52.04, 26.92 to 128.90). However, there is large variability in these estimates. Note for those trials where there existed marginal covariate information for both age and severity, the age covariate data in these trials was chosen to be included as this covariate had fewer data points available across all trials.
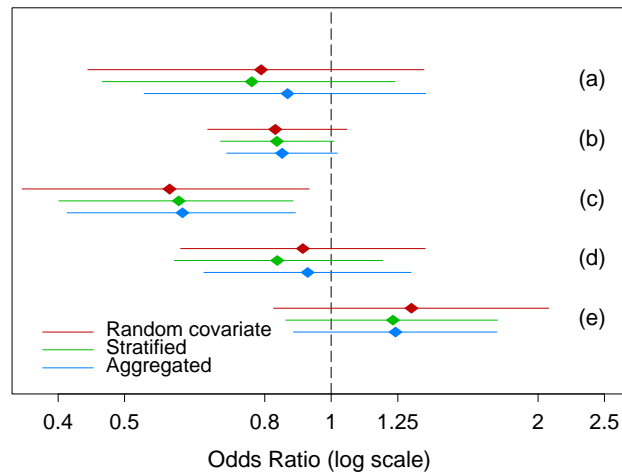
Figure 7.2: Posterior credible intervals for the odds ratios of treatment effect compared to general medical ward for each of the covariate, stratified and aggregate models. Treatments are (a) acute stroke ward; (b) comprehensive stroke ward; (c) rehabilitation stroke ward; (d) mixed rehabilitation ward; and (e) mobile stroke team.

7.5   Discussion

Described here is a mixed treatment comparisons model with covariate effects which incorporates data on collapsed and overlapping categories. The model allows the inclusion of studies with the joint distribution of covariates, those with marginal data for one or two covariates as well as those with no covariate information and provides an estimate of the covariate effects. The estimates of treatment effects are obtained using direct and indirect comparisons giving a more powerful comparison between treatment types.

The covariate model chosen to analyse the SUTC data was the random covariate model without interaction (model 2). Although the random covariate model with interaction (model 4) had a smaller DIC value, the interaction effect

magnitudes were not substantial in this model. Also, since the treatment effect estimates were small and unaffected by the difference in models, the simpler of the two models were chosen.

The treatment effect estimates obtained from the model were compared with those of the stratified covariate model (where each stratum can be considered a 'separate' trial) and a fully aggregated model (where no covariate data are used). Simple calculations show that bias may exist in the treatment effect if covariate effects are not taken into account. Therefore, the aggregate data analysis is expected to be most biased. The stratified model should perform better than the aggregated model but it should also be biased since there is still a high level of aggregation. This means that the estimated treatment odds ratios in the aggregate and stratified models should be closer to one than the covariate model.

In the example using the SUTC data, there appears to be no substantial bias in the estimation of treatment effect between aggregated, stratified and covariate models, although the odds ratios of the treatment effect from the aggregated model compared to the other models, shown in Figure 7.2, are slightly closer to 1. Figure 7.1 showed that a large covariate effect increased the size of the bias in the treatment effect. However, although the size of the covariate effect is fairly large in the example, it may be that there does not appear to be substantial bias because the size of the treatment effect is relatively small. This is also shown in Figure 7.1: when the within stratum treatment effect size is small (the bottom line), the bias is less than when the treatment effect is large.

Several authors have previously investigated the use of collapsed categories models when analysing data. For example Dominici (2000) introduced a multinomial logistic normal hierarchical model that combines contingency tables with

different dimensions in order to investigate the association between air pollution and mortality. This model combines all studies reporting the categorical variables, even when some included studies do not report all variables. However, Dominici assumes a logistic normal family to model the cell probabilities instead of a Dirichlet prior as the environmental data may have had positive correlations between cells.

Salanti et al. (2006) consider synthesising evidence from multiple epidemiological studies. They have also addressed the question of combining data on categorical variables where only one trial recorded the joint effect of all categorical variables and most only recorded one or two covariates of interest. However, unlike the collapsed categories model presented here, their model makes use of external data sources to inform the unobserved variables and they extend their model to include methods for case-control designs.

Rothman and Greenland (1998) also observed that ecological bias does not exclusively occur in meta-analysis of randomised clinical trials. They provide examples of ecological bias occurring in several observational studies with aggregated data.

Finally, the mixed treatment comparisons model can be easily adapted to include additional covariates, or covariates with greater than three levels. The model may also be adapted to include treatment-covariate interactions in addition to interactions between covariates.

Chapter 8

Mixed treatment comparison of different systems of organised inpatient (stroke unit) care

Chapter 4 conducted several comparisons of "more" organised inpatient services with "less" organised services. However, this analysis did not include information on indirect comparisons or the possible effects of age and severity on outcome. Using the model introduced in Chapter 7, a network meta-analysis is performed to explore whether any one system of care is most effective at improving patient outcomes, while also accounting for possible associations of age and severity with patient outcome. Given the results from Chapter 3, this Chapter assumes that the categorisation of severity from different measures of stroke severity are equivalent and that most of the prognostic information from the full scales is retained.

8.1 Background

It has been well established that organised inpatient (stroke unit) care reduces the risk of death, dependency and the need for institutional care following stroke (Stroke Unit Trialists' Collaboration, 2007). However, it has not been established which system of organised care is most effective. Stroke Unit Trialists' Collaboration (2007) show that "more" organised is better than "less" organised care, but generally the analyses cannot differentiate between the different systems of organised care.

However, in these analyses not all of the information available has been used. One of the obvious sources of information often not incorporated into analyses are patient subgroup information. Another source of information not utilised are indirect comparisons of treatment. In most organised inpatient care analyses, the most common comparison (organised care versus general medical ward) uses only trials where an organised service are compared to a general medical ward. However, data may be available from other trials that compare different systems of organised care. These trials hold indirect information about the comparisons between each organised care system and general medical ward; data from all trials can therefore be included in a network meta-analysis where direct and indirect information are used to obtain treatment effect estimates based on the maximum quantity of information.

A mixed treatment comparisons (MTC) meta-analysis including patient subgroup data was used to explore whether any one system of care is most effective in improving patient outcomes while also taking account of associations of age and severity with patient outcome.

8.2 Methods

8.2.1 Data

Data from the updated Stroke Unit Trialists Collaboration systematic review was used to conduct this analysis. In summary, the SUTC database currently contains 31 clinical trials (6936 subjects) and six care systems were identified: semi-intensive acute stroke ward; comprehensive stroke ward; rehabilitation stroke ward; mixed-rehabilitation ward; mobile stroke team; and general medical ward. The network of trials is given in Figure 4.1. Davis et al. (2000) also provided data. This trial compared an acute stroke ward system of care with a comprehensive stroke ward system of care.

The outcomes considered were death, death or dependency, and death or need for institutional care; subgroup data for age and severity were extracted. Age was classified as those under 75 versus those over 75 and severity was classified as moderate versus mild and severe versus mild, where mild, moderate and severe are defined by the Barthel Index in Table 3.1 or an equivalent measure of severity. Trials included in this analysis may have measured severity in a number of different ways. However, Chapter 3 shows the equivalence of several methods of categorising severity. Therefore, it assumed the severity categorisations used in each trial of this analysis are also equivalent.

8.2.2 Statistical methods

Using data from direct and indirect comparisons and Bayesian hierarchical models, odds ratios and 95% credible intervals (CrI) were calculated for outcomes in

each alternative system of care versus general medical wards. Systems of care were also ranked in order of most effective at reducing poor patient outcomes, giving the probability of each ward being the most effective at improving outcomes. The random covariate model without interactions (model 2 as described in Section 7.2) was used. In this model, the treatment and covariate effects are considered random and no interaction exists between the covariates. In this analysis there is one 2-level covariate (age) and one 3-level covariate (severity) as shown in Table 7.3. Residual deviance, as described in Section 5.2.2, was calculated to check model fit.

8.3 Results

Table 8.1 shows the number of trials that provided subgroup data for each of the main outcomes. For the outcome of death or dependency there were only

Table 8.1: Number of trials providing subgroup data for each of the main outcomes.

Subgroup	Death	Death or institutional care	Death or dependency
Joint* age and severity	2	2	2
Marginal† age and severity	9	9	1
Marginal severity only	15	11	21
Marginal age only	1	0	0
No breakdown‡	2	9	4
Total	29	31	28

*Joint distribution of covariates is equivalent to data in cells [1] to [6] of Table 7.3.

†Marginal distribution of covariates is equivalent to data in cells [A] to [C] for severity and [T] to [U] for age in Table 7.3.

‡No breakdown by covariate is equivalent to data in cell [Z] of Table 7.3.

Table 8.2: Number of data points and residual deviance for each outcome.

Outcome	Data points	Residual deviance
Death	364	326
Death or institutional care	348	325
Death or dependency	229	298

three trials that provided information on age. Therefore, due to lack of data, an estimate for the effect of age could not be obtained for this outcome. For all outcomes the fit of the model was adequate as given by residual deviance. Table 8.2 shows that the residual deviance is less than the number of data points.

Figure 8.1 shows that, compared to general medical ward, there were significant reductions in odds of death in rehabilitation stroke ward (OR 0.58; 95% CrI 0.35 to 0.93) and significant reductions in odds of death or institutional care in comprehensive stroke ward (0.72; 0.56 to 0.93). There appeared to be reductions in odds of death in comprehensive stroke ward (0.83; 0.66 to 1.05) and in odds of death or institutional care in rehabilitation stroke ward (0.67; 0.43 to 1.01), however these were not significant. For death or dependency there was a significant reduction of poor outcome in acute stroke ward compared to general medical ward (0.43; 0.21 to 0.84).

Table 8.3 gives the probabilities of each service type being the most effective in reducing the odds of poor outcome. For all three outcomes, general medical ward, mobile stroke team and mixed rehabilitation ward can be considered ineffective systems of care due to the loww probabilities associated with them.

The odds ratios for poor outcome of patient subgroups are given in Table 8.4. Notice that older patients tend to have an increased odds of death and death or institutional care. Also, as the severity of stroke increases, the odds of poor
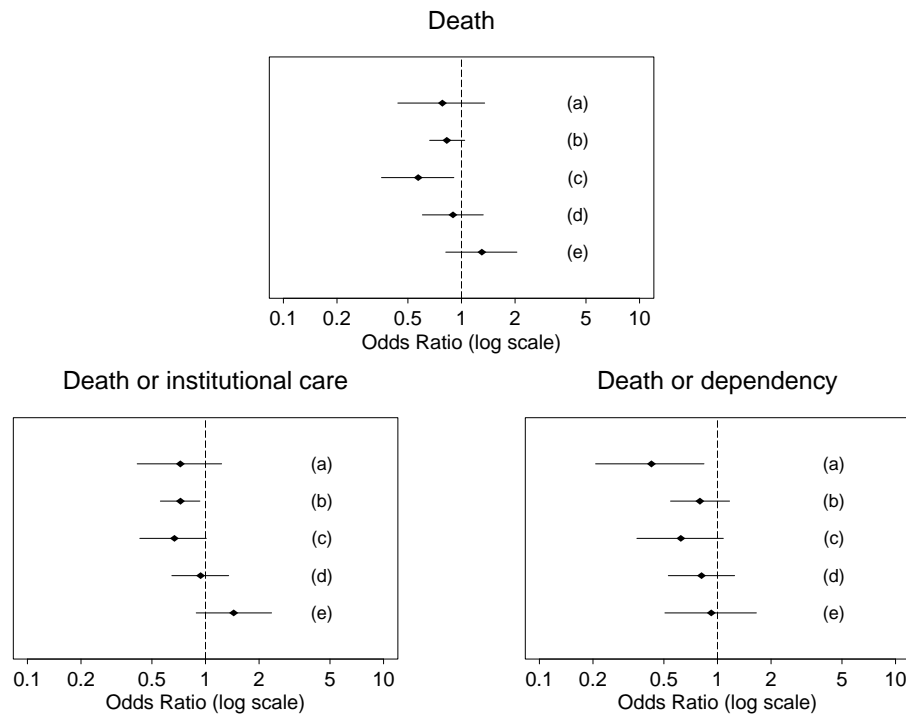
Figure 8.1: Comparing organised stroke unit care with general medical wards for the outcomes of death, death or institutional care and death or dependency. Comparisons are: (a) semi-intensive acute ward; (b) comprehensive stroke ward; (c) rehabilitation stroke ward; (d) mixed-rehabilitation ward; and (e) mobile stroke team. Results are presented as median odds ratios (shaded diamond) of organised stroke unit care versus general medical ward, with corresponding 95% credible intervals represented by the line. Odds ratios are plotted on the natural log-scale.

outcome also tend to increase for all outcomes. However, the credible intervals are wide so care must be taken when interpreting these results.

As in Chapter 4, sensitivity analysis was performed excluding the seven trials with an informal method of randomisation that is prone to bias (Ronning and Guldvog, 1998; Patel, 2000; Yagura et al., 2005; Cavallini et al., 2003; von Arbin et al., 1980; Strand et al., 1985; Hamrin, 1982). The conclusions for death and

Table 8.3: Probabilities of each service type being the most effective in reducing odds of poor outcome.

Service	Death	Death or institutional care	Death or dependency
General medical ward	<0.01	<0.01	<0.01
Mobile stroke team	<0.01	<0.01	0.02
Mixed rehabilitation ward	0.02	0.01	0.01
Rehabilitation stroke ward	**0.76**	**0.47**	0.17
Comprehensive stroke ward	0.04	0.20	0.01
Acute stroke ward	0.17	0.31	**0.80**

Due to rounding, probabilities do not sum to 1 in every outcome.

death or institutional care remain unchanged when trials with such informal randomisation procedures were excluded. However, this is not true for death or dependency. The results of the sensitivity analysis for death or dependency are given in Table 8.5. The reduction in odds of poor outcome in acute stroke wards is no longer significant, but rehabilitation stroke wards and comprehensive stroke wards have a possible reduction in the odds of death or dependency, but these were not significant. Again, general medical ward, mobile stroke team and mixed rehabilitation ward can be considered ineffective systems of care due to the low probabilities associated with them. There was no change in the conclusions

Table 8.4: Covariate odds ratios and 95% credible intervals.

Covariate/ Comparison	Death	Death or institutional care	Death or dependency
Age			
Old vs young	6.82 (3.41, 20.03)	3.99 (1.79, 8.74)	NA
Severity			
Moderate vs mild	6.83 (3.47, 14.62)	6.51 (3.88, 11.12)	9.16 (5.33, 16.79)
Severe vs mild	52.04 (26.92, 100>)	44.61 (26.95, 86.49)	69.27 (36.63, 100>)

Table 8.5: Odds ratios and probabilities each treatment is most effective for death or dependency sensitivity analysis.

Comparison	Odds ratios (95% CrI)		New probability most effective
	All data	Sensitivity	
Treatment			
General medical ward	-	-	<0.01
Mobile stroke team	0.92 (0.51, 1.65)	0.88 (0.48, 1.53)	0.11
Mixed rehabilitation ward	0.81 (0.53, 1.25)	0.89 (0.62, 1.28)	0.03
Rehabilitation stroke ward	0.62 (0.35, 1.08)	0.67 (0.41, 1.05)	**0.52**
Comprehensive stroke ward	0.80 (0.54, 1.17)	0.76 (0.54, 1.06)	0.15
Acute stroke ward	0.42 (0.21, 0.84)	0.82 (0.42, 1.48)	0.19
Severity			
Moderate vs mild	9.16 (5.33, 16.79)	8.55 (4.55, 17.11)	-
Severe vs mild	69.27 (36.63, 100>)	60.70 (28.50, 100>)	-

regarding the associations of covariates with outcome. The sensitivity analysis results are similar to the results given in Figure 8.1 for the outcomes of death and death or institutional care.

