Exploratory analyses to guide inclusion, limitation of sample size and strengthening of endpoints in clinical stroke trials

Rachael Louise Fulton

BSc (Hons), M.Sc (Statistics)

A thesis submitted in fulfilment of the requirements for the degree of doctor of philosophy

University of Glasgow

Institute of Cardiovascular and Medical Sciences

School of Medicine

College of Medical, Veterinary and Life Sciences

August 2013

Abstract

Clinical trials for treatment of acute ischaemic stroke require large numbers of patients and are expensive to conduct. Treatment is typically administered within the first hours or days after stroke onset. Outcome is usually assessed by a single measure, the most common being the modified Rankin scale (mRS) at day 90. Any strategy that can reduce cost or deliver more reliable answers on safety and efficacy of the investigational treatment would be welcome for future exploratory testing of novel treatments. This thesis focused on the impact of applying different methods of design, inclusion and outcome measurement to limit sample size and strengthen analysis in clinical trials in acute stroke.

Firstly, inclusion criteria were investigated to assess the impact on functional outcome. By assessing how the effect of thrombolysis changes over onset time to treatment (OTT) the relationship between OTT and age could be investigated. By looking across the entire range of OTT and assessing the interaction between the two covariates this provided complementary data to a previous VISTA analysis conducted by Mishra et al. It was found that across the full range of OTT, up to 3.5h, the treatment effect of thrombolysis in very elderly stroke patients (>80 years old) was comparable to that of their younger counterparts. The association of AF and modified Rankin Scale (mRS) at day 90 was then assessed. Multiple logistic regression analysis adjusted for age and baseline National Institutes of Health Stroke Scale (NIHSS) showed that history of AF had no independent impact on stroke outcome.

Deferred selection of subjects for neurorestorative therapies from hyperacute (<6h) to 24h was then explored using a simulation approach. The sample size required to detect a 'shift' in mRS outcome equivalent to a 5% absolute difference in proportion achieving mRS 0-2 versus 3-6 was modelled, setting power at 80% and assuming adjustment for entry age and NIHSS. It was found that extending the time window for patient selection provides a measurement which has a stronger more predictive relationship with outcome.

Trial inclusion was explored further by investigating selection for delayed treatment with thrombolysis. Prognostic scoring methods were proposed to identify a strategy for patient selection to be applied first to an existing trial dataset and then validated in the pooled RCT 4.5-6h data.). Prognostic score limits were chosen to optimise the sample for a net treatment benefit significant at p=0.01 by Cochran Mantel Haenszel test and by ordinal logistic regression. More inclusive limits were also defined based on p=0.05 criteria. After finalising prognostic score limits, for validation they were applied by an independent statistician to the pooled RCT data for 4.5-6h. The validation analysis based on ordinal outcomes failed to deliver a population in whom treatment >4.5h was safe and effective; analysis based on net benefit (mRS 0-1) showed significance.

Secondly, different strategies for endpoint selection were considered. In the past some trialists have investigated the use of earlier endpoints on single trial datasets and taken advantage of the fact that numerous outcome scales are available to measure various domains of neurological and functional recovery. The use of an earlier neurological endpoint for detecting futility in a trial was considered with validation on external RCT

data. Global endpoints, investigating different aspects of functional recovery at different time-points were then considered.

Simulations were undertaken to assess the relationship between sample size and power for ordinal scales and the corresponding global outcomes. Day 7 NIHSS was found to be the most sensitive individual ordinal endpoint. Dichotomised analyses supported these results. However this needed validation in a randomised trial dataset for use in exploratory stroke trials. The validation study reinforced the results from the non-randomised VISTA study. The global test combination of NIHSS90 with NIHSS7 appeared to offer incremental sensitivity to treatment effect compared to the ordinal scales alone. The combination of mRS90 with NIHSS7 did not increase the sensitivity to treatment effect when compared to NIHSS alone, but offers a broader clinical measure without loss of statistical power.

Finally, alternatives to the traditional RCT were considered. Abandoning the rigour of the blinded RCT carries substantial penalty in loss of reliability and should not be undertaken lightly.

If a placebo control is deemed impractical or unethical, investigators often consider comparisons against historical controls. A within-VISTA exploration of case-control matching is presented. The reliability of different matching methods and covariate combinations were assessed using a simulation approach. The results indicate that caution must be taken when using historical controls to generate a matched control group. Substantial further work matching to external data and validation to RCT data is needed.

Cluster randomised trials, which randomise patients by groups, are becoming a more widely used approach. When evaluating strategies to promote the transfer of research findings into clinical practice, i.e. in "Implementation Research", an RCT is impractical and a cluster randomised trial design is of advantage. Some elements in the design and sample size calculation of cluster randomised trials were considered. Intra cluster correlation coefficients (ICCs) were estimated from linear and generalised linear mixed models using maximum likelihood estimation for common measures used in stroke research. These estimates of relevant ICCs should assist in the design and planning of cluster randomised trials.

In conclusion, this research has shown that there are several areas in the design of clinical trials of acute stroke that merit further investigation. Several strategies have been highlighted that could potentially reduce sample size whilst retaining optimal levels of statistical power. However other aspects such as patient selection and the nature of the intervention under study can affect trial cost and statistical power and need to be taken under consideration.

Acknowledgements

I would like to thank my supervisor, Professor Kennedy Lees, to whom I am eternally grateful. Ken provided unreserved support during my PhD and has paved the way for my career in medical statistics. Perhaps most importantly, I thank Ken for his support during my heart surgery. His kindness and understanding helped me through a very difficult time. I would also like to thank my second supervisor Professor Stephen Senn for his guidance and support.

I am also greatly indebted to the many people who in some way contributed to the work within this thesis. First and foremost I thank my friend and colleague Dr Benedikt Frank and my hard working students Fraser Goldie and Daniel Kerr. I also appreciate the help provided by Dr Erich Bluhmki and Gabriele Biegert from Boehringer Ingelheim in the validation stages of some projects. Along with Asubio Pharmaceuticals, Johnson and Johnson and Wyeth/Pfizer for funding part of my PhD. Thank you also to Pamela Mackenzie, Elizabeth Colquhoun and Dorothy Ronney for their kindness and support over the last few years.

I would also like to thank Dr Myzoon Ali for her help and guidance particularly with the VISTA database; The VISTA steering committee, for their insight and assistance with proposals and manuscripts; Professor Matthew Walters and Dr Jesse Dawson for allowing me to take part in teaching and several important projects and finally my colleagues Rosalind Lees, Dr Karim Hajaar, Dr Kate McArthur and Dr Azmil Abdul-Rahim for allowing me to help with their projects and listening to me rant.

Finally my love goes to Dr Liam MacIsaac and my family whose support and encouragement has allowed me to complete this PhD, along with Elaine Rogers, Emma Armstrong, Dr David Reid, Dr Graham Morton, Dr James Buckle, Craig Spy, Dr Patricia Domingues, Alexander Mcdougall and Andrew Waugh for their friendship and laughter throughout my PhD.

Declaration

I declare that I am the sole author of this thesis entitled “Exploratory analyses to guide inclusion, limitation of sample size and strengthening of endpoints in clinical stroke trials.” This work has never previously been submitted for a higher degree. This work utilises anonymised data for tertiary analyses and is therefore exempt from Research Medical Ethics Approval.

The Onset time to treatment and Atrial Fibrillation projects presented in Chapter 3 and the ICC work presented in Chapter 9 were completed in collaboration with the VISTA research Fellow Dr Benedikt Frank.

The validation analysis presented in Chapter 5 was performed by external statisticians Dr Erich Bluhmki and Gabriele Biegert from Boehringer Ingelheim, Germany. This was for data security rules associated with the thrombolysis trial data.

The VISTA work presented in Chapter 6 was completed in collaboration with a BSc MedSci student, Daniel M Kerr as part of a summer project. The validation analysis in this chapter was performed by external statisticians Dr Erich Bluhmki and Gabriele Biegert from Boehringer Ingelheim, Germany. This was for data security rules associated with the thrombolysis trial data

The work presented in Chapter 7 was completed in collaboration with a BSc MedSci student Fraser C Goldie as part of a summer project.

All research was conducted at the Institute of Cardiovascular and Medical Sciences, University of Glasgow, under the supervision of Prof. K.R. Lees and Prof. S. Senn.

Rachael Fulton

Signature …….. Date …….

List of Publications

Papers

Kerr DM, Fulton RL, Higgins P, Bath PM, Shuaib A, Lyden P, et al. Response of blood pressure and blood glucose to treatment with recombinant tissue-type plasminogen activator in acute ischemic stroke: Evidence from the virtual international stroke trials archive. *Stroke*. 2011;10:10

Kerr DM, Fulton RL, Lees KR. Seven-day nihss is a sensitive outcome measure for exploratory clinical trials in acute stroke: Evidence from the virtual international stroke trials archive. *Stroke*. 2012;2:2

Lees KR, Bath PM, Schellinger PD, Kerr DM, Fulton R, Hacke W, Matchar, D, Sehra, R, Toni D. Contemporary outcome measures in acute stroke research: Choice of primary outcome measure. *Stroke*. 2012;43:1163-1170

Frank B, Fulton R, Weimar C, Shuaib A, Lees KR. Impact of atrial fibrillation on outcome in thrombolyzed patients with stroke: Evidence from the virtual international stroke trials archive (vista). *Stroke*. 2012;43:1872-1877

Fulton, R. L., K. R. Lees, et al. (2013). "Selection for delayed intravenous alteplase treatment based on a prognostic score." International Journal of Stroke: Jan 2013.

Goldie, F.C, Fulton, R.L, Dawson J, Bluhmki E, Lees KR. *(In Press International journal of stroke)* Exploration of time-course combinations of outcome scales for use in a global test of stroke recovery.

Frank, B, Fulton, R.L, Goldie, F.C, Hacke, W, Weimar, C, Diener, H,C, Lees, K,R. *(In Press International journal of stroke)* Intracluster Correlation Coefficients and Reliability of Randomised Multicentre Stroke Trials within VISTA

Conference proceedings

ISC 2011:(Moderated Poster presentation) Selection for delayed intravenous alteplase treatment based on prognostic score. R.L.Fulton, K.R.Lees, E.Bluhmki, G.Biegert, G.W.Albers, S.M.Davis, G.A.Donnan, J.Grotta, W.Hacke, M.Kaste, R.von.Kummer, A.Shuaib, D.Toni, on behalf of the VISTA Collaboration and ECASS, ATLANTIS and EPITHET Investigators

ESC 2011: (Oral presentation: Won the Young Investigators Award) Effect on study sample size of an extended time window for initiation of neurorestorative therapy after stroke: an exploratory analysis on the Virtual International Stroke Trials Archive (VISTA). R.L.Fulton, J.M.Gertner, P.D.Lyden, K.R.Lees.

ESC 2011: (Oral presentation) Response of Blood Pressure and Blood Glucose to Treatment With Recombinant Tissue-Type Plasminogen Activator in Acute Ischemic Stroke: Evidence From the Virtual International Stroke Trials Archive. Kerr, D. M. Fulton, R. L. Higgins, P. Bath, P. M. Shuaib, A. Lyden, P. Lees, K. R.

ISC 2012:(Poster presentation) Day 7 NIHSS is a Sensitive Outcome Measure for Exploratory Clinical Trials in Acute Stroke: Evidence from the Virtual International Stroke Trials Archive (VISTA). D.M.Kerr, R.L.Fulton, K.R.Lees for the VISTA collaborators.

ESC 2012: (Oral Presentation) Validation of 7 day NIHSS as a sensitive outcome measure for exploratory clinical trials in acute stroke. Evidence from pooled randomized control trials in Thrombolysis. R.L.Fulton, D.M.Kerr, E.Bluhmki, K.R.Lees, for the VISTA collaborators and the ECASS, NINDS, ATLANTIS and EPITHET investigators.

ESC 2012: (Oral Presentation) Impact of Atrial Fibrillation on Outcome in Thrombolysed Stroke Patients: Evidence from the Virtual International Stroke Trials Archive (VISTA). Benedikt Frank, Rachael Fulton, Christian Weimar, Ashfaq Shuaib, Kennedy R. Lees, for the VISTA Collaborators.

ESC 2012: (Poster presentation) Interdependence of Stroke Outcome Scales: availability of reliable estimates from the Virtual International Stroke Trials Archive (VISTA). Fraser C Goldie, Rachael Fulton, Benedikt Frank, Kennedy R Lees for the VISTA Collaborators.

ISC 2013: (Poster Presentation) Exploration of Time-course Combinations of Outcome Scales for Use in a Global Test of Stroke Recovery. Fraser C Goldie, Rachael L Fulton, Jesse Dawson, Kennedy R Lees for the VISTA Collaboration

ISC 2013: (Poster Presentation) The Effect of Time to Treatment on Outcome in Very Elderly Thrombolysed Stroke Patients. Benedikt Frank, Rachael L Fulton, Kennedy R Lees on behalf of the VISTA Collaboration.

ISC 2013: (Moderated poster presentation) Intracluster Correlation Coefficients and Reliability of Randomised Multicentre Stroke Trials within VISTA. Benedikt Frank, Rachael L Fulton, Fraser C Goldie, Werner Hacke, Christian Weimar, Kennedy R Lees on behalf of the VISTA Collaboration

Contents

List of Figures

List of Tables

Chapter 1
Introduction and background

1.1. What is Stroke?

1.1.1. Definition of Stroke

The World Health Organisation (WHO) defines acute Stroke as "rapidly developing clinical signs of focal (or global) disturbance of cerebral function, with symptoms lasting 24 hours or longer leading to death, with no apparent cause other than of vascular origin."[1] This definition includes stroke due to cerebral infarction (ischaemic), resulting from an arterial or venous blockage, and haemorrhagic strokes, resulting from a bleed[2]. Applying this definition excludes transient ischaemic attack (TIA), defined to last less than 24 hours, and patients with haematoma, haemorrhage or infarction caused by infection or tumor[3]. Stroke is the result of disruption of blood supply to the brain, occluding the supply of oxygen and nutrients, causing damage to brain tissue[4]. Due to the anatomy of the blood supply to the brain, certain patterns of focal neurological symptoms such as loss of body function, weakness down one side, visual problems, dysphasia and dysarthria are similar from patient to patient[5].

1.1.2. Pathology of stroke

Stroke can be due to an ischaemic or haemorrhagic disturbance of cerebral blood flow. Stroke can be classified into three groups; ischaemic, primary intracerebral haemorrhage (PICH) and certain instances of subarachnoid haemorrhage (SAH)[6]. As ischaemic and

haemorrhagic strokes often present in a similar manner but the causes and treatment are different, CT or MRI imaging is required to reliably distinguish between the two[2 5].

An ischaemic stroke or cerebral infarction accounts for approximately 85% of strokes[7-8] and occurs due to a critical reduction of the blood supply to the brain[5]. Ischemia can be sub-divided into three different categories: thrombosis , embolism and decreased system perfusion[9]. The location, extent and shape of the infarct depends on the size of the blocked vessel, mechanism of the obstruction and the compensatory capacity of the vascular region[10]. Thrombosis is the cause of approximately 50% of ischaemic strokes[5], leading to obstruction of the blood flow, often at sites of narrowing or partial occlusion of the blood vessels damaged by artheroma[5]. Embolic cerebral infarction is caused by material formed elsewhere in the vascular system lodging in an artery and causing a blockage to blood flow[9]. Decreased systemic perfusion is due to low systemic arterial pressure causing diminished flow to the brain[9].

Two prominent classifications for ischaemic stroke come from the criteria used in the Trial of Org 10172 in Acute Stroke Treatment (TOAST)[11] and the Oxfordshire Community Stroke Project classification (OCSP)[12]. The TOAST subtype classification system uses both clinical features, laboratory testing and CT/MRI imaging to categorise a stroke into one of five categories: large-artery atherosclerosis, cardio embolism, small artery occlusion (lacune), stroke of other determined aetiology and stroke of undetermined aetiology[11]. It can generally only be applied some days after hospitalisation. The OCSP classification system describes only four subtypes. It is based on a purely clinical assessment[13] with the subtypes defined as; total anterior circulation (TAC), partial anterior circulation (PAC), lacunar (LAC) and posterior circulation (POC), a final letter can be added as S for

syndrome or I for infarct[14]. Other clinical scales can be useful but many are very detailed and hence time consuming to apply[13].