8.4 Discussion

This analysis examined the effect of the five types of organised inpatient (stroke unit) care compared to general medical ward using more of the available information than previous analyses. The results show that patients treated in a dedicated stroke setting perform better than those in general medical wards. Rehabilitation stroke wards and comprehensive stroke wards have greatest reduction in odds of death, dependency and requirement for institutional care. General medical ward, mobile stroke team and mixed rehabilitation ward can be considered ineffective systems of care. Finally, the analysis shows that as age and stroke severity increase then so does the odds of poor outcome.

The model used to analyse the data has a number of strengths. It allows

a more powerful comparison between different types of organised care by using direct and indirect comparison data. For example, the most common direct comparison was between comprehensive stroke ward and general medical ward. The information provided by these direct comparisons along with the direct comparisons between comprehensive stroke ward and, say, acute stroke ward can be used to obtain indirectly an estimate of the comparison between acute stroke ward and general medical ward. This network meta-analysis uses more information than previous analyses, therefore giving more precise treatment effect estimates. It also allows each type of organised service to be compared to a common reference service so it can be seen more easily which type of organised service performs better. The model also includes the patient subgroups of age and severity, adjusting for these effects and providing an estimate of their magnitude. The model used can also be adapted to test for any interactions between covariates. The interaction term was not included here as no substantial interaction existed between age and severity.

In the sensitivity analysis, where trials of lower methodological quality were excluded, the conclusions for the death or dependency outcome were qualitatively different. A potential reason may have been that the trials excluded did not use a formal randomisation procedure, therefore, treatment effects could possibly have been exaggerated in these trials. Another potential reason could be due to less data being analysed in the sensitivity analysis. However, the results of the sensitivity analyses for death or dependency are consistent with the results achieved for death and death or institutional care.

Other limitations in the analysis are the wide credible intervals for the subgroups effects and the inability to estimate the effect of age on death or dependency. Individual patient data could not be obtained for all trials, therefore, lack of data prevented the precise estimation of subgroup effects. If individual patient data were available for all trials then this would allow more precise estimation of the effects of age and severity on all outcomes as the information available would be greater than that in the categorised age and severity used here.

Even with its limitations, this analysis provides further evidence that organised stroke unit care reduces the odds of death, death or institutional care and death or dependency compared to general medical wards. Dedicated wards such as the rehabilitation stroke wards, comprehensive stroke wards and acute stroke wards appear to perform better than mobile stroke teams and mixed rehabilitation wards, and such findings are robust to adjustment for case mix in the form of age and stroke severity. Future research should focus on evaluating the benefits of an acute system of care and rehabilitation stroke care to improve patient outcomes. Attempts should be made to obtain individual patient data and future studies should make individual patient data available for the purposes of meta-analysis to allow more accurate estimation of the effects of age and severity.

Chapter 9

Cost-utility analysis of different systems of organised inpatient (stroke unit) care

Chapter 4 shows that organised inpatient (stroke unit) care reduces deaths, dependency and requirement for institutional care and Chapter 8 shows that dedicated stroke unit services perform better than systems which are not restricted to stroke patients. However, the additional cost of stroke unit care may outweigh the benefits achieved. This Chapter implements a cost-utility analysis to determine which system of organised inpatient (stroke unit) care is most cost-effective.

9.1   Background

Stroke is the leading cause of long-term neurological disability in adults (Wolfe, 2000) meaning that stroke-related costs are one of the largest components in

health care expenditure in Scotland (Isard and Forbes, 1992). Given that the number of patients with first ever stroke is predicted to increase over the next 15 years (Malmgren et al., 1989), it is important to understand the relative cost-effectiveness of different models of inpatient care.

It has been shown that stroke units reduce deaths, requirement for institutionalisation and dependency in stroke patients (Chapter 4) and that stroke units with acute care, comprehensive care and rehabilitation perform better than care systems which are not restricted to patients with acute stroke (Chapter 8). However, it is unclear how the models of inpatient care compare in terms of costs and benefits. Therefore a cost-utility analysis (described in more detail in Section 9.2) is employed to evaluate the additional costs per quality-adjusted life year (QALY) gained by each of the different inpatient service models.

9.2   Economic evaluation

In an ideal world, improvements to service organisation would be accepted provided they improved patient outcomes. However, service improvements often come at a higher cost, and since this is not an ideal world, these costs must be taken into account when deciding which service improvements should be accepted.

Evers et al. (2000) describe four main types of economic evaluation: cost-minimisation analysis; cost-effectiveness analysis; cost-utility analysis; and cost-benefit analysis. A cost-minimisation analysis examines equally effective health care programs so that only costs need to be compared further. The outcome in a cost-effectiveness analysis can be assessed in a variety of ways (for example life years gained, decreased length of stay, cases prevented, pain free days), limiting

comparison between studies as the outcomes can be measured in different ways across different studies. However, a cost-utility analysis expresses health outcomes in a uniform way, the quality-adjusted life-year (QALY). A QALY takes into account both the quantity and the quality of life. The quantity of life (life years) is adjusted using utility values (where 1 corresponds to perfect health and 0 to death) to give a measure of the quality of the life-years. Finally, a cost-benefit analysis allows the analyst to see immediately whether benefits outweigh costs by expressing the outcome of the study in monetary terms so that costs and benefits are measured in the same unit.

A cost-minimisation analysis would not be appropriate as Chapter 8 shows that not all systems of care are equivalent. Additionally, it is difficult to assign monetary values to clinical outcomes, therefore a cost-benefit analysis was not performed. Finally, utility values were readily available, therefore, a cost-utility analysis was preferred over a cost-effectiveness analysis.

The cost-utility analysis considered National Health Service (NHS) costs and health benefits to individuals; it therefore used the same perspective as the National Institutes for Health and Clinical Excellence (NICE) in evaluating health technologies. Costs and benefits were estimated for a period of one year after the patient had the stroke; this was selected due to the lack of long term RCT follow-up.

9.3 Methods

9.3.1 Data

Data from the Stroke Unit Trialists Collaboration systematic review, described in Section 1.5.4 and Chapter 4, were used. In summary, this database includes 31 prospective trials using some form of random allocation of stroke patients to an organised system of inpatient (stroke unit) care or an alternative form of inpatient care. Six systems of care were identified: semi-intensive acute stroke ward; comprehensive stroke ward; rehabilitation stroke ward; mixed-rehabilitation ward; mobile stroke team; and general medical ward (GMW). In addition to the main outcomes (death, death or dependency and death or need for institutional care), data were also gathered on average length of stay and the timings of deaths.

In total 25 trials (Ronning and Guldvog, 1998; Vemmos et al., 2001; Ma et al., 2004; Patel, 2000; Stevens et al., 1984; Garraway et al., 1980; Svensson et al., 1993; Fagerberg et al., 2000; Sulter et al., 2003; Kaste et al., 1995; Cabral et al., 2003; Sivenius et al., 1985; Aitken et al., 1993; Juby et al., 1996; Kalra et al., 1993, 2000; Yagura et al., 2005; Cavallini et al., 2003; Hankey et al., 1997; von Arbin et al., 1980; Laursen et al., 1995; Ilmavirta, 1994; Indredavik et al., 1991; Strand et al., 1985; Hamrin, 1982) recorded length of stay, while 22 trials (Ronning and Guldvog, 1998; Vemmos et al., 2001; Stevens et al., 1984; Garraway et al., 1980; Fagerberg et al., 2000; Sulter et al., 2003; Kaste et al., 1995; Gordon and Kohn, 1966; Cabral et al., 2003; Sivenius et al., 1985; Dey et al., 2005; Wood-Dauphinee et al., 1984; Feldman et al., 1962; Aitken et al., 1993; Juby et al., 1996; Kalra et al., 2000; Hankey et al., 1997; Laursen et al., 1995; Ilmavirta, 1994; Indredavik et al., 1991; Strand et al., 1985; Hamrin, 1982) provided information on the total number

of deaths at varying time points. Unfortunately, data on the timings of those who were dependent or required institutional care (where requirement for institutional care is considered a surrogate measure of dependency) were not available in the Stroke Unit Trialists' Collaboration database. However, previous analysis by the SUTC (Stroke Unit Trialists' Collaboration, 1997b) provided information on the proportions of patients who were alive and living at home at several time points, including weeks 1, 2, 3, 4, 12, 26 and 52.

9.3.2 Costs

The variables that are assumed to affect NHS costs between types of inpatient stroke care were NHS staff time and length of stay. It was assumed that the use of medicines, tests, and other interventions would be similar in all types of unit. These aspects of care are less expensive than staff time and so it is believed they would not affect overall cost estimates. Additionally, Chapter 5 shows that the benefits of stroke unit care can be attributed to interventions to prevent complications, therefore, the hypothesis that observed differences in outcome are attributable to staff time and skills seems appropriate.

Kalra et al. (2000), Langhorne and Pollock (2002) and Rodgers et al. (2003) provide information on staffing levels for acute stroke ward, comprehensive stroke ward, rehabilitation stroke ward and mobile stroke team. No information on staffing levels for mixed rehabilitation was identified in the literature search.

Staff inputs of interest were allied health professionals including physiotherapists, occupational therapists and speech and language therapists; medical staff (junior and senior); and nursing staff (sister, staff nurse and untrained assistant).

It was assumed that the medical and nursing staff levels in a mobile stroke team would be the same as those in a general medical ward. All staff type and level information was converted to the form number of hours per patient per day. In order to convert the values it was assumed that a working week consists of 37.5 hours; that there were 3 shifts per day; and that the proportion of staff in each grade was the same for all systems of care. For example, it was assumed that the proportion of senior doctors and junior doctors is the same for all systems of care.

Costs for allied health professionals, medical and nursing staff in organised inpatient stroke unit care were obtained from Curtis (2007), while general medical ward costs were obtained from the Information Services Division, NHS Scotland (2007). Total costs per patient per day were calculated for each system of care. Total cost per patient per stay was calculated using the average length of stay in each system of organised inpatient care (Kalra et al., 2000; Langhorne and Pollock, 2002; Rodgers et al., 2003), while the average length of stay in general medical ward was estimated from the SUTC database. Of the 25 trials in the SUTC database where average length of stay was recorded, 18 compared some model of care with a general medical ward (Ronning and Guldvog, 1998; Vemmos et al., 2001; Ma et al., 2004; Patel, 2000; Garraway et al., 1980; Svensson et al., 1993; Fagerberg et al., 2000; Kaste et al., 1995; Cabral et al., 2003; Sivenius et al., 1985; Aitken et al., 1993; Juby et al., 1996; Hankey et al., 1997; von Arbin et al., 1980; Laursen et al., 1995; Indredavik et al., 1991; Strand et al., 1985; Hamrin, 1982). However, Stroke Unit Trialists' Collaboration (2007) also shows that there were minimal differences in length of stay between systems of care, so estimates of length of stay in each system will be examined in a sensitivity analysis.

9.3.3 Benefits

The proportion of deaths among patients treated in a general medical ward at each time point was calculated using those trials in the SUTC database where general medical ward was the comparator. Of the 22 trials where timing of deaths were recorded, 18 compared some model of organised inpatient (stroke unit) care to general medical wards (Ronning and Guldvog, 1998; Vemmos et al., 2001; Garraway et al., 1980; Fagerberg et al., 2000; Kaste et al., 1995; Gordon and Kohn, 1966; Cabral et al., 2003; Sivenius et al., 1985; Dey et al., 2005; Wood-Dauphinee et al., 1984; Feldman et al., 1962; Aitken et al., 1993; Juby et al., 1996; Hankey et al., 1997; Laursen et al., 1995; Indredavik et al., 1991; Strand et al., 1985; Hamrin, 1982). The proportion of deaths were calculated at weeks 1, 2, 3, 4, 12, 26 and 52.

Using data from Stroke Unit Trialists' Collaboration (1997*b*) (Figure 9.1), the
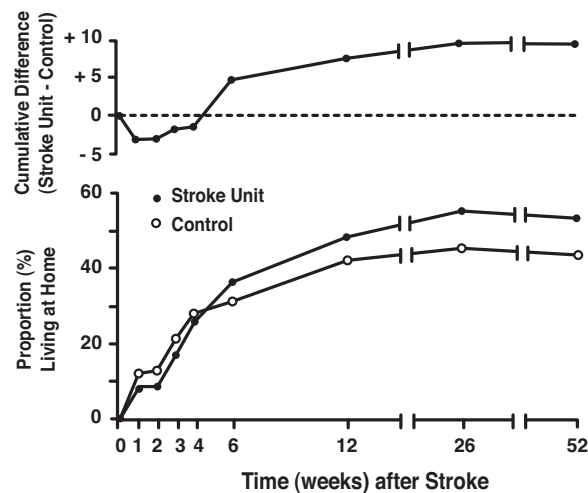
Figure 9.1: Proportion of patients living at home after the index stroke and cumulative difference between stroke unit and control subjects. Taken from Stroke Unit Trialists' Collaboration (1997*b*).
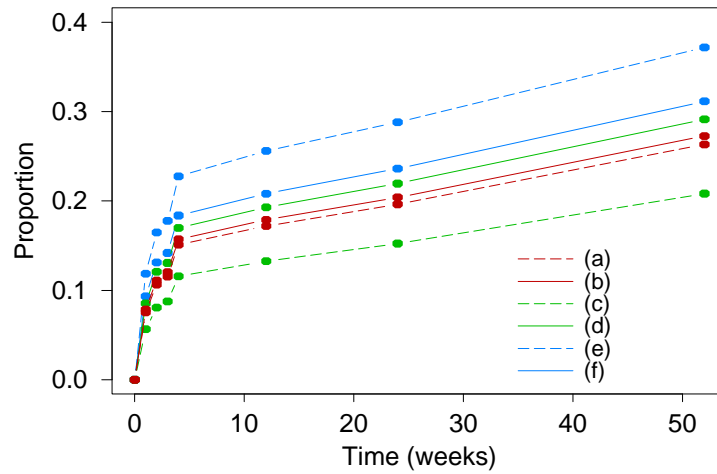
Figure 9.2: Probability of death over a 52 week period in each system of care: (a) acute stroke ward; (b) comprehensive stroke ward; (c) rehabilitation stroke ward; (d) mixed rehabilitation ward; (e) mobile stroke team; and (f) general medical ward.

proportions of patients who were alive and living at home in weeks 1, 2, 3, 4, 12, 26 and 52 were obtained. Together with the proportion of patients who were dead at each of these time points, as calculated from the SUTC database, the proportions of patients in general medical ward in each of the dead, dependent and independent categories at weeks 1, 2, 3, 4, 12, 26 and 52 were able to be calculated.

The Stroke Unit Trialists' Collaboration (2007) showed that benefits of stroke unit care remained up to five years post-stroke. It is therefore reasonable to assume that each treatment effect odds ratio for death and death or dependency obtained in Chapter 8 remains constant over the 52 week period for each system of care. The proportions of patients who were independent, dependent and dead could then be calculated at each time point in the remaining five systems of
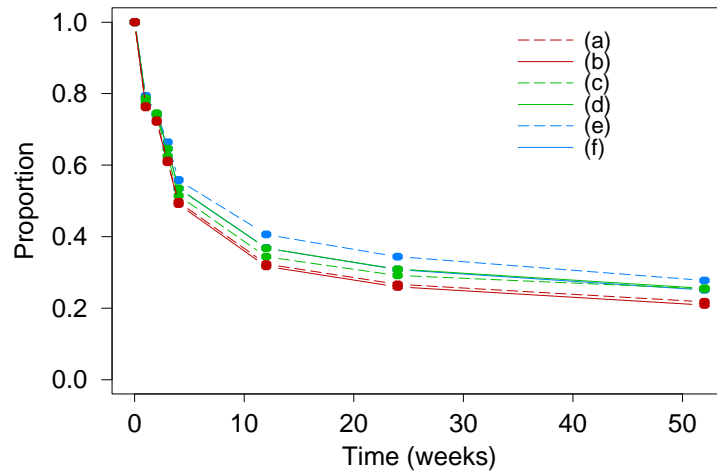
Figure 9.3: Probability of dependence over a 52 week period in each system
of care: (a) acute stroke ward; (b) comprehensive stroke ward;
(c) rehabilitation stroke ward; (d) mixed rehabilitation ward; (e)
mobile stroke team; and (f) general medical ward.

organised inpatient (stroke unit) care (Figures 9.2, 9.3 and 9.4). The number
of person-years saved can then be calculated for independent and dependent
patients.

Dorman et al. (2000) give utility values for dependent (EuroQoL EQ-5D mean
0.38; 95% CI 0.29 to 0.47), independent (0.74; 0.69 to 0.79) and recovered (0.80;
0.80 to 0.96) stroke patients. For the purposes of this analysis, the recovered
utility values were not used as the data available did not distinguish between
independent and recovered.

An estimate of the QALY can then be obtained by summing the saved life-
years adjusted using the utility values.

For the base-case, it is assumed that the odds ratios are equal to the medians
given in Chapter 8. Note that the sensitivity analysis results were used for the
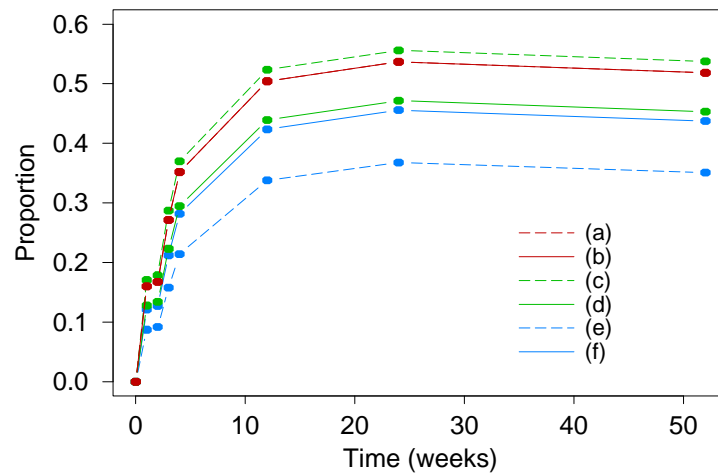
Figure 9.4: Probability of independence over a 52 week period in each system of care: (a) acute stroke ward; (b) comprehensive stroke ward; (c) rehabilitation stroke ward; (d) mixed rehabilitation ward; (e) mobile stroke team; and (f) general medical ward.

outcome of death or dependency as given in Table 8.5. Additionally, it is assumed that the utility values are the mean values reported.

9.3.4 Analysis

The incremental cost-effectiveness ratio (ICER) gives a measure of the additional cost per additional QALY gained. The ICER is calculated as

$$\text{ICER} = \frac{C_\text{B} - C_\text{A}}{Q_\text{B} - Q_\text{A}},$$

where C_A and C_B are the costs in treatments A and B, respectively, and Q_A and Q_B are the QALYs in treatments A and B, respectively. This gives the additional cost of treatment B compared to treatment A per additional QALY gained.

Incremental cost-effectiveness ratios can be calculated in two different ways to determine the most cost-effective treatment. Firstly, in a marginal analysis the treatments are ranked by their QALY estimates. The treatment with the lowest QALY estimate is compared to the treatment with the second lowest. The treatments are compared sequentially to the treatment next in rank to determine if the added cost is worth the additional gain in QALY. The second type of ICER calculated uses a comparative approach where all treatments are compared to general medical ward to determine which has the smallest extra cost per additional QALY gained.

9.3.5 Sensitivity analysis

A summary list of all assumptions are given in Table 9.1. In the sensitivity analysis, seven of the assumptions in Table 9.1 are varied and the effect on the results are examined.

In assumption 4., the staffing levels obtained from Kalra et al. (2000), Langhorne and Pollock (2002) and Rodgers et al. (2003) are assumed accurate. However, the level of medical staffing in acute stroke may be greater than that stated, therefore, the level of medical staffing in an acute stroke ward is increased to match that of general medical ward.

In assumption 5., it is assumed a working week is 37.5 hours long. However, it is possible that greater or less than 37.5 hours are worked per week. In assumption 6., it is assumed that there are 3 nursing shifts per day on average but it is also possible that there could be more or fewer shifts per day. For assumption 7., the unit costs for allied health professionals, medical and nursing staff in organised

Table 9.1: Summary of assumptions in the base-case model.

Description
1. Requirement for institutional care is a surrogate measurement for dependency.
2. Costs of medicines, tests and additional interventions were assumed to be similar for all systems of care.
3. Information on staffing levels for acute stroke ward, comprehensive stroke ward, rehabilitation stroke ward and mobile stroke team obtained from Kalra et al. (2000), Langhorne and Pollock (2002) and Rodgers et al. (2003).
4. Staff levels in mobile stroke team care are similar to general medical ward.
5. There are 37.5 hours in a working week.
6. There are 3 nursing shifts per day.
7. Average unit costs for allied health professionals, medical and nursing staff in organised inpatient stroke unit care taken from Information Services Division, NHS Scotland (2007).
8. Proportions of staff in each grade are equal for all systems of care.
9. Length of stay in each system of care are estimated from Kalra et al. (2000), Langhorne and Pollock (2002) and Rodgers et al. (2003).
10. Relative treatment effects are constant over a 52 week period.
11. Odds ratios are equal to the median values for treatment effect, as given in Figure 8.1 and Table 8.5.
12. Utilities are equal to the mean of the utility values reported.

inpatient stroke unit care are varied by 25%.

Stroke Unit Trialists' Collaboration (2007) have shown that there was not any substantial difference in length of stay across the different systems of care. Therefore, if it were to be assumed that length of stay was equal in all systems of care, how would this affect the conclusions of the cost-effectiveness analysis (assumption 9.)? It should also be noted here, however, that the length of stay is

recorded as total length of stay in hospital and not in any one component of care. For example, those entered into acute stay do not spend their entire hospital stay in acute care. Similarly, those entered into rehabilitation stroke care were first treated in an alternative form of care prior to entrance into rehabilitation. This will be discussed further in Sections 9.6 and 9.7.

In assumption 11., the odds ratio estimates used are the median values. However, for each system of care versus general medical ward, all odds ratios have credible intervals associated with them. In the cost-utility analysis the death, death or dependency and death or institutional care odds ratios for each treatment are varied to the upper and lower limits of the credible intervals and the effect on the cost-effectiveness results examined. If any odds ratio values are found to change the conclusions of the base-case analysis then these are varied in tandem to determine their combined effects.

As with the odds ratio values, the utility values each have an associated confidence interval (assumption 12.). The values for dependence and independence are varied individually and the results of the cost-effectiveness analysis examined.

9.4   Base-case results

Table 9.2 shows the average costs of health workers and total costs per patient per day in each system of care. Notice that rehabilitation stroke ward appears to be the least costly per patient per day, followed by general medical ward and mobile stroke team. Comprehensive and acute stroke wards appear to be most costly since these systems of care have a higher amount of medical and nursing staff, respectively.

Table 9.2: Average costs (£) per patient per day for health workers in each system of care.

Staff (cost per hour)	General medical ward	Mobile stroke team	Rehabilitation stroke ward	Comprehensive stroke ward	Acute stroke ward
Medical					
Senior doctor (£132)	46.67	46.67	17.68	53.03	14.14
Junior doctor (£41)	-	-	8.79	21.96	7.03
Nursing					
Sister (£31)	106.88	106.88	5.65	7.31	29.76
Staff nurse (£22)	-	-	21.04	27.23	58.08
Assistant (£14)	-	-	9.38	30.00	44.80
Allied health professions	19.79				
Physiotherapy (£38)	-	12.59	30.54	29.52	24.43
Occupational therapy (£41)	-	5.13	25.26	25.26	21.96
Speech/language therapy (£41)	-	5.89	8.79	10.98	7.69
Total costs	173.33	177.15	127.10	205.29	207.89

Results can be expressed as ICERs and cost versus QALY figures. Drawing conclusions from cost versus QALY figures can be more difficult to interpret than ICER results due to their informal nature, however, figures can also provide more information than ICERs. Possible treatments that have a high QALY estimate but low cost may be overlooked in a marginal ICER analysis due to the sequential nature of the comparisons. For example, a treatment may be more cost-effective than the next treatment in rank, but less cost-effective than a higher ranked treatment.

Table 9.3 and Figure 9.5 summarise the total cost per stay and mean QALY over one year post stroke for each system of care. Mobile stroke team has the smallest QALY while rehabilitation stroke ward has the largest QALY. However, also notice that of the systems of organised (stroke unit) care, rehabilitation stroke ward has the largest cost per patient per stay.

For the marginal ICER analysis, although the system of care with the lowest QALY is mobile stroke team, general medical ward is chosen as the initial system of care. The reason being that general medical ward can be considered a control

Table 9.3: Estimates of QALY and total costs per patient for each system of care

System of care	QALY	Cost per patient per day (£)	Average length of stay (days)*	Cost per stay (£)
General medical ward	0.443	173.33	39	6760
Mobile stroke team	0.395	177.15	16	2834
Mixed rehabilitation ward	0.454	-	-	-
Rehabilitation stroke ward	0.507	127.10	52	6609
Comprehensive stroke ward	0.482	205.29	24	4927
Acute stroke ward	0.485	207.89	9	1871

*Average length of stay in general medical ward estimated from SUTC database. Length of stay for other systems taken from Kalra et al. (2000), Langhorne and Pollock (2002) and Rodgers et al. (2003).
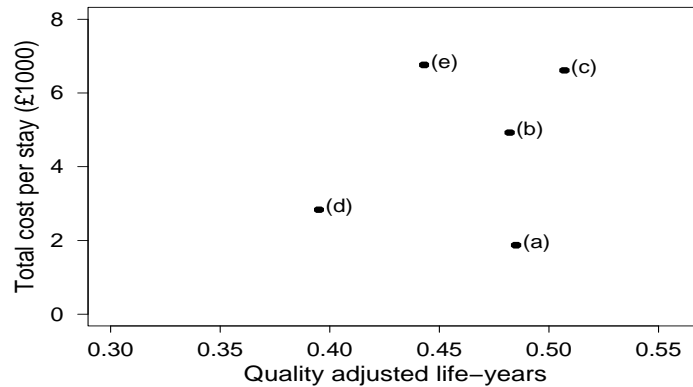
Figure 9.5: Cost and quality adjusted life-years in each system of care: (a) acute stroke ward; (b) comprehensive stroke ward; (c) rehabilitation stroke ward; (d) mobile stroke team; (e) general medical ward.

treatment as it contains none of the aspects of care associated with organised inpatient (stroke unit) care. General medical ward is compared to the next most effective treatment, comprehensive stroke ward. The ICER is then calculated as

$$
\begin{aligned}
\text{ICER} &= \frac{C_{\text{Comprehensive}} - C_{\text{GMW}}}{Q_{\text{Comprehensive}} - Q_{\text{GMW}}} \\
&= \frac{£4927 - £6760}{0.482 - 0.443} \\
&= -£46,878.
\end{aligned}
$$

Therefore, one would expect to save £46,878 per additional QALY gained if comprehensive stroke ward was chosen over general medical ward, so it is worthwhile accepting a comprehensive stroke ward as a new treatment compared to a general medical ward. Table 9.4 shows the ICER results for the marginal analysis (ICER$_1$). Comprehensive stroke ward is compared to the next most effective

Table 9.4: Incremental cost-effectiveness ratios for the base-case.

System of care	ICER$_1$ (£)*	ICER$_2$ (£)†
General medical ward	-	-
Comprehensive stroke ward	-46,878	-46,878
Acute stroke ward	**-1,099,926**	**-116,738**
Rehabilitation stroke ward	215,040	-2,356

*Marginal analysis, where treatments are compared sequentially.

†Comparitive analysis where all treatments are compared to general medical ward.

treatment, acute stroke ward. Finally, acute stroke ward is compared to rehabilitation stroke ward where one would expect to pay £215,040 per additional QALY gained. Typically a maximum of £20,000 per additional QALY is taken as the threshold for accepting new treatments and £30,000 or more can be acceptable under some circumstances (National Institutes for Health and Clinical Excellence, June 2008), meaning that rehabilitation stroke ward is not cost-effective compared with acute stroke ward.

For the comparative ICER approach, mobile stroke team is again not considered here since its QALY is lower than the QALY for general medical ward. In Table 9.4, ICER$_2$ compares comprehensive stroke ward, acute stroke ward and rehabilitation stroke ward to general medical ward. Again acute stroke ward appears to be the most cost-effective system of care since it has the highest saving.

9.5 Sensitivity results

For the sensitivity analysis the changes in ICER$_1$, the ICER results for the marginal analysis, were examined. As with the base-case, general medical ward is taken as the initial treatment for all ICER calculations.

9.5.1 Assumption 4.: medical costs in acute stroke unit and general medical ward

Table 9.5 gives the results of the sensitivity analysis when it is assumed the cost of medical staff in an acute stroke ward is equal to that of a general medical ward. Varying these assumptions does not affect the QALY estimates given in Table 9.3. Table 9.5 shows that the conclusions obtained from the base-case do not change: acute stroke unit care still appears to be the most cost-effective system of care.

Table 9.5: Incremental cost-effectiveness ratios for the sensitivity analysis on assumption 4.: equal cost of medical staff in an acute stroke ward and general medical ward.

System of care	QALY	Cost	ICER (£)
General medical ward	0.443	6760	-
Comprehensive stroke ward	0.482	4927	-46878
Acute stroke ward	**0.485**	**2100**	**-1,017,338**
Rehabilitation stroke ward	0.507	6609	204,627

QALY values remain unchanged from base-case. Costs are calculated per stay and ICER calculated using marginal analysis, where treatments are compared sequentially.

9.5.2 Assumptions 5., 6. and 7.: hours per week, shifts per day and unit costs

Table 9.6 gives the results of the sensitivity analysis when assumptions 5., 6. and 7. are varied independently: assumption 5. varies the number of hours in a working week between 30 hours per week and 45 hours per week; for assumption 6., the number of nursing shifts per day are varied between 2 and 4 shifts per day; and for assumption 7. the unit costs for allied health professionals, doctors and

Table 9.6: Incremental cost-effectiveness ratios for the sensitivity analysis on assumptions: 5. hours per week; 6. shifts per day; and 7. unit costs for staff.

System of care	Lower value (£)		Upper value (£)	
	Cost	ICER	Cost	ICER
Hours per week*				
General medical ward	6760	-	6760	-
Comprehensive stroke ward	3942	-72,079	5912	-21,677
Acute stroke ward	1736	**-794,007**	2006	**-1,405,845**
Rehabilitation stroke ward	5288	161,197	7931	268,883
Shifts per day†				
General medical ward	6760	-	6760	-
Comprehensive stroke ward	4927	-46,878	4927	-47,000
Acute stroke ward	1473	**-1,243,149**	2269	**-956,704**
Rehabilitation stroke ward	6609	233,099	6609	196,982
Medical staff‡				
General medical ward	6760	-	6760	-
Comprehensive stroke ward	4477	-58,387	5377	-35,370
Acute stroke ward	1823	**-955,104**	1919	**-1,244,749**
Rehabilitation stroke ward	6265	201,589	6953	228,492
Nursing staff‡				
General medical ward	6760	-	6760	-
Comprehensive stroke ward	4540	-56,783	5314	-37,077
Acute stroke ward	1572	**-1,068,003**	2169	**-1,131,871**
Rehabilitation stroke ward	6140	207,309	7078	222,770
Allied health professionals‡				
General medical ward	6760	-	6760	-
Comprehensive stroke ward	4546	-56,611	5322	-36,788
Acute stroke ward	1754	**-1,005,182**	1993	**-1,198,141**
Rehabilitation stroke ward	5801	183,687	7449	247,619

QALY values remain unchanged from base-case. Costs are calculated per stay and ICER calculated using marginal analysis, where treatments are compared sequentially.