Intracerebral haemorrhage (ICH) reflects bleeding directly into the brain[9]. ICH accounts for approximately 10-15% of all strokes.[8 10 15] It and is often caused by elevated blood pressure often referred to as hypertension damaging or weakening the blood vessels[5 9]. ICH can also be caused by lack of blood clotting factors, often due to prior anticoagulant medication such as Warfarin[5]. The degree of damage from ICH depends on the location, volume and pressure of the bleeding[9]. ICH is characterised by a severe prognosis, particularly in the short term and there is no specific targeted therapy available[16-17].

Subarachnoid haemorrhage (SAH) accounts for approximately 3% of all strokes[18]. In subarachnoid haemorrhage (SAH), blood leaks out onto the surface of the brain and is dispersed into the subarachnoid spaces around the brain[9]. Approximately 75% of SAH are caused by an aneurysm. SAH is a significant cause of morbidity and mortality worldwide[19]. However approximately 15% of patients presenting with SAH have no obvious source of bleeding. This is known as a nonaneurysmal SAH (NA-SAH).[20] NA-SAH carries a more favourable prognosis than aneurysmal SAH[20].

1.1.3. Pathophysiology of stroke

The brain is a metabolically active organ which uses approximately 25% of the body's energy supply. The sole substrates for energy metabolism in the brain are glucose and oxygen; the brain requires approximately 100 mg of glucose and 500mL of oxygen each minute[21]. This means that the brain requires oxygenated blood containing adequate sugar. Cerebral blood flow is closely related to metabolic regulation but it is important for the survival of brain tissue that the flow rate remains reasonably constant despite the

blood pressure changing, this is controlled by a process known as autoregulation[9]. When an ischaemic stroke occurs, decreased or absent circulating blood deprives neurons of the necessary substrates, glucose and oxygen.

Because cerebral blood flow is closely related to cerebral metabolism, autoregulation may compensate for relative ischemia by increasing the extraction of oxygen and glucose from the blood. However, when cerebral blood flow is severely diminished, cell membranes and functions are severely affected and neurons cannot survive for long. This can result in irreversible injury. The progression and extent of ischemic injury can be influenced by diverse factors such as rate of onset and duration of ischaemia, collateral circulation, health of the systemic circulation, haematological factors, temperature and glucose metabolism[21]. While irreversible damage will occur in any deeply ischaemic area there may be an adjacent zone known as the penumbra in which blood flow is less severely reduced and where the glial and neuronal tissue survives for longer[22-23]. This can persist for longer than the acute phase of the stroke[24].

In contrast ICH causes damage to brain tissue by disrupting connecting pathways and causing localised pressure injury[21 23]. The pathophysiology of an ICH is an acute space occupying lesion compressing and disrupting the surrounding tissue. This causes an increase in intracranial pressure and may lead to herniation; moving brain tissues, cerebrospinal fluid and blood vessels inside the skull[10].

1.1.4. Signs and symptoms

Clinical features that arise depend on the normal function of the specific part of the brain that has lost blood supply. Common symptoms and signs include weakness of one side of the body (hemiparesis), difficulty with swallowing (dysphagia), incoordination (ataxia),

difficulty understanding or expressing speech (dysphasia), slurred speech (dysarthria) and problems with vision such as loss of vision to one side affecting both eyes (hemianopia). Other symptoms such as confusion, weakness and blackouts are also common but are non-focal[2]. Pancioli et al conducted a telephone survey looking at public awareness of stroke risk factors and warning signs and risk factors of stroke. Only 57% of respondents could name one of the 5 established warning signs and only 68% of respondents could name at least one risk factor for stroke[25]. To increase public awareness of stroke in the UK, and encourage rapid medical intervention in acute stroke, the Stroke Association set up the FAST campaign[26].

The FAST campaign concentrates on three main symptoms Face, Arm and Speech: can the person smile? Has their face drooped? Can the person raise both arms? Can the person speak properly? The final letter denotes Time. This is a crucial factor as not only do brain cells die over time but for ischaemic stroke, treatment is most effective when given early[27]. In 2012 Robinson et al conducted a new survey to assess knowledge of symptoms and awareness of the FAST campaign. It was found that 70% of the population were aware of the FAST campaign and over 80% of respondents were aware of the individual signs[28].

1.1.5. Investigation

The diagnosis of stroke is first made clinically based on the neurological symptoms. If the symptoms are focal, have a negative effect (i.e. loss of vision, loss of movement rather than positive such as pain) and develop suddenly rather than over a period of hours or days the likelihood of a stroke is high. Once a patient has been admitted to hospital, a detailed medical history is taken to assess for prior risk factors. Because acute therapies for stroke have a very narrow time window assessment of neurological impairment needs

to be performed quickly; the most evaluated scale is the National Institutes of Health Stroke Scale (NIHSS)[4] which can be completed in 8 minutes[29].

The most widespread diagnostic tool used in the assessment of acute stroke is neuroimaging. Brain imaging is essential to differentiate haemorrhagic from ischaemic strokes and also to exclude stroke mimics such as tumours. Imaging can help to select a patient most likely to benefit from treatment and exclude those with greater risk of complications. The most common and accessible imaging technique used is CT scanning. While a CT scan is sensitive at detecting ICH, its sensitivity to tissue damage in the acute period of ischaemic stroke is reasonably poor[30]. It gives a suboptimal view of the posterior fossa. However CT is widely available, practical, quick and easy to use[31]. Compared to CT, MRI can image the entire brain and detect microbleeds without radiation or iodinated contrast agents[32]. However many patients may be ineligible for MRI due to confusion or they may have a metallic foreign body such as a pacemaker.

The SIGN guidelines for Scotland suggest that all patients with suspected stroke should have brain imaging upon presentation[4]. A CT scan is recommended for most patients in the acute phase of stroke though MRI with diffusion weighted imaging and gradient echo sequences is recommended if available and applicable[4]. For minor strokes and TIA, MR is the imaging modality of choice. Scans are usually interpreted by a trained radiologist.

1.2. Epidemiology of stroke

1.2.1. Stroke surveillance

Stroke is a non-communicable disease of increasing socioeconomic importance in ageing populations[33]. While in recent years imaging techniques have been introduced for diagnosis, stroke can be defined clinically allowing past trends to be observed without

access to specific laboratory or imaging equipment[34]. Surveillance systems are said to be most appropriate when the disease or health concern is both a major public health problem and preventable or modifiable[35]. Stroke meets many of the criteria warranting a surveillance system. Documenting the size of stroke burden in relation to other diseases provides a foundation for improvement of stroke prevention and treatment[36].

Surveillance allows trends to be observed within populations, showing epidemiological changes in stroke impact over time. The WHO global burden of disease project[37] gives us reliable data measured across different WHO member states. Studies such as the WHO MONICA project[1], the Global and Regional Burden of Stroke Study[6] and the Global Stroke Initiative[36] attempt to document the size of stroke burden in relation to other diseases.

1.2.2. Incidence and prevalence of Stroke

It has been reported that approximately 15 million stroke cases occur each year[38]. The most reliable epidemiological data come from population based studies. Incidence measures the number of strokes (new or recurrent) per unit of time (usually year)[39]. Stroke incidence and prevalence are highly associated with age. There is a continuous trend between increasing age and increasing stroke incidence[2 33]. This is illustrated in Figure 1-1 below. There is also an observed trend between socioeconomic status and stroke incidence; with an observed increase of 100% reported in lower-middle income countries over the past four decades[40]. The WHO MONICA project found geographical differences in stroke incidence[41] but a review of several population based studies by Feigin et al only observed small differences in age-standardised stroke incidence[33]. These differences in reported incidence rates could be due to changes in medical care and exposure to risk factors[42].