*Hours per week: lower value = 30; upper value = 45.

†Shifts per day: lower value = 2; upper value = 4.

‡Staff costs: lower value = unit cost − 25%; upper value = unit cost + 25%
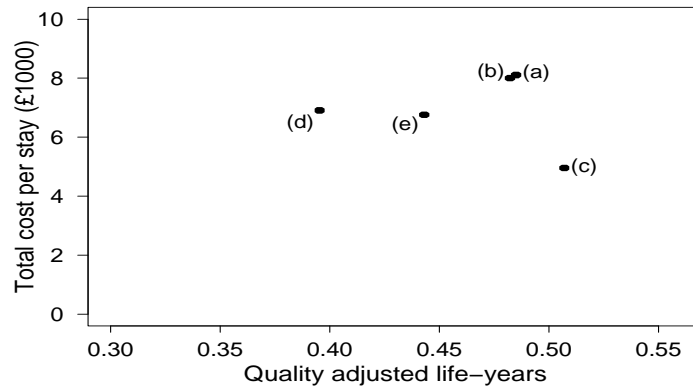
Figure 9.6: Cost and quality adjusted life-years in each system of care when
length of stay is assumed to be 39 days in each system of care: (a)
acute stroke ward; (b) comprehensive stroke ward; (c) rehabili-
tation stroke ward; (d) mobile stroke team; (e) general medical
ward. QALY values remain unchanged from base-case and costs
are calculated per stay.

nursing staff are varied by 25% in either direction. Varying these assumptions
do not affect the QALY estimates given in Table 9.3, but affect the cost of each
system of care. However, Table 9.6 shows that the conclusions obtained from
the base-case do not change: acute stroke unit care still appears to be the most
cost-effective system of care.

9.5.3   Assumption 9.: length of stay

Previously, it was assumed that the average length of stay in each system of care
was the average obtained from Kalra et al. (2000), Langhorne and Pollock (2002)
or Rodgers et al. (2003). It is now assumed that the length of stay is equal to
the general medical ward length of stay of 39 days, as obtained from Stroke Unit
Trialists' Collaboration (2007).

Figure 9.6 shows that rehabilitation stroke ward system of care is the most cost-effective when the length of stay is assumed to be equal in all systems of care. However, as will be discussed in more detail in Section 9.7, the assumption regarding length of stay is complicated by the patient pathway through hospital.

9.5.4   Assumption 11.: odds ratios

In the base-case, the odds ratios for death, death or dependency and death or institutional care are assumed to be the median odds ratio values for each treatment. For the sensitivity analysis, the odds ratios for death, death or dependency and death or institutional care are varied simultaneously but independently for each treatment using the values given by the credible intervals. Varying the odds ratios does not alter the cost per stay in each system of care, but it does affect the QALY estimate gained for each care system.

Figure 9.7 shows that as the odds ratios in mobile stroke team, rehabilitation stroke ward and comprehensive stroke ward are varied, the conclusions are again unaffected: acute stroke ward appears to be most cost-effective. This is also true when the acute stroke ward odds ratio is assumed to be the lower value of its credible interval. However, when the upper value of the odds ratio credible interval is chosen, the most cost-effective treatment appears to be comprehensive stroke ward.

9.5.5   Assumption 12.: utility values

As with the odds ratios, the utility values obtained from Dorman et al. (2000) also have confidence intervals associated with them. The utility values for dependence
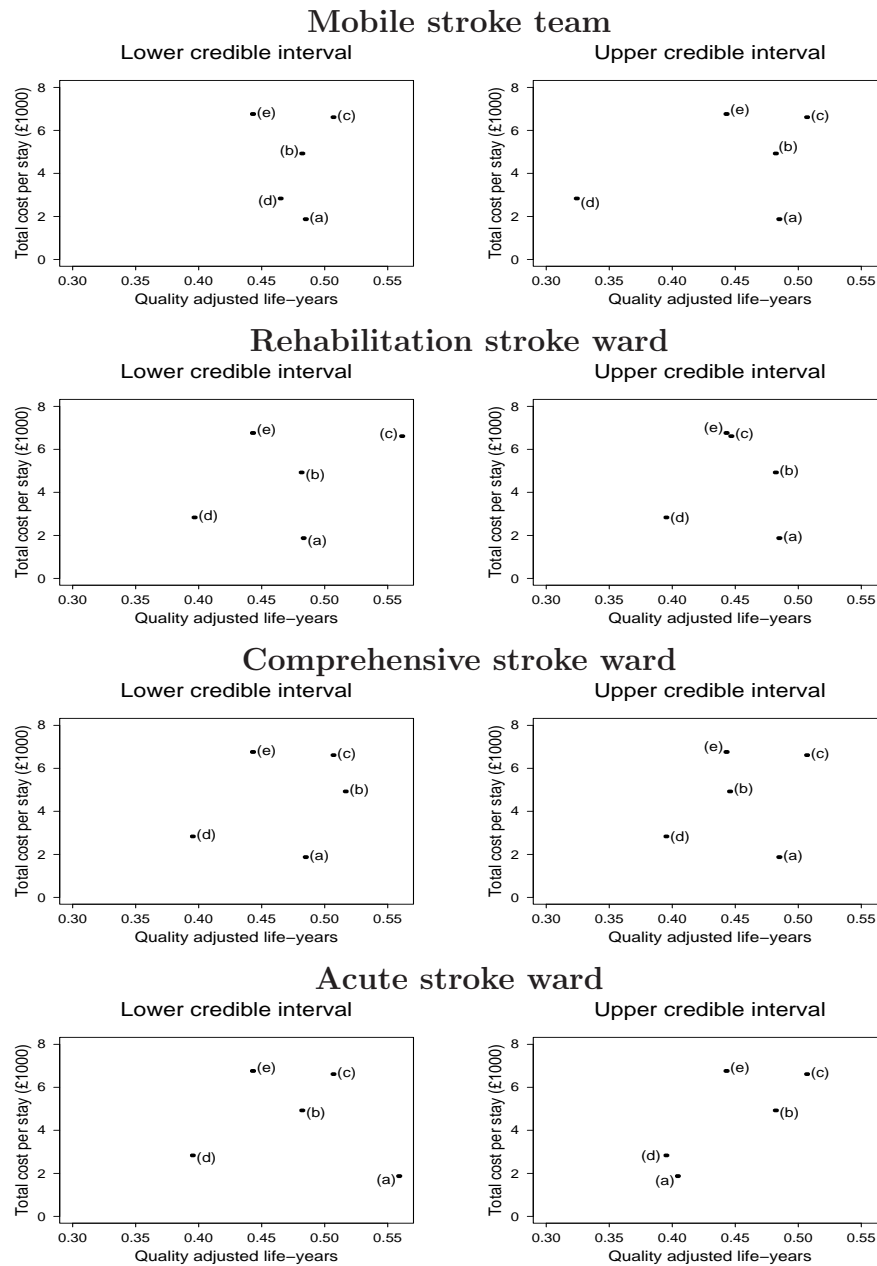
Figure 9.7: Cost and quality adjusted life-years in each system of care when each odds ratio is varied by its credible interval: (a) acute stroke ward; (b) comprehensive stroke ward; (c) rehabilitation stroke ward; (d) mobile stroke team; (e) general medical ward.
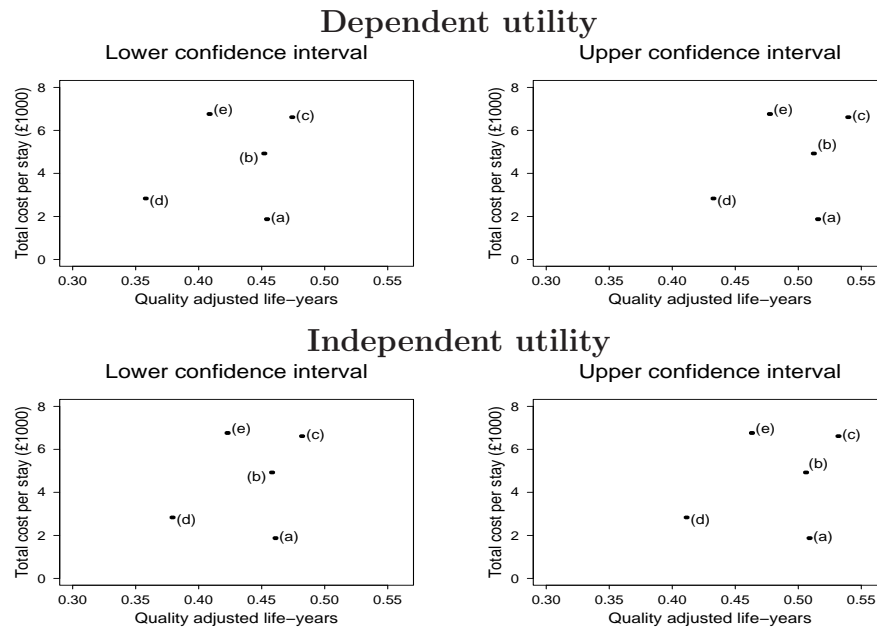
Figure 9.8: Cost and quality adjusted life-years in each system of care when utility values for dependent and independent stroke survivors are varied by their confidence intervals: (a) acute stroke ward; (b) comprehensive stroke ward; (c) rehabilitation stroke ward; (d) mobile stroke team; (e) general medical ward.

and independence are independently varied using these confidence intervals. The results are shown in Figure 9.8.

As with previous sensitivity analysis, the conclusions are not changed after varying the utility values. Acute stroke ward still appears to be the most cost-effective.

9.6 Patient pathway analysis

As mentioned in Section 9.5.3 and discussed in Section 9.7, the total costs per stay in each system of care are not strictly comparable. Patients admitted to

hospital following stroke may be treated in more than one system of care. For example, patients admitted to an acute stroke ward may then be transferred to an alternative system of care if the patient requires further inpatient care. This section attempts to estimate the cost of patient pathways through hospital in order to compare directly the total cost per stay for each system of care.

After acute care a patient may either die, return home or require further care. Langhorne (2003-2007) estimates that 30% (25% - 35%) of patients will require further treatment in hospital following initial treatment. The pathways of care considered in this section are: acute stroke ward followed by rehabilitation stroke ward; comprehensive ward only; mobile stroke team only; and general medical ward only.

Table 9.3 summarises the QALY, average length of stay and cost per patient per day in each system of care. First, consider the pathway of acute stroke ward followed by rehabilitation stroke ward. The average length of stay in an acute stroke ward is 9 days (Rodgers et al., 2003) and the average length of stay in a rehabilitation ward is 52 days (Langhorne and Pollock, 2002). However, only 30% of patients are admitted to a rehabilitation stroke ward. Therefore, the total cost per stay is calculated as the acute stroke ward cost per stay plus 30% of the rehabilitation stroke ward cost per stay.

In order to calculate a comparable total cost per stay for the other systems of care, the average total lengths of stay must be estimated. This is given as the average length of stay in acute ward (9 days) added to 30% of the average length of stay in a rehabilitation ward (52 days). This gives the total average length of stay as 24.6 days. Assuming the total average length of stay to be similar in all systems of care (Stroke Unit Trialists' Collaboration (2007) have shown
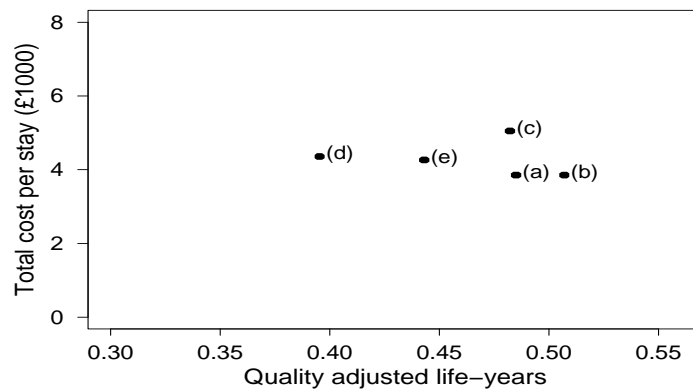
Figure 9.9: Cost and quality adjusted life-years in each pathway of care:
(a) acute stroke ward and rehabilitation stroke ward (acute
stroke ward QALY); (b) acute stroke ward and rehabilitation
stroke ward (rehabilitation stroke ward QALY); (c) comprehen-
sive stroke ward; (d) mobile stroke team; (e) general medical
ward.

that there was not any substantial difference in length of stay across the different
systems of care), the total cost per stay can be estimated.

Figure 9.9 shows that acute stroke ward followed by rehabilitation stroke ward
is the most cost-effective pathway of care. Although an estimate of the QALY
for this pathway was not obtained, it is reasonable to assume that it will lie
somewhere between that of the acute stroke ward care and the rehabilitation
stroke ward care QALYs. The conclusions do not change when both QALY
estimates for acute and rehabilitation stroke ward are examined.

There is uncertainty around the estimate of the proportion of patients who
require further hospital care. Therefore, Figure 9.10 shows how the estimate
of total cost changes when this proportion is varied. Acute care followed by
rehabilitation care remains the most cost-effective pathway of care.
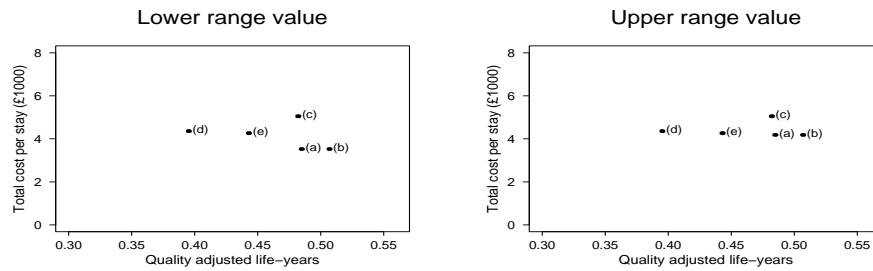
Figure 9.10: Cost and quality adjusted life-years in each pathway of care when the proportion requiring care is varied around its range: lower value 25%, upper value 35%. Pathways are: (a) acute stroke ward and rehabilitation stroke ward (acute stroke ward QALY); (b) acute stroke ward and rehabilitation stroke ward (rehabilitation stroke ward QALY); (c) comprehensive stroke ward; (d) mobile stroke team; (e) general medical ward.

9.7 Discussion

The Stroke Unit Trialists' Collaboration (2007) has shown that organised inpatient (stroke unit) care reduces deaths, dependency and requirement for institutional care, with more organised care proving better than less organised care. Chapter 8 shows that rehabilitation stroke wards are most effective at improving outcomes followed by acute stroke wards and comprehensive stroke wards. However, the higher level of organisation also incurs greater expense due to higher staffing levels. Despite this, acute stroke unit care appears to be the most cost-effective system of care.

Social work costs were not considered in this analysis. However, social work costs may be correlated to NHS costs, therefore, it should not affect the conclusions gained.

Several assumptions are made during this analysis. Firstly, the costs of medicines, tests and additional interventions are assumed to be similar for all

systems of care. This seems reasonable since patients with complications related to stroke will be treated with the same medicines regardless of which system of care they receive. Secondly, it seems reasonable to assume that those who require institutional care are also dependent, therefore requirement for institutional care can be considered a surrogate measurement for dependency. Thirdly, since mobile stroke teams are usually operated within a general medical ward, staffing levels within these two systems of care will be similar. Fourthly, the proportions of staff within each grade are assumed to be equal in all systems of care. This means, for example, that the ratio of senior and junior doctors is equal in all systems of care. This allowed the estimation of costs within some systems of care where medical staffing was given but not staffing levels of each grade. This was also assumed for nursing staffing levels. Finally, the Stroke Unit Trialists' Collaboration (2007) show that the benefits of organised stroke unit care appear to remain for up to 5 years post stroke, therefore, it seems reasonable to assume that the relative benefits of stroke unit care will remain constant over a 52 week period.