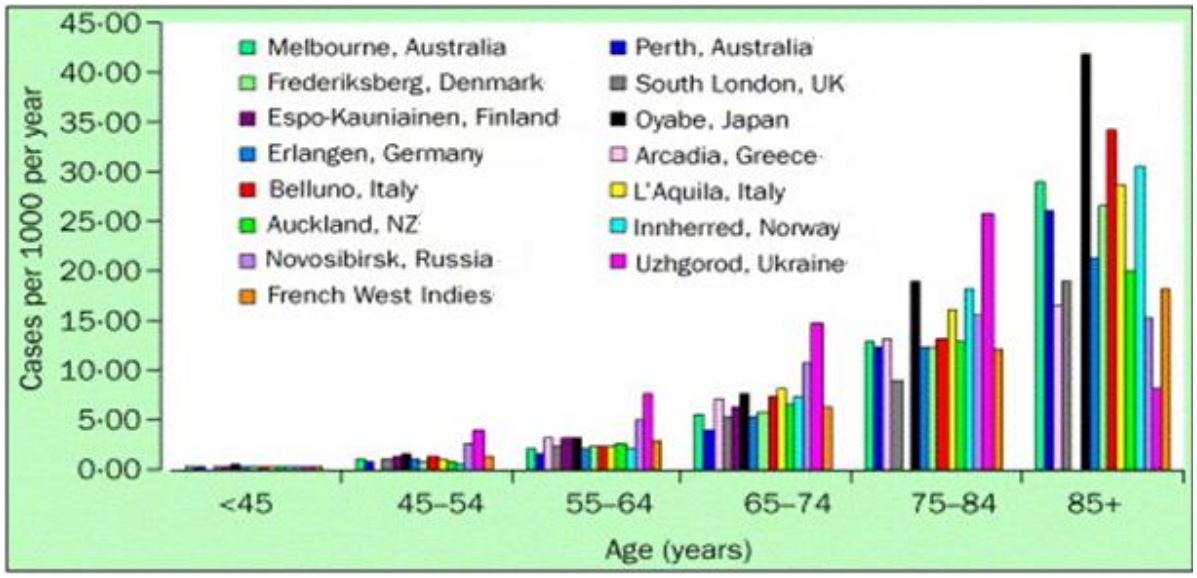

Figure 1-1: Annual incidence by age per 1000 population in all types of stroke. reproduced from Feigin et al (2003)[33].

Feigin et al[42] conducted a review of 56 population based studies and found a pooled estimate of age-adjusted incidence of first ever stroke for high income countries was 94 per 100,000 population per year and for low to middle income countries 117 per 100,000 population per year in the period 2000-2008. These estimates had decreased from 121 and 164 per 100,000 population per year respectively in the period 1990-1999[42]. With the recent improvements in surveillance of stroke these estimates will be more reliable than earlier estimates of incidence[39], the categorisation for high income and low-middle income countries was taken from the world bank definitions 2004[43]. The incidence of stroke in the United Kingdom has been calculated from the General Practice research database (GPRD) to have fallen from 148 per 100,000 person years to 104 per 100,000 person years over the period 1999 to 2008[44]. This decrease in stroke incidence over time is most often attributed to better treatment of risk factors[45].

The prevalence of stroke depends on incidence, mortality and mean duration of survival after stroke[46]. In 2002 the World health Organisation (WHO) estimated that there were

15.3 million strokes worldwide[37]. The centre for disease control (CDC) calculated the prevalence of stroke in the United States in 2006 to be 2.7% and 2010 to be 2.6%[46]. UK-specific prevalence according to the GPRD has increased from 0.64% to 0.72% in the period 1999 to 2008[44].

1.2.3. Stroke related mortality and case fatality

Stroke caused an estimated 5.7 million deaths in 2005[47], in 2011 stroke and other cerebrovascular diseases were the second leading cause of death worldwide causing an estimated 6.15 million deaths[48]. Johnston et al[39] derived global mortality rates using data from the WHO global burden of disease project[49] to range from 24.5 to 251 per 100,000 population (0.024% to 0.251%) depending on the country. These estimates were based on data obtained from 192 WHO member states with reliable data and adjusted for age and sex. As can be seen in Figure 1-2 below Western Europe and North America are among those with the lowest and Eastern Europe and North Asia are amongst those with the highest stroke mortality rates.

Stroke mortality varies widely from region to region as it is determined by many factors such as incidence, stroke sub-type, age and gender of the population studied[2]. With life expectancy increasing and populations in low-middle income countries ageing there has been an epidemiological shift resulting in stroke becoming a major health problem in such countries[47]. Compared to people in high income countries, those from low-middle income countries experience a higher stroke mortality rate. Approximately 85% of all stroke deaths are registered from these regions[47 50-51].

Early case fatality, coded as death within 21 days to 1 month after event, from a population based study was found to be 17-30% for high income countries and 18-35%

for low-middle income countries for the period 2000-2008[42]. UK specific early case-fatality coded as death within 56 days of first stroke was found to be 15% from the GPRD[44]. From Figure 1-3 a decreasing trend in stroke mortality can be observed over a period of 10 years, the mortality rate for females appearing slightly higher than males in both higher (≥ 80) and lower (<80) age groups. The information presented in both Figure 1-2 and Figure 1-3 include early mortality rates for all stroke subtypes and are not specific to ischaemic stroke.

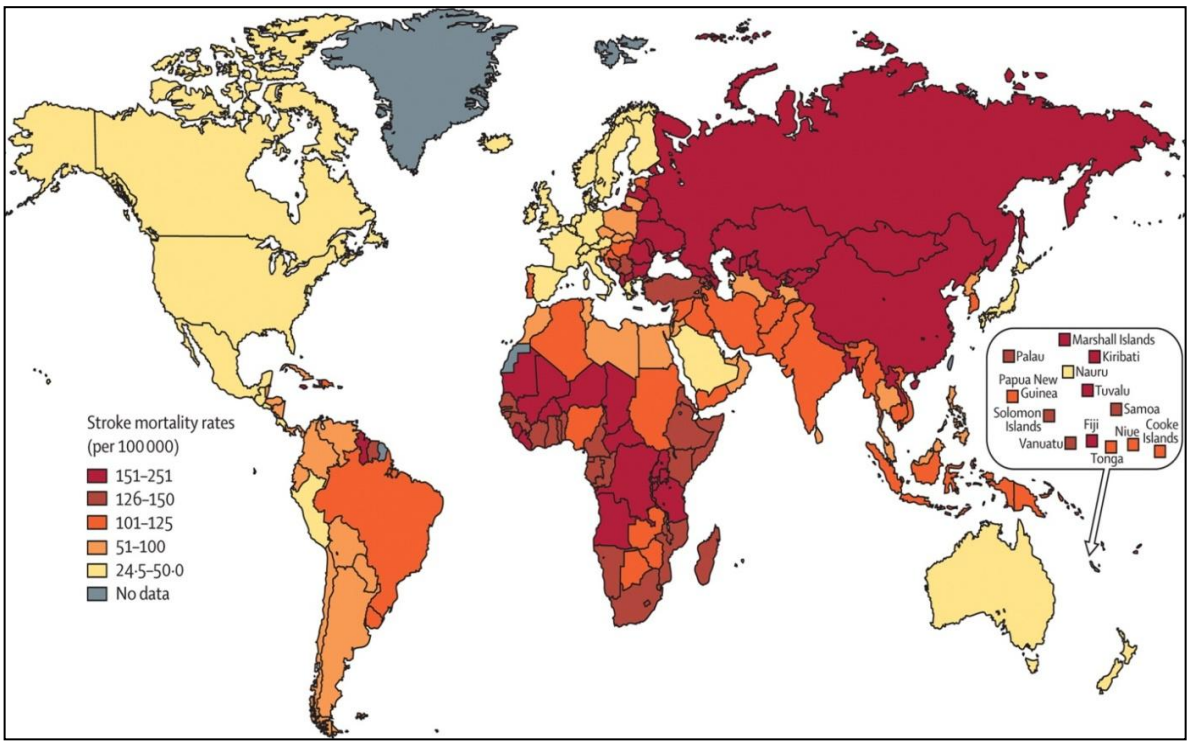

Figure 1-2: Rates of age and sex-adjusted stroke mortality, reproduced from Johnston et al (2009) [39].