Sensitivity analysis was performed based on the remaining assumptions. The results appear to be robust when the assumptions affecting staffing levels and costs are varied. In the base-case, the estimate of cost of medical staff in an acute stroke ward appears to be smaller than that of a general medical ward. A sensitivity analysis assuming the medical staff costs are the same in these systems of care revealed no change in the conclusions obtained in the base-case. The results are also robust when assumptions affecting estimation of benefits (that is, the odds ratios and utility values) are varied. Acute stroke unit care appears to be most effective in these circumstances. However, one major limitation becomes

apparent for the length of stay sensitivity analysis.

The length of stay estimates for the base-case results are collected from different sources (Kalra et al., 2000; Langhorne and Pollock, 2002; Rodgers et al., 2003). The studies are relatively old and estimates of length of stay may have changed since publication. The length of stay estimates from these studies are the average lengths of stay within that particular component of care. However, the studies do not give an indication of where the patient was either before entering care (in the case of rehabilitation stroke ward patients could have originally been admitted to a general medical ward or acute stroke ward care and then transferred to a rehabilitation stroke ward) or after leaving care (in case of acute stroke ward patients could have gone on to rehabilitation stroke ward, general medical ward, died or been discharged home or with relatives). This means that the costs per stay calculated for each system of care (particularly acute stroke wards and rehabilitation stroke wards) are not strictly comparable as they do not take into account all the components of inpatient care a patient may have received.

On the other hand, the estimates of length of stay obtained from the Stroke Unit Trialists' Collaboration (2007) are the average total length of stays, not just the length of stay for a particular component. However, as previously stated, a patient may not spend their entire inpatient stay in one component of care. Therefore, if the total length of stay is assumed to be equal in all systems of care then the cost of acute care will be over-estimated, as shown in Figure 9.6. Additionally, the cost of rehabilitation stroke care may be incorrectly estimated since patients will be transferred to rehabilitation care after receiving treatment in another system of care. Again, this means that the costs of each system of

care are not strictly comparable.

Identification of where the benefit of organised inpatient care is achieved is also limited for similar reasons. For those studies randomising to acute care, the differences achieved can be attributed to that care system since this is the only systematic difference between groups. However, for other systems of care the conclusion cannot be so easily determined. Therefore, given the available data, it is not possible to determine if the benefit of inpatient care is gained in the first few weeks of care (that is, in an acute care setting), or if the benefit is achieved by rehabilitation.

Section 9.6 calculates the total cost per stay for different pathways of care in order to compare the cost-effectiveness. It was shown that acute stroke ward care followed by rehabilitation stroke ward care was the most cost-effective pathway of care compared to comprehensive stroke ward care, mobile stroke team care and general medical ward care alone. Although an estimate of the QALY was not available for the acute plus rehabilitation pathway of care, by analysing QALY estimates from both components it was shown that this pathway is still most cost effective. The conclusion also holds when the proportion of those requiring further care is varied.

Several other pathways of care are possible. However, there are several reasons why these were not analysed. In some cases, such as comprehensive stroke ward care, it would not be relevant to analyse this in addition to another component, since this system of care includes acute care and rehabilitation if required. Lack of data prevented the estimation of total cost per stay of patients transferred to and from general medical ward care. The proportions of patients requiring further care may be different for these pathways than those for acute stroke wards. Also,

if patients were entered into a general medical ward for their acute care it may be reasonable to speculate that the hospital may have limited resources and further stroke care, such as rehabilitation, may not be available.

Although the costs of each system of care in this analysis are not strictly comparable, there is a clear conclusion that organised inpatient (stroke unit) care is cost-effective. Rehabilitation stroke ward care appears to be the most effective system of care, but acute stroke ward care appears to be most cost-effective for patients with shorter lengths of stay. Finally, acute stroke ward care followed by rehabilitation stroke ward care, if required, appears to be the most cost-effective pathway of care compared to the other pathways analysed. Stroke Unit Trialists' Collaboration (2007) found that the benefit of organised stroke unit care may remain for up to 5 or even 10 years post stroke, suggesting that the results of the economic analysis may also be stable for this period of time.

Future trials could randomise the pathway of care a patient receives while in hospital instead of randomising to a particular component of care to determine where the benefit of inpatient stroke unit care is achieved and to estimate accurately the total cost of stay for a patient suffering from stroke. Length of stay in each component of care should be recorded to aid the costing of inpatient stay. Acute stroke ward, comprehensive stroke ward and rehabilitation stroke ward care should be examined in particular to determine the benefits of each system of care, while cost estimates should be obtained for the entire patient journey through care, and not just for a particular component of care. Individual patient data should be made available wherever possible to determine costs of patient care and also to allow the inclusion of the data into a meta-analysis to give a more precise estimate of cost-effectiveness.

Chapter 10

Conclusions

10.1   Summary of clinical results

In Chapter 2, a novel stroke severity scale, the IMAGES Stroke Scale (ISS) was introduced and its ability to predict outcome was compared to that of current stroke severity scales. It was found that the IMAGES Stroke Scale is easily derived from routinely available data and requires no specific training for its use. This offers an alternative to the commonly used National Institutes of Health stroke scale (NIHSS), where language barriers or limited access to training material may prevent it being implemented. The modest correlation between ISS and NIHSS is possibly explained by the fact that ISS measures a range of aspects of stroke severity. Despite this, however, the wide familiarity and standardised video training make the NIHSS the preferred acute stroke scale.

Chapter 3 investigates the amount of prognostic information retained when stroke severity is categorised into 3 groups: mild, moderate and severe. This study showed that the categorisation of four currently used stroke scales did not

substantially reduce the predictive ability of the scales. Furthermore, the four stroke scales stratified in this way appear to be equivalent to each other, and although the prognostic accuracy for longer follow-up is lower, it is not further reduced by categorisation of the stroke scales.

Chapter 4 shows that the addition of new trials to the Stroke Unit Trialists' Collaboration (2006) systematic review does not alter the conclusions found in this previous version. The updated review illustrates that patients receiving organised (stroke unit) care are more likely to survive, return home and regain independence than those receiving a less organised service, with a suggestion that the benefit may last for five or even ten years post stroke. However, Chapter 4 does not explain why stroke unit care improves patient outcomes: possible explanations include the prevention of complications or more intense monitoring of acute patients, as explored in Chapters 5 and 6, respectively.

Chapter 5 examines how organised inpatient care improves patient outcomes by exploring the use of interventions to prevent complications. The results indicate that stroke unit care appears to reduce complications of immobility (in particular, infections), although there were also reductions in stroke recurrence or progression. The findings suggest that some of these reductions could be explained by a more comprehensive implementation of measures to prevent complications.

Chapter 6 compares a conventional stroke unit approach with conventional stroke unit care plus continuous monitoring in the acute phase and addresses whether routine automated monitoring for, and treatment of physiological abnormalities reduce adverse outcomes in stroke patients. Routine monitoring may reduce the risk of stroke progression with the reduction in the monitoring group

being greater in mild stroke patients than in moderate or severe patients. The reduction in death or dependency in the monitoring group was greatest in those patients classed as low risk and the reduction in death was greatest in those who did not suffer heart failure. However, caution is needed when interpreting these conclusions since they are based on relatively small data sets.

Chapters 4 to 6 showed that organised inpatient care improves patient outcomes and provides possible explanations for how this is achieved, but there are several different systems of organised inpatient care. Using the mixed treatment comparisons model presented in Chapter 7, Chapter 8 explores whether any one system of care is more effective in improving patient outcomes while also taking account of associations of age and severity with patient outcome. It was found that dedicated wards such as the rehabilitation stroke wards, comprehensive stroke wards and acute stroke wards appear to perform better than mobile stroke teams and mixed rehabilitation wards, and such findings are robust to adjustment for case mix in the form of age and stroke severity.

Finally, given which systems of care are most effective, Chapter 9 implements a cost-utility analysis to determine which system of inpatient care is most cost-effective. Although the costs of each system of care in this analysis were not strictly comparable, there is substantial evidence that organised inpatient (stroke unit) care is cost-effective. Rehabilitation stroke ward care appears to be the most effective system of care, but acute stroke ward care appears to be most cost-effective for patients with shorter lengths of hospital stay. An attempt was made to estimate the cost of different pathways of care and found that care in an acute stroke ward followed by rehabilitation stroke ward was the most cost-effective of the pathways analysed.

10.2 Summary of meta-analysis techniques

In Chapters 4 through 8, various meta-analysis techniques are used; the complexity of the techniques increases with every successive Chapter. Initially, a basic frequentist approach to meta-analysis was employed in Chapter 4 to analyse aggregated data and incorporated fixed and random effect estimates as required. Chapter 5 also analyses aggregated data but introduces a Bayesian random-effects meta-analysis model to calculate odds ratios and absolute risk differences of treatment effect. This model was developed further in Chapter 6 to include covariates in a meta-regression while also expanding the data into an individual patient data format.

Finally, a novel Bayesian meta-analysis model was presented in Chapter 7 that incorporates direct and indirect treatment comparisons along with covariate effects in a network meta-analysis. The inclusion of indirect comparison data increases the ability of the model to estimate precisely the effect of each system of care by utilising the information provided by the combinations of direct comparisons.

This Chapter also shows that not adjusting for covariate effects in a meta-analysis may lead to bias in the estimate of the treatment effect. The data for two covariates were therefore incorporated into the analysis. The model may be expanded to account for more than two covariates. Not all trials included in a meta-analysis provide information on the joint distribution of covariates. However, the model allows for the possibility of some trials only providing marginal data for one or more covariates or no data on the breakdown of covariates. This makes use of all the available data (individual patient data and aggregated data)

for analysis by including trials that do not provide covariate information.

The model describes the estimation of treatment effects using random effects. Covariate effects may be estimated using either fixed or random effects and interactions between the covariates may also be introduced.

10.3 Alternative methodological techniques

The novel Bayesian meta-analysis model presented in Chapter 7 can be developed further. The model can be adapted to examine relative risks and absolute risk differences as well as odds ratios. Additionally, the model could also be adapted to include treatment-covariate interaction terms. As shown in Chapter 6, patients with mild stroke appear to improve more than patients with moderate or severe strokes. Therefore, including treatment-covariate interaction terms may show that certain patient types improve more depending on the type of inpatient care they receive.

Access to individual patient data not only increases the ability to predict precisely the effect of each system of care and the relationship of covariates with outcome, but also allows the use of other meta-analysis techniques. For example, dependency is categorised using the equivalent of the Barthel Index < 19, but individual patient data could allow the use of continuous outcomes meta-analysis (Higgins et al., 2001). Care must be taken here though as this type of analysis assumes equal spacing of the categories, which may not be reasonable in dependency data.

Whitehead and Jones (1994) describe a meta-analysis model for an ordinal response and the model is extended by Whitehead et al. (2001) for the use of

individual patient data. They consider a simple categorical response and a small number of categories with a clear ordering to the categories. With appropriate data, the combined outcomes described in this thesis could be analysed using this approach. Instead of death or dependence, the outcome could be death, dependence and independence. Similarly, death or institutional care could be analysed as death, institutional care and home. With individual patient data, the full dependency scale could even be used, though Whitehead et al. (2001) warn that there is little to gain in efficiency by using more than five categories.

Berkey et al. (1998) present a random effects approach to meta-regression of multiple outcomes compared to fixed effects meta-analysis and meta-regression with separate and multiple outcomes. They find that the random-effect multiple outcomes models provide a more realistic estimate of effects compared to fixed-effect models and the multiple outcomes models provide a more efficient test. Their model may be extended to incorporate several treatments and outcomes. This type of model may also be implemented given the appropriate data, and could possibly be combined with the ordinal outcomes technique.

The sensitivity analysis for the cost-utility analysis in Chapter 9 was performed by varying the values of some of the assumptions given in Table 9.1. This is a straightforward method of sensitivity analysis and as a result there are a number of potential problems, particularly regarding the robustness of the results. Varying one parameter at a time in a series of one-way sensitivity analyses may underestimate the uncertainty, leading to false claims of robustness, whereas multi-way sensitivity analysis (where two or more parameters are varied simultaneously) is difficult to present and interpret (Claxton, 2008). The uncertainty may be better characterised using probabilistic sensitivity analysis.

Instead of point estimates, the input parameters in the model are given specified probability distributions so that the uncertainty surrounding their values can be reflected. The choice of distribution is determined by the form of the data, the type of parameter, and how the parameter is estimated (Claxton et al., 2005).

10.4 Future research

Although the stroke severity scales tested in Chapter 3 appear to be equivalent, they are only four of the currently used measurements. Others scales used in practice, such as the Glasgow Coma Scale (Jennett and Teasdale, 1977), the Canadian Neurological Scale (Côté et al., 1989) or the Edinburgh Stroke Predictor (Weir et al., 2003), should also be examined. Also, the follow-up times of the scales examined in Chapter 3 are fairly short: 1 and 3 months. Follow-up lengths of time in meta-analyses are often longer, for example six or twelve months. Therefore, it would be of interest to perform comparisons of stroke severity scales at later follow-up times to assess if the equivalence of full and categorised scales is maintained.

The findings in Chapter 5 emphasise the potential importance of complications as a treatable factor in stroke outcome. Future research should explore the best ways of preventing and managing specific complications, particularly those that seem to carry a high risk of causing harm. Chapter 6 shows that routine automated monitoring may reduce the risk of stroke in progression and reduce the length of hospital stay. However, the magnitude of benefits may be small, therefore, a large multi-centred trial is required to clarify the uncertainties found in this meta-analysis and to identify if the added costs and efforts of such a policy

are justified by better patient recovery. The trial should be sufficiently large to examine which subgroups of patients and subgroups by co-morbidity may benefit most from routine automated monitoring.

Chapter 8 showed that dedicated wards such as the rehabilitation stroke wards, comprehensive stroke wards and acute stroke wards appear to perform better than mobile stroke teams and mixed rehabilitation wards, and such findings are robust to adjustment for case mix in the form of age and stroke severity. Future research should incorporate covariate-treatment interaction terms to examine if any subgroups of patients respond better to any one treatment. Trials should focus on evaluating the benefits of an acute system of care and rehabilitation, independently and in combination, to improve patient outcomes.

Finally, future trials should randomise patients to pathways of care instead of particular components to determine how the benefit of inpatient stroke unit care is achieved and accurately estimate the total cost of stay for a patient suffering from stroke. Length of stay in each component of care should be recorded to aid in the costing of inpatient stay. Acute stroke ward, comprehensive stroke ward and rehabilitation stroke ward care should be examined in particular to determine the benefits of each system of care, while cost estimates should be obtained for the entire patient journey, and not just for a particular component of that pathway of care. Early supported discharge services should also be considered as part of the patient pathway. Although not considered here, Early Supported Discharge Trialists (2008) show that stroke patients who receive input from an early supported discharge service not only return home earlier than those who received conventional care, but are also more likely to be alive, independent and still living at home 6 months post stroke.

Attempts should be made to obtain individual patient data and future studies should make individual patient data available for the purposes of meta-analysis to allow more accurate estimates of the effects of age and severity and to determine costs of organised inpatient (stroke unit) care.

10.5   Summary

Overall, the novel mixed treatment comparison model presented here compares several systems of organised inpatient (stroke unit) care, but also allows the estimation of covariates. It was found that patients treated in organised inpatient (stroke unit) care have reduced odds of death, dependence and requirement for institutional care. Rehabilitation stroke ward care appears to be the most effective system of care, but acute stroke ward care appears to be most cost-effective for patients with shorter lengths of hospital stay. Finally, acute stroke ward care followed by rehabilitation stroke ward care, if required, appears to be the most cost-effective pathway of care compared to the other pathways analysed.