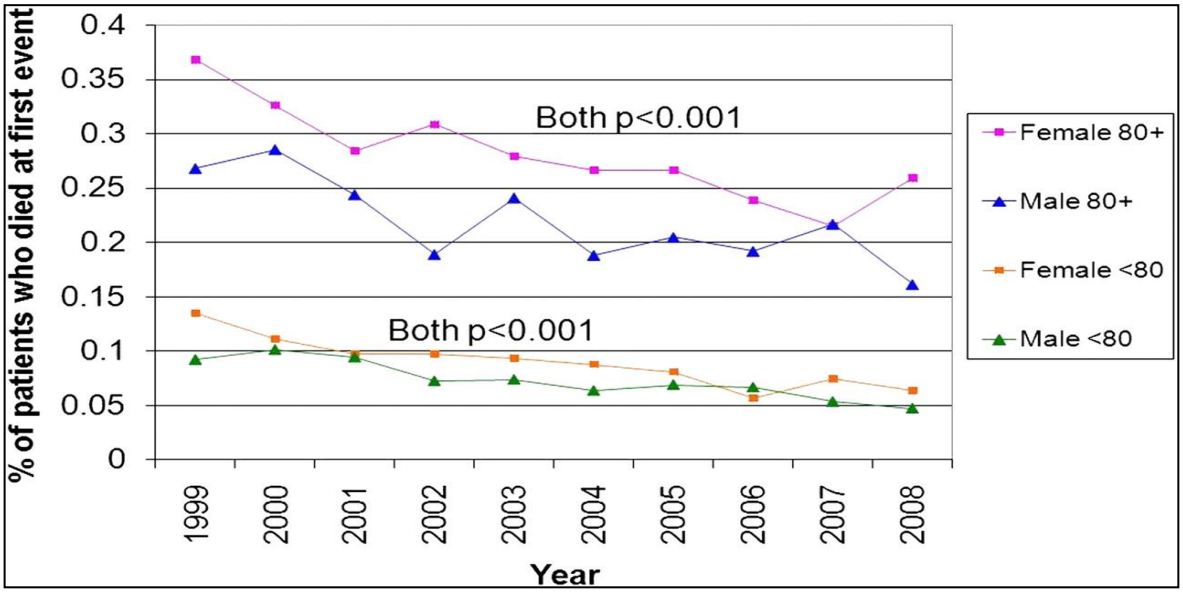

Figure 1-3: Proportion of patients who die within 56 days of first stroke by age group and gender. Reproduced from Lee et al (2011)[44].

1.2.4. The burden of stroke

In addition to mortality, long term morbidity is also a growing issue leaving stroke survivors with moderate or severe disabilities who are then dependant on others to carry on with daily life[52]. Stroke is a leading cause of serious, long term disability[53] causing a greater range of disability than any other acquired condition[54]. Approximately 40% of survivors are left with some degree of functional impairment[55]. Worldwide over 5 million people are left permanently disabled after stroke; however stroke also causes emotional and intellectual problems with approximately 1 in 3 survivors suffering from depression[27].

Due to different surveillance approaches the comparison of disability rates over time and between populations is difficult. Reports of incidence and prevalence come mainly from developed countries with strong epidemiological traditions[34]. One of the first global assessments of burden was the global burden of disease study[56] looking to understand the long term impact of chronic, morbid conditions[35]. The unit of measurement used to

measure burden is the disability-adjusted life year (DALY), in which one DALY represents the loss of one year of healthy life[49].

The World Health Organisation burden of disease project (2004 update) found that worldwide cerebrovascular diseases contributed to 3.1% of total DALYs, placing stroke in 6th place among the leading causes of burden of disease[49]. DALYs due to stroke vary among regions due to geographical differences in incidence, risk factors, access to healthcare and mortality rates[39]. Figure 1-4 shows how DALYs differ for 20 different WHO member states; similar to trends in mortality, high income countries such as UK and USA have a lower rate of DALY from stroke than lower-middle income countries such as Russia. Looking at Figure 1-4 similar geographic trends can be seen to those observed in mortality rates can be seen for DALYs.

The true burden of stroke is often under-represented, possibly due to a lack of understanding of its nature and the misrepresentation that stroke is a disease only of the elderly. Stroke is a major cause of dementia, depression and other secondary medical problems such as falls and fractures[57]. Stroke has a major impact on everyday life from the more obvious physical effects and emotional wellbeing to family life, social skills and vitality[27].

Since stroke causes morbidity more often than mortality, patients require longer hospital stays followed by ongoing support[57]. However as well as human cost, stroke has a major economic burden worldwide. It was estimated in 2006/7 that the cost of stroke to the health and social care system was over £2.5 billion in the UK alone, over 80% of these costs for inpatient hospital care[58]. In the United States the total direct cost of stroke was estimated at $25.2 billion. In the 27 EU countries this figure was estimated to be €18.5

billion. These costs are nearly twice the cost for coronary heart disease [53]. Stroke accounts for between 2 and 4% of total healthcare expenditure in developed countries[59]. The lengthy hospital stay of a small proportion of patients goes some way to explaining this disproportionate use of resources[3]. However the costs of stroke are more diverse than other conditions. These costs reflect the high rates of disability and dependence experienced by stroke survivors. Indirect costs, considering lost productivity due to morbidity and mortality, account for over 30% of the total cost of stroke[51].

Despite the enormous and growing burden of stroke, there is evidence that research into the condition does not receive the funding it deserves[36]. Spending on stroke research is low in comparison to the other two most common causes of death in the developed world, heart disease and cancer[57]. Funding comes primarily from three sources: disease-specific charities, disease-specific government agencies and the pharmaceutical industry[60]. In 2005 it was estimated that in the US alone cancer received approximately 16 times the amount of funding received by stroke. Figure 1-5 shows condition-specific NIH funding over a 3 year period. These patterns are mirrored in the UK and across Europe[60-61].

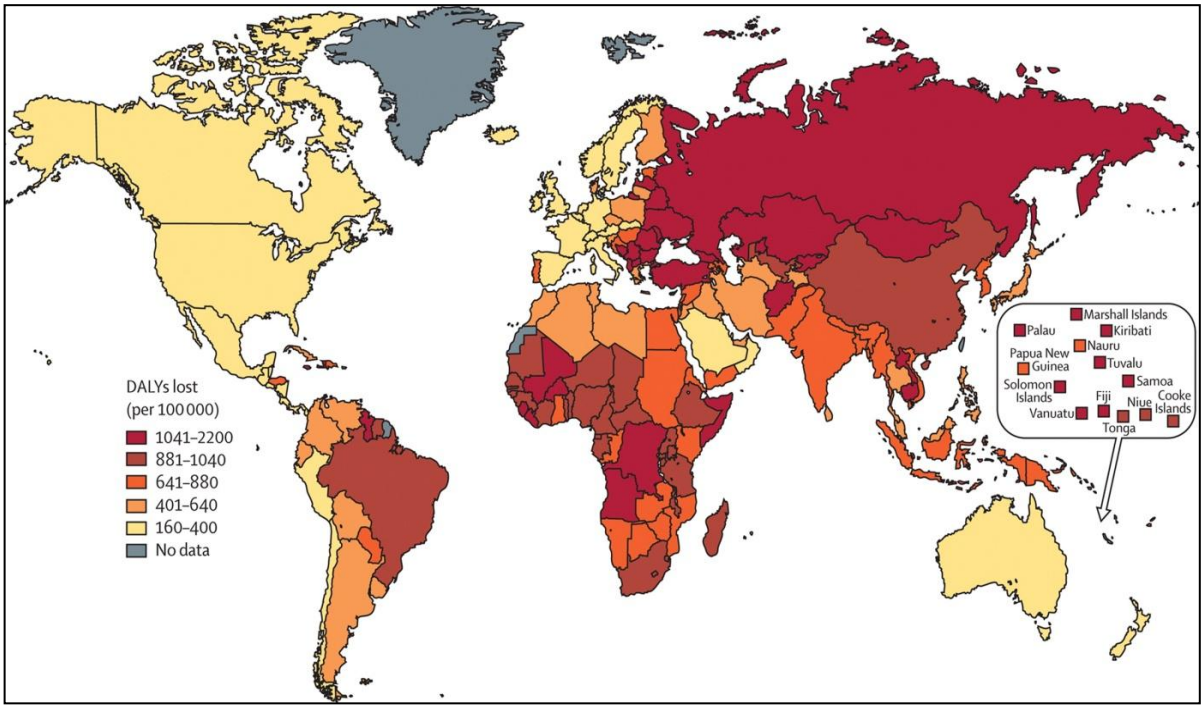

Figure 1-4: Age and sex-adjusted DALY loss attributed to stroke. Reproduced from Johnston et al (2009) [39].