Appendices

Appendix A

Complications and interventions odds ratio WinBUGS code

WinBUGS 1.4 code for the odds ratio model used in Chapter 5 is given below. Notice that the treatment effect variance has been given the Inverse-Gamma prior.

```
model{

    for(i in 1:m){

        # model events as Binomial
        rC[i] ~ dbin(pC[i],nC[i])
        rT[i] ~ dbin(pT[i],nT[i])

        # log odds ratio model
        logit(pC[i]) <- mu[i]
        logit(pT[i]) <- mu[i] + delta[i]

        # priors for mu and delta
        mu[i] ~ dnorm(0,1.0E-6)
        delta[i] ~ dnorm(d,prec)

    }
```

```
   # priors for delta parameters
   d ~ dnorm(0,1.0E-6)


   # Inverse-Gamma treatment variance
   prec ~ dgamma(0.001,0.001)
   tau.sq <- 1/prec
   tau <- sqrt(tau.sq)



   # residual deviance for C group
   for(i in 1:m){
      rChat[i] <- pC[i] * nC[i]
      devC[i] <- 2* (rC[i]*(log(rC[i]/rChat[i]))
                + (nC[i]-rC[i])*(log((nC[i]-rC[i])/(nC[i]-rChat[i])))))
   }
   resdevC <- sum(devC[])


   # residual deviance for T group
   for(i in 1:m){
      rThat[i] <- pT[i] * nT[i]
      devT[i] <- 2* (rT[i]*(log(rT[i]/rThat[i]))
                + (nT[i]-rT[i])*(log((nT[i]-rT[i])/(nT[i]-rThat[i])))))
   }
   resdevT <- sum(devT[])


   # total residual deviance
   resdev <- resdevC + resdevT


}
```

Appendix B

Complications and interventions

absolute risk difference

WinBUGS code

WinBUGS 1.4 code for the absolute risk difference model used in Chapter 5 is given below. Notice that the treatment effect variance has been given the Inverse-Gamma prior.

```
model{

  for(i in 1:m){

    # model events as Binomial
    rC[i] ~ dbin(pC[i],nC[i])
    rT[i] ~ dbin(pT[i],nT[i])

    # absolute risk difference model
    pC[i] <- mu[i]
    pT[i] <- mu[i] + min(max(delta[i],-pC[i]),(1-pC[i]))

    # priors for mu and delta
    mu[i] ~ dunif(0,100)
```

```
    delta[i] ~ dnorm(d,prec)


  }


  # priors for delta parameters
  d ~ dnorm(0,1.0E-6)

  # Inverse-Gamma treatment variance
  prec ~ dgamma(0.001,0.001)
  tau.sq <- 1/prec
  tau <- sqrt(tau.sq)



  # residual deviance for C group
  for(i in 1:m){
     rChat[i] <- pC[i] * nC[i]
     devC[i] <- 2* (rC[i]*(log(rC[i]/rChat[i]))
              + (nC[i]-rC[i])*(log((nC[i]-rC[i])/(nC[i]-rChat[i]))))
  }
  resdevC <- sum(devC[])


  # residual deviance for T group
  for(i in 1:m){
     rThat[i] <- pT[i] * nT[i]
     devT[i] <- 2* (rT[i]*(log(rT[i]/rThat[i]))
              + (nT[i]-rT[i])*(log((nT[i]-rT[i])/(nT[i]-rThat[i]))))
  }
  resdevT <- sum(devT[])


  # total residual deviance
  resdev <- resdevC + resdevT

}
```

Appendix C

Physiological monitoring covariate WinBUGS code

WinBUGS 1.4 code for the covariate odds ratio model used in Chapter 6 is given below. Notice that the treatment effect variance has been given the Inverse-Gamma prior.

```
model{

    # IPD model
    for (i in 1:m) {

        for (j in 1:Nsubj[i]) {
            # model events as Bernoulli
            Y[i,j] ~ dbern(pi[i,j])
            # log-odds ratio model with covariate main effect and interaction
            logit(pi[i,j]) <- mu[i] + delta[i]*t[i,j]
                            + beta0[i]*x[i,j] + beta*x[i,j]*t[i,j]
        }

        # priors for mu, beta0 and delta
        mu[i] ~ dnorm(0,1.0E-6)
        beta0[i] ~ dnorm(b,bprec)
        delta[i] ~ dnorm(d,prec)
```

```
    }

    # priors for delta parameters
    d ~ dnorm(0,1.0E-6)

    # Inverse-Gamma treatment variance
    prec ~ dgamma(0.001,0.001)
    tau.sq <- 1/prec
    tau <- sqrt(tau.sq)

    # prior for beta
    beta ~ dnorm(0,1.0E-6)

    # priors for beta0 parameters
    b ~ dnorm(0,1.0E-6)

    # Inverse-Gamma covariate variance
    bprec ~ dgamma(0.001,0.001)
    btau.sq <- 1/bprec
    btau <- sqrt(btau.sq)

    # storing pooled treatment and covariate effects
    # treatment effect only
    delta[(m+3)] <- d

    # covariate effect only
    beta0[(m+3)] <- b

    # treatment effect and covariate effect
    delta[(m+4)] <- d + b

    # treatment effect and covariate effect with interaction
    delta[(m+5)] <- d + b + beta

}
```

Appendix D

MTC random covariate

WinBUGS code

WinBUGS 1.4 code for the random covariate odds ratio model without interaction terms (model 2) used in Chapter 6 is given below. Notice that the treatment effect variance has been given the Inverse-Gamma prior.

```
model{

    # i = study, j = treatment arm, k = cell of Sev*Age table

    # pi ~ dirichlet
    for (k in 1:6){
        lkappa[k] ~ dnorm(0,0.1)I(0,)
        log(kappa[k]) <- lkappa[k]
        sumfrom[k] <- sum(kappa[k:6])
    }
    for (i in 1:ns){
        pi[i,1] ~ dbeta(kappa[1],sumfrom[2])
        for (k in 2:5){
            phi[i,k] ~ dbeta(kappa[k],sumfrom[k+1])
            pi[i,k] <- (1 - sum(pi[i,1:(k-1)])) * phi[i,k]
        }
```

```
    pi[i,6] <- 1 - sum(pi[i,1:5])
}


# true successes modelled logistically
for (i in 1:ns) {
    lam[i,1] <- mu[i]
    lam[i,2] <- mu[i] + ddA[i]
    lam[i,3] <- mu[i] + ddS2[i]
    lam[i,4] <- mu[i] + ddS2[i] + ddA[i]
    lam[i,5] <- mu[i] + ddS3[i]
    lam[i,6] <- mu[i] + ddS3[i] + ddA[i]

    for (j in nf[i]:na[i]){
        for (k in 1:6) {
            # log-odds ratio model
            logit(LAM[i,j,k]) <- lam[i,k] + delta[i,t[i,j]]*(1-equals(j,1))
        }

        # equations for marginal event probabilities
        XAM[i,j,1] <- (pi[i,1]*LAM[i,j,1] + pi[i,2]*LAM[i,j,2]) / XP[i,1]
        XAM[i,j,2] <- (pi[i,3]*LAM[i,j,3] + pi[i,4]*LAM[i,j,4]) / XP[i,2]
        XAM[i,j,3] <- (pi[i,5]*LAM[i,j,5] + pi[i,6]*LAM[i,j,6]) / XP[i,3]
        XAM[i,j,4] <- (pi[i,1]*LAM[i,j,1] + pi[i,3]*LAM[i,j,3]
                        + pi[i,5]*LAM[i,j,5]) / XP[i,4]
        XAM[i,j,5] <- (pi[i,2]*LAM[i,j,2] + pi[i,4]*LAM[i,j,4]
                        + pi[i,6]*LAM[i,j,6]) / XP[i,5]
        XAM[i,j,6] <- pi[i,1]*LAM[i,j,1] + pi[i,2]*LAM[i,j,2]
                        + pi[i,3]*LAM[i,j,3] + pi[i,4]*LAM[i,j,4]
                        + pi[i,5]*LAM[i,j,5] + pi[i,6]*LAM[i,j,6]
    }

    # probabilities of being in each marginal cell
    XP[i,1] <- pi[i,1] + pi[i,2]
    XP[i,2] <- pi[i,3] + pi[i,4]
    XP[i,3] <- pi[i,5] + pi[i,6]
    XP[i,4] <- pi[i,1] + pi[i,3] + pi[i,5]
    XP[i,5] <- pi[i,2] + pi[i,4] + pi[i,6]

}
```

```
# read in and model data

#######################
# age * severity (N6) #
#######################

for (i in 1:N6){
   for (j in nf[i]:na[i]){
      for (k in 1:6){

         # model events as Binomial
         r[i,j,k] ~ dbin(LAM[i,j,k],n[i,j,k])

         # r deviance contribution
         rhat[i,j,k] <- LAM[i,j,k] * n[i,j,k]
         dev.r.N6[(i-lowN6+1),(j-nf[i]+1),k]
                     <- 2* (r[i,j,k]*(log(r[i,j,k])-log(rhat[i,j,k]))
                         + (n[i,j,k]-r[i,j,k])*(log(n[i,j,k]-r[i,j,k])
                         - log(n[i,j,k]-rhat[i,j,k])))
      }

      # model n as multinomial
      n[i,j,1:6] ~ dmulti(pi[i,1:6],ntot[i,j])

      for (k in 1:6){
         # n deviance contribution
         nhat[i,j,k] <- pi[i,k]*ntot[i,j]
         dev.n.N6[(i-lowN6+1),(j-nf[i]+1),k]
                     <- 2* n[i,j,k]*log(n[i,j,k]/nhat[i,j,k])
      }
   }
}

# residual deviance for N6 r and n
sumdev.r.N6 <- sum(dev.r.N6[,,])
sumdev.n.N6 <- sum(dev.n.N6[,,])
```

```
#####################
# age + severity (N4) #
#####################

for (i in (N6+1):(N6+N4)){
   for (j in nf[i]:na[i]){
      for (k in 1:4){

         # model events as Binomial
         r[i,j,k] ~ dbin(XAM[i,j,k],n[i,j,k])

         # r deviance contribution
         rhat[i,j,k] <- XAM[i,j,k] * n[i,j,k]
         dev.r.N4[(i-lowN4+1),(j-nf[i]+1),k]
                     <- 2* (r[i,j,k]*(log(r[i,j,k])-log(rhat[i,j,k]))
                        + (n[i,j,k]-r[i,j,k])*(log(n[i,j,k]-r[i,j,k])
                        - log(n[i,j,k]-rhat[i,j,k])))
      }

      # model n as multinomial
      # severity breakdown
      n[i,j,1:3] ~ dmulti(XP[i,1:3],ntot[i,j])
      # age breakdown
      n[i,j,4] ~ dbin(XP[i,4],ntot[i,j])

      for (k in 1:3){
         # n deviance contribution - severity
         nhat[i,j,k] <- XP[i,k] * ntot[i,j]
         dev.n.N4s[(i-lowN4+1),(j-nf[i]+1),k]
                     <- 2* n[i,j,k]*log(n[i,j,k]/nhat[i,j,k])
      }
      # n deviance contribution - age
      nhat[i,j,4] <- XP[i,4] * ntot[i,j]
      dev.n.N4a[(i-lowN4+1),(j-nf[i]+1),1]
                  <- 2* (n[i,j,4]*(log(n[i,j,4])-log(nhat[i,j,4]))
                     + (ntot[i,j]-n[i,j,4])*(log(ntot[i,j]-n[i,j,4])
                     - log(ntot[i,j]-nhat[i,j,4])))
}
 }
```

```
# residual deviance for N4 r and n (age and severity)
sumdev.r.N4 <- sum(dev.r.N4[,,])
sumdev.n.N4s <- sum(dev.n.N4s[,,])
sumdev.n.N4a <- sum(dev.n.N4a[,,])




####################################
# old age only + severity (N4old) #
####################################

for (i in (N6+N4+1):(N6+N4+N4old)){

   # probabilities in each severity marginal
   for (k in 1:3){
      XP3[i,k] <- pi[i,k*2] / XP[i,5]
   }

   for (j in nf[i]:na[i]){
      for (k in 1:3){

         # model events as Binomial
         r[i,j,k] ~ dbin(LAM[i,j,k*2],n[i,j,k])

         # r deviance contribution
         rhat[i,j,k] <- LAM[i,j,k*2] * n[i,j,k]
         dev.r.N4old[(i-lowN4old+1),(j-nf[i]+1),k]
                  <- 2* (r[i,j,k]*(log(r[i,j,k])-log(rhat[i,j,k]))
                      + (n[i,j,k]-r[i,j,k])*(log(n[i,j,k]-r[i,j,k])
                      - log(n[i,j,k]-rhat[i,j,k])))
      }

      # model n as multinomial
      n[i,j,1:3] ~ dmulti(XP3[i,1:3],ntot[i,j])

      for (k in 1:3){
         # n deviance contribution
         nhat[i,j,k] <- XP3[i,k] * ntot[i,j]
         dev.n.N4old[(i-lowN4old+1),(j-nf[i]+1),k]
```

```
                             <- 2* n[i,j,k]*log(n[i,j,k]/nhat[i,j,k])
      }
    }
}


# residual deviance for N4old r and n
sumdev.r.N4old <- sum(dev.r.N4old[,,])
sumdev.n.N4old <- sum(dev.n.N4old[,,])



#######################
# severity only (N3) #
#######################

for (i in (N6+N4+N4old+1):(N6+N4+N4old+N3)){
    for (j in nf[i]:na[i]){
       for (k in 1:3){

          # model events as Binomial
          r[i,j,k] ~ dbin(XAM[i,j,k],n[i,j,k])

          # r deviance contribution
          rhat[i,j,k] <- XAM[i,j,k] * n[i,j,k]
          dev.r.N3[(i-lowN3+1),(j-nf[i]+1),k]
                       <- 2* (r[i,j,k]*(log(r[i,j,k])-log(rhat[i,j,k]))
                           + (n[i,j,k]-r[i,j,k])*(log(n[i,j,k]-r[i,j,k])
                           - log(n[i,j,k]-rhat[i,j,k])))
       }

       # model n as multinomial
       n[i,j,1:3] ~ dmulti(XP[i,1:3],ntot[i,j])

       for (k in 1:3){
          # n deviance contribution
          nhat1[i,j,k] <- XP[i,k]*ntot[i,j]
          dev.n.N3[(i-lowN3+1),(j-nf[i]+1),k]
                      <- 2* n[i,j,k]*log(n[i,j,k]/nhat[i,j,k])
       }
    }
```

```
}

# residual deviance for N3 r and n
sumdev.r.N3 <- sum(dev.r.N3[,,])
sumdev.n.N3 <- sum(dev.n.N3[,,])



#################################
# mild and moderate only (N3mm) #
#################################

for (i in (N6+N4+N4old+N3+1):(N6+N4+N4old+N3+N3mm)){

    # probability in mild
    XPmm[i] <- XP[i,1]/(XP[i,1] + XP[i,2])

    for (j in nf[i]:na[i]){
        for (k in 1:2){

            # model events as Binomial
            r[i,j,k] ~ dbin(XAM[i,j,k],n[i,j,k])

            # r deviance contribution
            rhat[i,j,k] <- XAM[i,j,k] * n[i,j,k]
            dev.r.N3mm[(i-lowN3mm+1),(j-nf[i]+1),k]
                    <- 2* (r[i,j,k]*(log(r[i,j,k])-log(rhat[i,j,k]))
                        + (n[i,j,k]-r[i,j,k])*(log(n[i,j,k]-r[i,j,k])
                        - log(n[i,j,k]-rhat[i,j,k])))
        }

        # model n (mild) as binomial
        n[i,j,1] ~ dbin(XPmm[i],ntot[i, j])

        # n deviance contribution
        nhat[i,j,1] <- XPmm[i]*ntot[i,j]
        dev.n.N3mm[(i-lowN3mm+1),(j-nf[i]+1),1]
                  <- 2* (n[i,j,1]*(log(n[i,j,1])-log(nhat[i,j,1]))
                      + (ntot[i,j]-n[i,j,1])*(log(ntot[i,j]-n[i,j,1])
                      - log(ntot[i,j]-nhat[i,j,1])))
```

```
      }
   }


   # residual deviance for N3mm r and n
   sumdev.r.N3mm <- sum(dev.r.N3mm[,,])
   sumdev.n.N3mm <- sum(dev.n.N3mm[,,])



   #########################
   # moderate only (N3mod) #
   #########################

   for (i in (N6+N4+N4old+N3+N3mm+1):(N6+N4+N4old+N3+N3mm+N3mod)){
      for (j in nf[i]:na[i]){

         # model events as Binomial
         r[i,j,2] ~ dbin(XAM[i,j,2],n[i,j,2])

         # r deviance contribution
         rhat[i,j,2] <- XAM[i,j,2] * n[i,j,2]
         dev.r.N3mod[(i-lowN3mod+1),(j-nf[i]+1),1]
                     <- 2* (r[i,j,2]*(log(r[i,j,2])-log(rhat[i,j,2]))
                         + (n[i,j,2]-r[i,j,2])*(log(n[i,j,2]-r[i,j,2])
                         - log(n[i,j,2]-rhat[i,j,2])))
      }
   }

   # residual deviance for N3mod r
   sumdev.r.N3mod <- sum(dev.r.N3mod[,,])



   ###################################
   # moderate and severe only (N3ms) #
   ###################################

   for (i in (N6+N4+N4old+N3+N3mm+N3mod+1):(N6+N4+N4old+N3+N3mm+N3mod+N3ms)){

      # probability in moderate
      XPms[i] <- XP[i,2]/(XP[i,2] + XP[i,3])
```

```
    for (j in nf[i]:na[i]){
       for (k in 2:3){

          # model events as Binomial
          r[i,j,k] ~ dbin(XAM[i,j,k],n[i,j,k])

          # r deviance contribution
          rhat[i,j,k] <- XAM[i,j,k] * n[i,j,k]
          dev.r.N3ms[(i-lowN3ms+1),(j-nf[i]+1),k-1]
                     <- 2* (r[i,j,k]*(log(r[i,j,k])-log(rhat[i,j,k]))
                         + (n[i,j,k]-r[i,j,k])*(log(n[i,j,k]-r[i,j,k])
                         - log(n[i,j,k]-rhat[i,j,k])))
       }

       # model n (moderate) as binomial
       n[i,j,2] ~ dbin(XPms[i],ntot[i,j])

       # n deviance contribution
       nhat[i,j,2] <- XPms[i] * ntot[i,j]
       dev.n.N3ms[(i-lowN3ms+1),(j-nf[i]+1),1]
                  <- 2* (n[i,j,2]*(log(n[i,j,2])-log(nhat[i,j,2]))
                      + (ntot[i,j]-n[i,j,2])*(log(ntot[i,j]-n[i,j,2])
                      - log(ntot[i,j]-nhat[i,j,2])))
    }
}

# residual deviance for N3mm r and n
sumdev.r.N3ms <- sum(dev.r.N3ms[,,])
sumdev.n.N3ms <- sum(dev.n.N3ms[,,])


################
# age only (N2) #
################

for (i in (N6+N4+N4old+N3+N3mm+N3mod+N3ms+1):
                             (N6+N4+N4old+N3+N3mm+N3mod+N3ms+N2)) {
    for (j in nf[i]:na[i]){
```

```
    for (k in 1:2){

        # model events as Binomial
        r[i,j,k] ~ dbin(XAM[i,j,3+k],n[i,j,k])

        # r deviance contribution
        rhat[i,j,k] <- XAM[i,j,3+k] * n[i,j,k]
        dev.r.N2[(i-lowN2+1),(j-nf[i]+1),k]
                    <- 2* (r[i,j,k]*(log(r[i,j,k])-log(rhat[i,j,k]))
                        + (n[i,j,k]-r[i,j,k])*(log(n[i,j,k]-r[i,j,k])
                        - log(n[i,j,k]-rhat[i,j,k])))
        }

        # model n (young) as binomial
        n[i,j,1] ~ dbin(XP[i,4],ntot[i,j])

        # n deviance contribution
        nhat[i,j,1] <- XP[i,4] * ntot[i,j]
        dev.n.N2[(i-lowN2+1),(j-nf[i]+1),1]
                <- 2* (n[i,j,1]*(log(n[i,j,1])-log(nhat[i,j,1]))
                    + (ntot[i,j]-n[i,j,1])*(log(ntot[i,j]-n[i,j,1])
                    - log(ntot[i,j]-nhat[i,j,1])))
    }
}

# residual deviance for N2 r and n
sumdev.r.N2 <- sum(dev.r.N2[,,])
sumdev.n.N2 <- sum(dev.n.N2[,,])


#####################
# no breakdown (N1) #
#####################

for (i in (N6+N4+N4old+N3+N3mm+N3mod+N3ms+N2+1):
                            (N6+N4+N4old+N3+N3mm+N3mod+N3ms+N2+N1)){
    for (j in nf[i]:na[i]){

        # model events as Binomial
```

```
      r[i,j,1] ~ dbin(XAM[i,j,6],n[i,j,1])

      # r deviance contribution
      rhat[i,j,1] <- XAM[i,j,6] * n[i,j,1]
      dev.r.N1[(i-lowN1+1),(j-nf[i]+1),1]
                  <- 2* (r[i,j,1]*(log(r[i,j,1])-log(rhat[i,j,1]))
                      + (n[i,j,1]-r[i,j,1])*(log(n[i,j,1]-r[i,j,1])
                      - log(n[i,j,1]-rhat[i,j,1])))
   }
}


# residual deviance for N1 r
sumdev.r.N1 <- sum(dev.r.N1[,,])



# model treatment effect
for (i in 1:ns){
   w[i, 1] <- 0
   delta[i,t[i,1]] <- 0
   mu[i] ~ dnorm(0,0.0001)

   # delta prior, mixed effect comparison and multi-arm adjustment
   for (j in 2:na[i]){
      # trial-specific LOR distributions
      delta[i,t[i,j]] ~ dnorm(md[i,t[i,j]], dprec[i,t[i,j]])
      # mean of LOR distributions
      md[i,t[i,j]] <- d[t[i,j]] - d[t[i,1]] + sw[i,j]
      # cumulative adjustment for multi-arm trials
      sw[i,j] <- sum(w[i,1:j-1])/(j-1)
      # adjustment, multi-arm trials
      w[i,j] <- (delta[i,t[i,j]] - d[t[i,j]] + d[t[i,1]])
      # precision of LOR distributions
      dprec[i,t[i,j]] <- prec*2*(j-1)/j
   }
}

# vague priors for delta parameters
d[1]<-0
for (j in 2:nt){
```

```
      d[j] ~ dnorm(0,0.0001)
   }


   prec ~ dgamma(0.01,0.01)
   tau.sq <- 1/prec
   tau <- sqrt(tau.sq)


   # vague priors for age and severity parameters
   for (i in 1:ns){
      ddA[i] ~ dnorm(dA,cprec)
      ddS2[i] ~ dnorm(dS2,cprec)
      ddS3[i] ~ dnorm(dS3,cprec)
   }


   dA ~ dnorm(0,0.001)
   dS2 ~ dnorm(0,0.001)
   dS3 ~ dnorm(0,0.001)


   cprec ~ dgamma(0.01,0.01)
   ctau.sq <- 1/cprec
   ctau <- sqrt(ctau.sq)


   # ranking treatment means to find best intervention
   for (j in 1:6){
      pbest[j]<- equals(rank(d[],j),1)
   }


   # residual deviance for all r
   resdev.r <- sumdev.r.N6 + sumdev.r.N4 + sumdev.r.N4old + sumdev.r.N3
               + sumdev.r.N3mm + sumdev.r.N3mod + sumdev.r.N3ms
               + sumdev.r.N2 + sumdev.r.N1
   # residual deviance for all n
   resdev.n <- sumdev.n.N6 + sumdev.n.N4s + sumdev.n.N4a + sumdev.n.N4old
               + sumdev.n.N3 + sumdev.n.N3mm + sumdev.n.N3ms + sumdev.n.N2
   # total residual deviance
   resdev <- resdev.r + resdev.n


}
```

Bibliography

Ades, A. E. and Sutton, A. J. (2006). Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169**(1), 5–35.

Aitken, P. D., Rodgers, H., French, J. M., Bates, D. and James, O. F. W. (1993). General medical or geriatric unit care for acute stroke? A controlled clinical trial. *Age Ageing* **22**(suppl 2), 4–5.

Altman, D. G., Lausen, B., Sauerbrei, W. and Schumacher, M. (1994). Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute* **86**(11), 829–835.

Armitage, P. and Colton, T. (1998). *Encyclopedia of Biostatistics*. John Wiley & Sons. Chichester.

Bamford, J., Dennis, M., Sandercock, P., Burn, J. and Warlow, C. (1990). The frequency, causes and timing of death within 30 days of a first stroke: the Oxfordshire Community Stroke Project. *Journal of Neurology, Neurosurgery & Psychiatry* **53**(10), 824–829.

Bamford, J., Sandercock, P., Dennis, M., Burn, J. and Warlow, C. (1991). Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet* **337**(8756), 1521 – 1526.

Barber, M., Wright, F., Stott, D. J. and Langhorne, P. (2004). Predictors of early neurological deterioration after ischaemic stroke: a case-control study. *Gerontology* **50**(2), 102–109.

Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F. and Colditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine* **17**(22), 2537–2550.

Bland, J. M. and Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* **8**(2), 135–160.

Boissel, J. P., Blanchard, J., Panak, E., Peyrieux, J. C., Sacks, H. and the Meta-analysis Cooperative Group (1989). Considerations for the meta-analysis of randomized clinical trials: Summary of a panel discussion. *Controlled Clinical Trials* **10**(3), 254–281.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**(4), 434–455.

Brott, T., Adams Jr, H. P., Olinger, C. P., Marler, J. R., Barsan, W. G., Biller, J., Spilker, J., Holleran, R., Eberle, R. and Hertzberg, V. (1989). Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* **20**(7), 864–870.

Cabral, N. L., Moro, C., Silva, G. R., Scola, R. H. and Werneck, L. C. (2003).

Study comparing the stroke unit outcome and conventional ward treatment: a randomised study in Joinville, Brazil. *Arquivos de Neuro-psiquatria* **61**(2-A), 188–193.

Cavallini, A., Micieli, G., Marcheselli, S. and Quaglini, S. (2003). Role of monitoring in management of acute ischemic stroke patients. *Stroke* **34**(11), 2599–2603.

Cicchetti, D. V. and Allison, T. (1971). A new procedure for assessing reliability of scoring EEG sleep recordings. *American Journal of EEG Technology* **11**, 101–109.

Claxton, K. (2008). Exploring uncertainty in cost-effectiveness analysis. *PharmacoEconomics* **26**(9), 781–798.

Claxton, K., Sculpher, M., McCabe, C., Briggs, A., Akehurst, R., Buxton, M., Brazier, J. and O'Hagan, T. (2005). Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics* **14**(5), 339–347.

Cochrane Collaboration (2003). Review Manager (RevMan). Version 4.2 for Windows. http://www.cc-ims.net/RevMan.

Cochrane Collaboration (2005). Review Manager 4.2 User guide.

Cohen, J. (1968). Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin* **70**(4), 213–220.

Connor, R. J. (1972). Grouping for testing trends in categorical data. *Journal of the American Statistical Association* **67**(339), 601–604.

Côté, R., Battista, R. N., Wolfson, C., Boucher, J., Adam, J. and Hachinski, V. (1989). The Canadian Neuological Scale: validation and reliability assessment. *Neurology* **39**(5), 638–643.

Cox, D. R. (1957). Note on grouping. *Journal of the American Statistical Association* **52**(280), 543–547.

Curtis, L. (2007). Unit costs in health and soical care 2007. http://www.pssru.ac.uk/pdf/uc/uc2007/uc2007.pdf.

Davis, M., Chambers, B., Barer, D. and Birschel, P. (2000). A pilot study of monitoring in acute stroke. Proceedings of the 4th World Stroke Congress.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**(3), 177–188.

Dey, P., Woodman, M., Gibbs, A., Steele, R., Stocks, S. J., Wagstaff, S., Khanna, V. and Chaudhuri, M. D. (2005). Early assessment by a mobile stroke team: a randomised controlled trial. *Age Ageing* **34**(4), 331–338.

Dickersin, K. and Berlin, J. A. (1992). Meta-analysis: state-of-the-science. *Epidemiologic Reviews* **14**(1), 154–176.

Dominici, F. (2000). Combining contingency tables with missing dimensions. *Biometrics* **56**(2), 546–553.

Dorman, P., Dennis, M. and Sandercock, P. (2000). Are the modified "simple questions" a valid and reliable measure of health related quality of life after stroke?. *Journal of Neurology, Neurosurgery & Psychiatry* **69**(4), 487–493.

Dromerick, A. W., Edwards, D. F. and Diringer, M. N. (2003). Sensitivity to changes in disability after stroke: a comparison of four scales useful in clinical trials. *Journal of Rehabilitation Research and Development* **40**(1), 1 – 8.

Early Supported Discharge Trialists (2008). Services for reducing duration of hospital care for actue stroke patients. *Cochrane Database of Systematic Reviews. Issue 1.*

Evers, S. M. A. A., Ament, A. J. H. A. and Blaauw, G. (2000). Economic evaluation in stroke research: a systematic review. *Stroke* **31**(5), 1046–1053.

Fagerberg, B., Claesson, L., Gosman-Hedstrom, G. and Blomstrand, C. (2000). Effect of acute stroke unit care integrated with care continuum versus conventional treatment: a randomized 1-year study of elderly patients: The Goteborg 70+ Stroke Study. *Stroke* **31**(11), 2578–2584.

Feldman, D. J., Lee, P. R., Unterecker, J., Lloyd, K., Rusk, H. A. and Toole, A. (1962). A comparison of functionally orientated medical care and formal rehabilitation in the management of patients with hemiplegia due to cerebrovascular diseaes. *Journal of Chronic Diseases* **15**, 297–310.

Garraway, W. M., Akhtar, A. J., Hockey, L. and Prescott, R. (1980). Management of acute stroke in the elderly: follow up of a controlled trial. *British Medical Journal* **281**(6244), 827–829.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian data analysis. Texts in Statistical Science.* Chapman & Hall. London.

Gordon, E. E. and Kohn, K. H. (1966). Evaluation of rehabilitation methods in the hemiplegic patient. *Journal of Chronic Diseases* **19**(1), 3–16.

Govan, L. and for the Stroke Monitoring Trialists (2007). A meta-analysis of clinical trials of routine automated physiological monitoring in acute stroke. *Cerebrovascular Diseases* **23**(suppl 2), 7.

Govan, L., Langhorne, P., Weir, C. J. and for the Stroke Unit Trialists' Collaboration (2006). The prevention of complications may explain the survival benefit of organised (stroke unit) care: a secondary analysis of a systematic review. *Cerebrovascular Diseases* **21**(suppl 4), 30.

Govan, L., Langhorne, P., Weir, C. J. and for the Stroke Unit Trialists' Collaboration (2007*a*). Does the prevention of complications explain the survival benefit of organized inpatient (stroke unit) care? Further analysis of a systematic review. *Stroke* **38**(9), 2536–2540.

Govan, L., Langhorne, P., Weir, C. J. and for the Stroke Unit Trialists' Collaboration (2007*b*). Organised inpatient (stroke unit) care for stroke. *Cerebrovascular Diseases* **23**(suppl 2), 13.

Govan, L., Weir, C. J., Langhorne, P., Ades, A. E. and for the Stroke Unit Trialists' Collaboration (2008). Comparisons of different systems of organised (stroke unit) care: a secondary analysis of systematic review. *Cerebrovascular Diseases* **25**(suppl 2), 60.