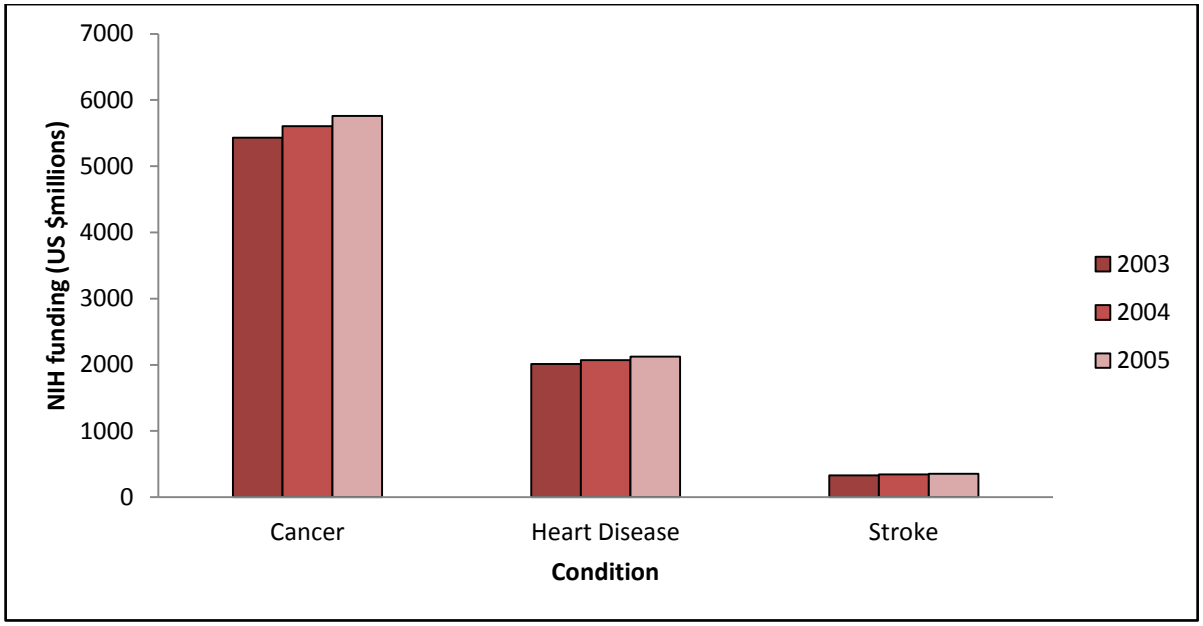

Figure 1-5: Research spending by the National institutes of health by condition. Reproduced from data provided in Broderick (2004)[62].

1.2.5. Trends over time, looking to the future

The global population aged over 65 years of age was estimated to be over 600 million in 2003[38] and is projected to be increasing by 9 million per year[38]. Given that age is one of the most substantiated risk factors for stroke[63] it follows that with the ageing population and improvements in life expectancy, stroke will affect an increasing number of people[27] [45]. However there have been measures to counteract the increasing prevalence with major developments in identification and control of risk factors as well as the management of acute stroke[45].

The Framingham study observed a population free of prevalent strokes over a period of 50 years. During the three periods of this study (1950-1977,1978-1989 and 1990-2004) age adjusted incidence reduced significantly from 0.076 to 0.053 in men and 0.062 to 0.051 in women per 100,000 person years. Thirty day mortality significantly decreased in men over the three periods from 23% to 14% but barely changed in women, with rates of 21% to 20%.

Trends in stroke incidence and mortality vary widely among different populations, making predictions difficult[40]. Data from the WHO show an increased incidence of stroke during the period 1970-2008 in lower-middle income countries increasing from approximately 52 to 117 per 100,000 person years which is a disproportionate increase when compared to high income countries. Figure 1-6 shows a comparison of these rates over four decades for low-middle income countries and high income countries. In many developed populations, death from stroke has fallen dramatically in the past 30 years[27]; however an inverse association has been observed between socioeconomic status and mortality. Figure 1-7 shows the changes over four decades in the mortality rates from stroke separated by income group. Much of the disparity may be due to the lack of reliable data

for developing countries[64] rather than just the limited resources available for healthcare. These data suggest that for the impact of stroke to be reduced globally, prevention strategies targeted to lower socioeconomic groups need to be introduced. In particular more cost-effective measures are needed, promoting access to effective intervention and care[40].

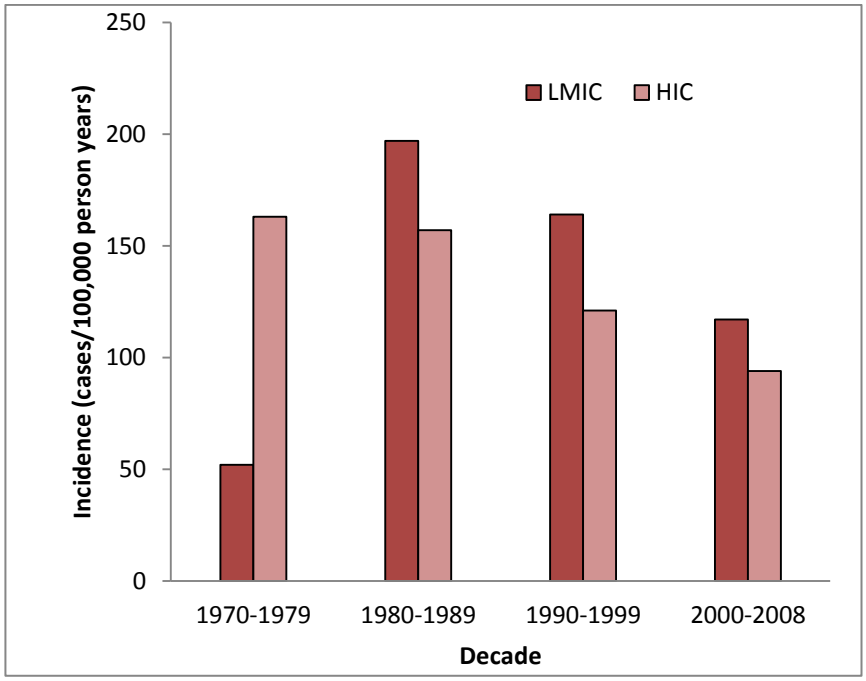

Figure 1-6: Variation in incidence of stroke for lower-middle income countries (LMIC) and high income countries (HIC). Reproduced from data provided in Addo et al (2012)[40].

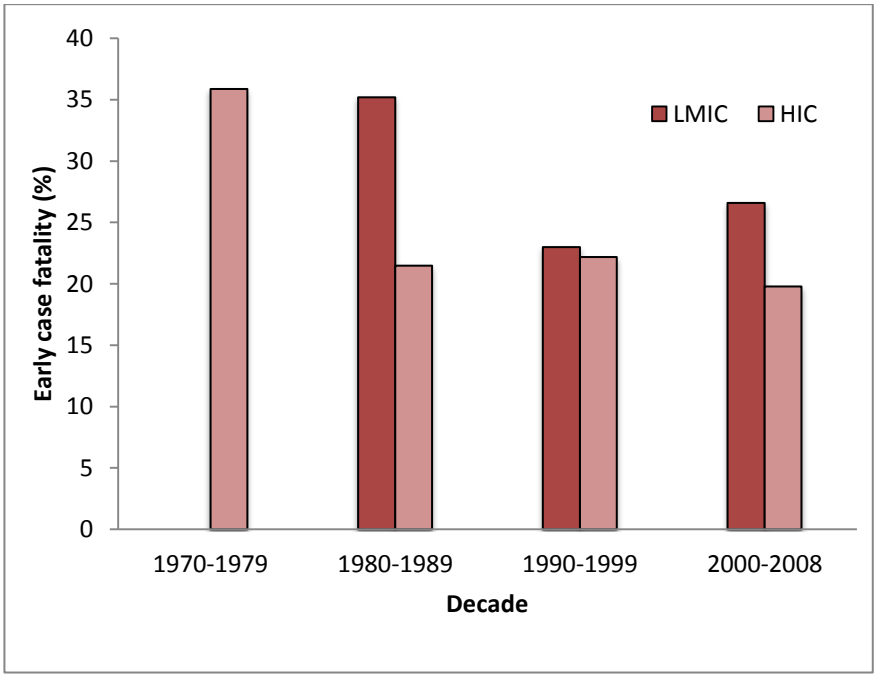

Figure 1-7: Variation in early case fatality of stroke for lower middle income countries and high income countries. Reproduced from data provided in Addo et al (2012)[40].

1.3. Risk factors for stroke

A risk factor is a characteristic associated with an increased incidence of stroke in an individual or population compared to those without the characteristic. The effects of risk factors on stroke incidence may be additive or multiplicative, so the presence of several puts an individual at very high risk.

1.3.1. General risk factors

Stroke is known as a disease of the elderly, with nearly three quarters of all strokes occurring in people over the age of 65[53]. There is a continuous trend observed between stroke incidence and age. Stroke incidence rates rise exponentially with each increasing decade of age[3].

Stroke is approximately 25 times more common in those aged 75-84 than those aged 45-54[2]. Patients who are recovering from a mild stroke or who have experienced a transient

ischaemic attack are at high risk of recurrent stroke. The long-term stroke recurrence rates range from 4%-14% annually, with 2 year cumulative recurrence rate approximately 14.1%.[65]. Recurrent stroke is most common within 30 days of the event, approximately 30% of recurrent strokes occur within this time frame[66].

Several medical conditions are well known to contribute to the risk of stroke; the most consistent and powerful is high blood pressure or hypertension[67]. The association between blood pressure and stroke incidence has been widely studied[68-69]. Diastolic blood pressure (DBP) is independently associated with the incidence of stroke, particularly at higher levels[70-72]. There have been major efforts to reduce hypertension in the general population by improving knowledge of hypertension and the development of new anti-hypertensive medications[73-74]. Population wide reductions in blood pressure and mortality rates of stroke highlights the influence that the methods to reduce blood pressure have had[73]; however it is important that these methods are sustained[75].

Atrial Fibrillation (AF) increases the risk of stroke. The attributable risk of stroke for AF increases from 1.5% to 23.5% with increasing age from 50 to >79[76]. Atrial Fibrillation reduces cardiac output and in turn reduces cerebral blood flow exerting a significant impact on the risk of stroke independently of other cardiac risk factors[76]. There have been strategies in place for the management of AF and the reduction in the rate of stroke in those with AF using anticoagulant therapies such as warfarin[77]. Warfarin reduces the risk of stroke in patients with AF. Though anticoagulation remains the established approach, it is suboptimal in many cases. New antiarrhythmic and antithrombotic agents have now been approved[78] and will have an increasing role in the future[79].

Hypertension often coexists with diabetes[80-81]. Several studies have confirmed diabetes to be an independent risk factor for ischaemic stroke. In non-diabetic patients impaired glucose tolerance was found to be an independent risk factor of stroke, almost doubling the risk and nearly tripling it in those with diabetes[82].

1.3.2. Lifestyle risk factors

There are many risk factors for stroke can be attributed to lifestyle; it has been shown that a healthy lifestyle is more effective in lowering cardiovascular disease than any one single factor[83]. The main factors that are considered to contribute to a high risk lifestyle are smoking, exercise, diet, body mass index (BMI) and alcohol consumption[83]. In a large cohort of women it was found that controlling these risk factors was associated with a substantial decrease in the risk of stroke, even after adjustment for other risk factors such as diabetes, cholesterol and hypertension, with a hazard ratio of 0.45 compared to those with a high-risk lifestyle[84].

INTERSTROKE is a case-control study designed to look at the association of traditional and emerging risk factors with stroke[85]. Hypertension, smoking, obesity, diet and physical inactivity accounted for 80% of the global risk of all stroke[85].

Guidelines for the primary prevention of ischaemic stroke published by the American Heart Association and American Stroke Association summarised existing evidence on risk factors for stroke. From epidemiological and retrospective cohort studies it was found that diets rich in fruit and vegetables with reduced sodium and potassium, regular exercise, reduction in weight, smoking cessation and low to moderate alcohol consumption all reduce the risk of stroke[86]. Many of these lifestyle factors are associated

with a reduced risk of other factors such as high blood pressure and diabetes, further reducing the risk of stroke.

1.3.3. Race as a risk factor

There is some evidence to suggest that there is a racial disparity in stroke incidence for those living in western countries[2]. The Reasons for GEographic And Racial Differences in Stroke (REGARDS) study sought to investigate the disparities in stroke incidence and mortality between racial groups[87]. The incident stroke risk was 2.9 times greater in blacks than whites, remaining around two-fold higher after adjusting for other risk and socioeconomic factors[88].

The age adjusted stroke incidence rates for those aged 45 to 84 were 660 per 100,000 population in black men and 360 in white men, 490 in black women and 230 in white women. These rates have not changed significantly over time; however the case fatality rate is similar in both racial groups. The age adjusted incidence of first ischaemic stroke per 100,000 was 88 in whites, 191 in blacks and 149 in Hispanics, taken from the Northern Manhattan Stroke Study[53].

It is thought that risk factor prevalence and control differs between racial groups, contributing heavily to the inequality in stroke incidence[89]. From the REGARDS study, race differences were found to be large for blood pressure, diabetes and use of antihypertensive therapy. Hypertension is more prevalent and more difficult to control in blacks and often presents at an earlier age[90]. Diabetes is also more prevalent and rates of diabetes are increasing faster in blacks than whites[89]. Data from the Northern Manhattan Stroke Study found that the prevalence of stroke risk factors varied widely by race

particularly when considering diabetes and hypertension the prevalence and risk was higher in both blacks and Hispanics than in whites[91].

1.4. Acute treatment

For the purposes of this thesis the primary focus will be acute ischaemic stroke.

1.4.1. Thrombolysis

1.4.1.1. What is thrombolytic therapy

One of the biggest advances in the last few decades has been the use of thrombolytic therapy in the early stages of ischaemic stroke. The interest for its use arose from the use of thrombolytic agents in myocardial infarction and an increased understanding of stroke mechanisms [10 92]. Thrombolytic agents can be used to help dissolve any recent thrombus quickly, restoring blood flow to the affected area[92]. Recombinant tissue plasminogen activator (rt-PA) is the only thrombolytic treatment licensed for stroke use in Europe. It has been shown to increase the odds of survival without dependency[93].

1.4.1.2. Is it safe and effective?

The use of intravenous thrombolytic therapy for acute stroke was approved by the U.S Food and drug administration (FDA) in 1996[94]. Wardlaw et al conducted a Cochrane review of thrombolysis for acute ischaemic stroke[95], including 26 trials. Thrombolysis gave a net reduction in the proportion of patients dead or dependant. The benefits of intravenous thrombolysis are greatest when given between 0 and 3 hours of stroke onset as shown in the National Institute of Neurological Disorders and Stroke (NINDS) rt-PA trial[96]. The odds of achieving more favourable outcome for those treated with thrombolysis were 1.7, with treated patients 30% more likely to have minimal or no

disability[96]. It has since been shown in the third European Cooperative acute stroke study (ECASS III) that treatment between 3 and 4.5 hours is also safe and effective, with odds of favourable outcome of 1.32.