Govan, L., Weir, C. J., Langhorne, P. and for the Stroke Unit Trialists' Collaboration (2008). Organized inpatient (stroke unit) care for stroke. *Stroke* **39**(8), 2402–2403.

Govan, L., Weir, C. J., Muir, K. W., Lees, K. R. and for the IMAGES Study Group (2006). Does the IMAGES Stroke Scale provide more information than a simple clinical classification?. *Cerebrovascular Diseases* **21**(suppl 4), 125.

Greiner, M., Pfeiffer, D. and Smith, R. D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine* **45**(1-2), 23–41.

Hamrin, E. (1982). Early activation in stroke - does it make a difference?. *Scandinavian Journal of Rehabilitation Medicine* **14**(3), 101–109.

Hankey, G. J., Deleo, D. and Stewart-Wynne, E. G. (1997). Stroke units: an Australian perspective. *Australian and New Zealand Journal of Medicine* **27**(4), 437–438.

Higgins, J. P. T. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**(11), 1539–1558.

Higgins, J. P. T. and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* **15**(24), 2733–2749.

Higgins, J. P. T., Whitehead, A., Turner, R. M., Omar, R. Z. and Thompson, S. G. (2001). Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* **20**(15), 2219–2241.

Ilmavirta, M. (1994). Stroke unit and outcome of brain infarction. Doctor of medicine. University of Tampere. Tampere, Finland.

Indredavik, B., Bakke, F., Solberg, R., Rokseth, R., Haaheim, L. and Holme, I. (1991). Benefit of a stroke unit: a randomized controlled trial. *Stroke* **22**(8), 1026–1031.

Information Services Division, NHS Scotland (2007). Scottish Health Service Costs (Costs Book). http://www.isdscotland.org/isd/costs-overview.jsp?pContentID=797&p_applic=CCC&p_service=Content.show&.

Intravenous Magnesium Efficacy in Stroke (IMAGES) Study Investigators (2004). Magnesium for acute stroke (Intravenous Magnesium Efficacy in Stroke trial): randomised controlled trial. *Lancet* **363**(9407), 439–445.

Isard, P. A. and Forbes, J. F. (1992). The cost of stroke to the National Health Service in Scotland. *Cerebrovascular Diseases* **2**(1), 47–50.

Jennett, B. and Teasdale, G. (1977). Aspects of coma after severe head injury. *Lancet* **1**(8017), 878–881.

Juby, L. C., Lincoln, N. B. and Berman, P. (1996). The effect of a stroke rehabilitation unit on functional and psychological outcome. *Cerebrovascular Diseases* **6**(2), 106–110.

Kalra, L. and Eade, J. (1995). Role of stroke rehabilitation units in managing severe disability after stroke. *Stroke* **26**(11), 2031–2034.

Kalra, L., Dale, P. and Crome, P. (1993). Improving stroke rehabilitation. A controlled study. *Stroke* **24**(10), 1462–1467.

Kalra, L., Evans, A., Perez, I., Knapp, M., Donaldson, N. and Swift, C. G. (2000).

Alternative strategies for stroke care: a prospective randomised controlled trial. *Lancet* **356**(9233), 894–899.

Kasner, S. E. (2006). Clincal interpretation and use of stroke scales. *Lancet Neurology* **5**(7), 603–612.

Kaste, M., Palomaki, H. and Sarna, S. (1995). Where and how should elderly stroke patients be treated? A randomized trial. *Stroke* **26**(2), 249–253.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R. and Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distribution in MCMC using WinBUGS. *Statistics in Medicine* **24**(15), 2401–2428.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**(1), 159–174.

Langhorne, P. (2003-2007). Glasgow Royal Infirmary stroke services report.

Langhorne, P. and Dennis, M. S. (2004). Stroke units: the next 10 years. *Lancet* **363**(9412), 834–835.

Langhorne, P. and Pollock, A. (2002). What are the components of effective stroke unit care?. *Age Ageing* **31**(5), 365–371.

Laursen, S. O., Henriksen, I. O., Dons, U., Jacobsen, B. and Gundertofte, L. (1995). Intensive rehabilitation following stroke: controlled pilot study. *Ugeskr Laeger* **157**, 1996–1999.

Lindley, R. I., Warlow, C. P., Wardlaw, J. M., Dennis, M. S., Slattery, J. and Sandercock, P. A. (1993). Interobserver reliability of a clinical classification of acute cerebral infarction. *Stroke* **24**(12), 1801–1804.

Lu, G. and Ades, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* **101**(474), 447–459.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**(4), 325–337.

Lyden, P. D. and Hanston, L. (1998). Assessment scales for the evaluation of stroke patients. *Journal of Stroke and Cerebrovascular Diseases* **7**(2), 113–127.

Lyden, P. D. and Lau, G. T. (1991). A critical appraisal of stroke evaluation and rating scales. *Stroke* **22**(11), 1345–1352.

Lyden, P. D., Lu, M., Levine, S. R., Brott, T. G., Broderick, J. and Cote, R. (2001). A modified National Institutes of Health Stroke Scale for use in stroke clinical trials: preliminary reliability and validity. *Stroke* **32**(6), 1310–1317.

Lyden, P., Lu, M., Jackson, C., Marler, J., Kothari, R., Brott, T. and Zivin, J. (1999). Underlying structure of the National Institutes of Health Stroke Scale: results of a factor analysis. *Stroke* **30**(11), 2347–2354.

Lyden, P., Raman, R., Liu, L., Grotta, J., Broderick, J., Olson, S., Shaw, S., Spilker, J., Meyer, B., Emr, M., Warren, M. and Marler, J. (2005). NIHSS

training and certification using a new digital video disk is reliable. *Stroke* **36**(11), 2446–2449.

Ma, R. H., Wang, Y. J., Qu, H. and Yang, Z. H. (2004). Assessment of the early effectiveness of a stroke unit in comparison to the general ward. *Chinese Medical Journal* **117**(6), 852–855.

Mahoney, F. I. and Barthel, D. W. (1965). Functional evaluation: the Barthel Index. *Maryland State Medical Journal* **14**, 61–65.

Malmgren, R., Bamford, J., Warlow, C., Sandercock, P. and Slattery, J. (1989). Projecting the number of patients with first ever strokes and patients newly hanidcapped by stroke in England and Wales. *British Medical Journal* **298**(6674), 656–660.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* **22**(4), 719–748.

Mead, G. E., Lewis, S. C., Wardlaw, J. M., Dennis, M. S. and Warlow, C. P. (2000). How well does the Oxfordshire Community Stroke Project classification predict the site and size of the infarct on brain imaging?. *Journal of Neurology, Neurosurgery & Psychiatry* **68**(5), 558–562.

Morris, R. W., Whincup, P. H., Emberson, J. R., Lampe, F. C., Walker, M. and Shaper, A. G. (2003). North-south gradients in Britain for stroke and CHD: are they explained by the same factors?. *Stroke* **34**(11), 2604–2609.

Muir, K. W., Weir, C. J., Murray, G. D., Povey, C. and Lees, K. R. (1996). Comparison of neurological scales and scoring systems for acute stroke prognosis. *Stroke* **27**(10), 1817–1820.

Murray, C. J. L. and Lopez, A. D. (1997). Mortaility cause for eight regions of the world: Global Burden of Disease Study. *Lancet* **349**(9061), 1269–1276.

National Institutes for Health and Clinical Excellence (June 2008). Guide to the methods of technology appraisal.

Patel, N. (2000). Organised care of acute stroke at Goort Schuur Hospital. MPhil in Epidemiology. University of Cape Town. Cape Town, South Africa.

Peacock, P. B., Riley, C. H. P., Lampton, T. D., Raffel, S. S. and Walker, J. S. (1972). *Trends in epidemiology: the Birmingham stroke, epidemiology and rehabilitation study.* Thomas. Springfield, Illinois. pp. 231–345.

Rankin, J. (1957). Cerebral vascular accidents in Patients over the age of 60: II. Prognosis. *Scottish Medical Journal* **2**(5), 200–215.

Rodgers, H., Dennis, M., Cohen, D. and Rudd, A. (2003). British Association of Stroke Physicians: benchmarking survey of stroke services. *Age Ageing* **32**(2), 211–217.

Ronning, O. M. and Guldvog, B. (1998). Stroke unit versus general medical wards, II: neurological deficits and activities of daily living - A quasi-randomized controlled trial. *Stroke* **29**(3), 586–590.

Rothman, K. J. and Greenland, S. (1998). *Modern Epidemiology (2nd edition).* Lippincott-Raven. Philadelphia.

Salanti, G., Higgins, J. P. T. and White, I. R. (2006). Bayesian synthesis of epidemiological evidence with different combinations of exposure groups: Application to a gene-gene-environment interaction. *Statistics in Medicine* **25**(24), 4147–4163.

Sandercock, P., Algra, A., Anderson, C., Bereczki, D., Berge, E., Bowen, A., Forster, A., Fraser, H., Hankey, G. J., Langhorne, P., Lewis, S., Liu, M., McInnes, A., Pollock, A., Prasad, K., Ricci, S., Sudlow, C. and Thomas, B. (2008). Cochrane Stroke Group. *About the Cochrane Collaboration (Cochrane Review Groups (CRGs))*.

SAS Institute Inc. (2007). Sample 25017: Nonparametric comparison of areas under correlated ROC curves. Knowledge Base for SAS Customer Support, http://support.sas.com/kb/25/017.html. Accessed 22/05/08.

Scandinavian Stroke Study Group (1985). Multicenter trial of hemodilution in ischemic stroke–background and study protocol. Scandinavian Stroke Study Group. *Stroke* **16**(5), 885–890.

Scandinavian Stroke Study Group (1987). Multicenter trial of hemodilution in acute ischemic stroke. I. Results in the total patient population. Scandinavian Stroke Study Group. *Stroke* **18**(4), 691–699.

Schmid, C. H. (2001). Using Bayesian inference to perform meta-analysis. *Evaluation and the Health Professions* **24**(2), 165–189.

Sellars, C., Bowie, L., Bagg, J., Sweeney, M. P., Miller, H., Tilston, J., Langhorne, P. and Stott, D. J. (2007). Risk factors for chest infection in acute stroke: a prospective cohort study. *Stroke* **38**(8), 2284–2291.

Sidik, K. and Jonkman, J. N. (2002). A simple confidence interval for meta-analysis. *Statistics in Medicine* **21**(21), 3153–3159.

Sivenius, J., Pyorala, K., Heinonen, O., Salonen, J. and Riekkinen, P. (1985). The significance of intensity of rehabilitation of stroke - a controlled trial. *Stroke* **16**(6), 928–931.

Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation.* John Wiley & Sons. Chichester.

Spiegelhalter, D. J., Best, N. G. and Carlin, B. P. (1998). Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. Technical report.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model compexity and fit. *Journal of the Royal Statistical Society* **64**(4), 583–639.

Stevens, R. S., Ambler, N. R. and Warren, M. D. (1984). A randomized controlled trial of a stroke rehabilitation ward. *Age Ageing* **13**(2), 65–75.

Strand, T., Asplund, K., Eriksson, S., Hagg, E., Lithner, F. and Wester, P. (1985). A non-intensive stroke unit reduces functional disability and the need for long-term hospitalization. *Stroke* **16**(1), 29–34.

Stroke Unit Trialists' Collaboration (1997*a*). Collaborative systematic review of the randomised trials of organised inpatient (stroke unit) care after stroke. *British Medical Journal* **314**(7088), 1151–1159.

Stroke Unit Trialists' Collaboration (1997*b*). How do stroke units improve patient outcomes? A collaborative systematic review of the randomized trials. *Stroke* **28**(11), 2139–2144.

Stroke Unit Trialists' Collaboration (2006). Organised inpatient (stroke unit) care for stroke. *Cochrane Database of Systematic Reviews. Issue 2.*

Stroke Unit Trialists' Collaboration (2007). Organised inpatient (stroke unit) care for stroke. *Cochrane Database of Systematic Reviews. Issue 4.*

Sulter, G., Elting, J. W., Langedijk, M., Maurits, N. M. and De Keyser, J. (2003). Admitting acute ischemic stroke patients to a stroke care monitoring unit versus a conventional stroke unit: a randomized pilot study. *Stroke* **34**(1), 101–104.

Sutton, A. J., Abrams, K. R. and Jones, D. R. (2001). An illustrated guide to the methods of meta-analysis. *Journal of Evaluation in Clinical Practice* **7**(2), 135–148.

Sutton, A. J. and Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* **10**(4), 277–303.

Sutton, A. J. and Higgins, J. P. T. (2008). Recent developements in meta-analysis. *Statistics in Medicine* **27**(5), 625–650.

Sutton, A. J., Kendrick, K. and Coupland, C. A. C. (2008). Meta-analysis of individual- and aggregate-level data. *Statistics in Medicine* **27**(5), 651–669.

Svensson, A., Harmsen, P. and Wilhelmsen, L. (1993). Goteborg Ostra stroke unit trial. Personnal Communication.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science* **240**(4857), 1285–1293.

Taylor, J. M. G. and Yu, M. (2002). Bias and efficiency loss due to categorizing an explanatory variable. *Journal of Multivariate Analysis* **83**(1), 248–263.

van Swieten, J. C., Koudstaal, P. J., Visser, M. C., Schouten, H. J. and van Gijn, J. (1988). Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* **19**(5), 604–607.

Vemmos, K., Takis, K., Madelos, D., Synetos, A., Volotasiou, V. and Tzavellas, H. (2001). Stroke unit treatment versus general medical wards: long term survival. *Cerebrovascular Diseases* **11**(suppl 4), 8.

von Arbin, M., Britton, M., deFaire, U., Helmers, C., Miah, K. and Murray, V. (1980). A study of stroke patients treated in a non-intensive stroke unit or in general medical wards. *Acta Medica Scandinavica* **208**(1-2), 81–85.

Warn, D. E., Thompson, S. G. and Spiegelhalter, D. G. (2002). Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Statistics in Medicine* **21**(11), 1601–1623.

Weir, N. U., Counsell, C. E., McDowall, M., Gunkel, A. and Dennis, M. S. (2003). Reliability of the variables in a new set of models that predict outcome after stroke. *Journal of Neurology, Neurosurgery & Psychiatry* **74**(4), 447–451.

Whitehead, A. and Jones, N. M. B. (1994). A meta-analysis of clinical trials involving different classifications of response into ordered categories. *Statistics in Medicine* **13**(23-24), 2503–2515.

Whitehead, A., Omar, R. Z., Higgins, J. P. T., Savaluny, E., Turner, R. M. and Thompson, S. G. (2001). Meta-analysis of ordinal outcomes using individual patient data. *Statistics in Medicine* **20**(15), 2243–2260.

WHO MONICA Project Principal Investigators (1988). The World Health Organisation MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. *Journal of Clincal Epidemiology* **41**(2), 105–114.

Wilson, J. L., Hareendran, A., Grant, M., Baird, T., Schulz, U. G., Muir, K. W. and Bone, I. (2002). Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin Scale. *Stroke* **33**(9), 2243–2246.

Wolfe, C. D. A. (2000). The impact of stroke. *British Medical Bulletin* **56**(2), 275–286.

Wood-Dauphinee, S., Shapiro, S., Bass, E., Fletcher, C., Georges, P., Hensby, V. and Mendelsohn, B. (1984). A randomized trial of team care following stroke. *Stroke* **15**(5), 864–872.

Yagura, H., Miyai, I., Suzuki, T. and Yanagihara, T. (2005). Patients with severe stroke benefit most by interdisciplinary rehabilitation team appraoch. *Cerebrovascular Diseases* **20**(4), 258–263.

Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. (1985). Beta-blockade during and after myocardial-infarction - an overview of the randomized trials. *Progress in Cardiovascular Diseases* **27**(5), 335–371.