As the goal of thrombolytic therapy is to dissolve a thrombus, the most common complication that occurs is bleeding[97]. The principal adverse event from treatment with thrombolysis for stroke is symptomatic Intracerebral haemorrhage (SICH) [98]. From the NINDS trial the rate of SICH is approximately 6%[96] in stroke patients treated with thrombolysis. However different definitions are used to define SICH so this estimate varies[99]. Other side effects include extracranial haemorrhage, infection, recurrent stroke and anaphylaxis. Specific trials will be discussed in section 1.6.2.

1.4.1.3. Guidelines for use

The European Stroke Organisation (ESO) published guidelines for treatment of ischaemic stroke in 2008[100] and updated in 2009 to include treatment with thrombolysis between 3 and 4.5h[101]. These guidelines suggest patients eligible for treatment should follow certain characteristics. These include: Diagnosis of ischaemic stroke presenting less than 4.5 h after onset, showing no minor or rapidly improving symptoms and no symptoms indicative of SAH. No recent head trauma, major surgery, arterial puncture, previous stroke or myocardial infarction (MI). Patients outwith the ages of 18 and 80 may be treated but it is outside current European guidelines. Patient must have normal blood pressure, have taken no oral anticoagulant and have no evidence of active bleeding or acute trauma. Platelet count should be greater than 100,000mm^3, blood glucose greater than 2.7mmol/L and if heparin has been recently administered aPTT must be normal. Finally the CT scan does not show a multilobar infarction and the patient or family

members understand both the benefits and the risks of thrombolytic therapy. Guidelines for the United States, published by the American Heart Association, follow similar restrictions but licensing only allows for treatment below 3h[94 102].

1.4.2. Aspirin

1.4.2.1. Is it safe and effective?

The small benefit associated with prolonged use of aspirin in the prevention of cardiovascular events has been well documented[103]. The use of aspirin for the treatment of acute stroke was assessed in two large multicentre trials, the International Stroke Trial (IST)[104] and the Chinese Acute Stroke Trial (CAST). The use of aspirin in conjunction with thrombolysis might increase the risk of bleeding and so it is only recommended to be given to patients in the acute phase of stroke who are ineligible for thrombolysis[105].

1.4.2.2. Guidelines for use

Aspirin is the only oral antiplatelet agent that has been evaluated for the treatment of acute ischaemic stroke[94]. While the benefits of aspirin after acute stroke are small, the intervention is available to the majority of patients[10] and is recommended as part of management after acute stroke[94].

1.4.3. Mechanical embolus removal and other endovascular approaches

1.4.3.1. Is it safe and effective?

The Mechanical Embolectomy Removal in Cerebral Ischaemia (MERCI) trial[106] showed that clot extraction after large vessel intracranial occlusion was successful in 53.7% of those treated with the device alone. In those treated with a combination of the device and intra arterial (IA) thrombolysis recanalisation was achieved in 69.5% of patients

treated. However there is a risk of SICH, in approximately 9.8% of those given the device SICH was observed. Clinically significant procedural complications occurred in 5.5% of patients and the device related serious adverse event rate was 2.4%[106]. Other mechanical embolectomy trials include the Penumbra System[107], the Interventional Management of Stroke (IMS) trial[108] and the MR RESCUE trial[109]. Both the IMS and MR RESCUE trials found that embolectomy was not superior to standard treatment with intravenous thrombolysis[108 110]. Other endovascular approaches have been investigated such as intra-arterial thrombolysis with pro-urokinase in the PROACT I and II trials[111-112] and urokinase in the MELT study[113]. Some smaller studies have also been conducted[114-115]. The benefits and drawbacks of intra-arterial thrombolysis are discussed further in several systematic reviews[95 116-121].

1.4.3.2. Guidelines for use

Both the MERCI retriever device and the Penumbra system are approved for use in the US and Europe[122]. These are two of the few devices approved for clot extraction from an occluded artery[94]. Other endovascular approaches such as intra-arterial thrombolysis have also been approved for use in current practice[94]. However, a recent study by the SYNTHESIS expansion investigators found that endovascular therapy is not superior to intravenous thrombolysis[123]. Mechanical thrombectomy provides a therapeutic option for those who are ineligible for or unresponsive to IV thrombolysis. In carefully selected patients the device is recognised as a reasonable intervention. However since all studies for the devices have been single-arm, the effectiveness of the device for improving outcomes after stroke is unclear. Further studies including randomised clinical trials are needed to fully validate these findings[94 122 124].

1.5. Outcome measurement

1.5.1. Functional outcome measures

There are several dimensions to post stroke recovery: neurological deficit, functional disability, quality of life and physical impairment. There is no single outcome measure that describes or predicts all of these dimensions. This has led to scales being used inconsistently among trials[125]. Assessing the influence of a treatment for acute stroke requires valid, reliable and responsive outcome measures[126]. Accurate assessment of patient outcomes is particularly important for clinical trials as the effect of an intervention is being sought whether positive or negative. There are a wide range of functional outcome measures available. A recent review by Quinn et al found that 47 different outcome measures were described in 126 trials published in the years 2001-2006[127]. This heterogeneity in the use of scales presents a problem in the comparison of studies.

1.5.1.1. Most prominent outcome measures used

Quinn et al found that the three most prominent outcome measures used were the modified Rankin Scale (mRS), the Barthel Index (BI) and the National Institutes of Health Stroke Scale (NIHSS). The mRS was primary endpoint in 26.2% of trials, the BI in 7.9% and the NIHSS in 11.9%. The majority of trials recorded at least two outcomes and so each of these scales was used as secondary outcome in several studies[128]. Each of these three scales measure a different domain of disability.

The mRS, shown to be the most prevalent of all outcome scales, is a tool to assess functional independence after stroke. The scale is graded based on a short (approximately 5 minute) interview of the patient or caregivers. The scale is scored on an

ordinal basis into one of seven categories with 0 indicating no symptoms at all and 6 records death. The nature of the mRS leads to difficulty in distinction between grades so there is often a high rate of interobserver variability[129-130]. This variability is one of the main weaknesses of the mRS as it can lead to misclassification of the outcome measure[130]. Formal training and certification is available and recommended for use of the mRS, particularly for use in a clinical trial setting[131]. The lack of specificity in the mRS is another pitfall; it does not take into account factors such as cognition, language and emotional disability[125].

The NIHSS is a 15-item scale measuring neurological impairment. The individual items of the scale measure deficits affecting level of consciousness, motor function, ataxia, language, visual fields, dysarthria, sensory, extraocular movements, facial palsy and neglect[125 132]. The score ranges from 0 (no deficit) to 42 (maximal deficit). The scale is assessed based on a series of tasks graded by a trained observer. The NIHSS is widely used and has high interobserver reliability when performed by trained investigators[133]. Formal training and certification are available online and the latter is required for participation in clinical trials[134]. One of the main pitfalls of the NIHSS is that it is not indicative of a patient's functional ability. The scale favours left hemisphere strokes i.e. such patients receive a higher score for a given volume of infarction[135].

The BI measures activities of daily living and mobility by assessing the patient's ability to perform 10 basic tasks. These tasks are feeding, chair/bed transfer, grooming, toilet, bathing, ambulation, stair climbing, dressing, bowel control and bladder control. These each attract a score of 0, 5 or 10 points that are summed to give a score ranging from 0 to 100. Lower scores indicate greater dependency[136]. The BI is widely used in clinical trials and there is training and certification available online[125 127]. The intra and interobserver

reliability have been reported to be very high, along with internal consistency of the scale[125 136]. While the BI looks at activities of daily living, many aspects of functional independence are not included which could have a significant bearing on independence such as language, emotional function and visual impairment. The scale also has a 'ceiling effect' where patients who may have several aspects of functional disability may perform well in all of the BI categories, with the patient receiving a high score but in reality being unable to function independently[136].

1.5.1.2. Time point of outcome measurement

In most acute stroke trials the primary end point is measured at 90 days post stroke. Within the Virtual International Stroke Trials Archive (VISTA) 24 acute stroke trials had an outcome measurement at approximately 90 days (26,898 patients)[132].

1.5.1.3. Other methods of measurement

There are several other scales that can be used to measure outcomes after stroke. Scales such as the Stroke Impact Scale (SIS) have been developed from the perspectives of the patient or caregiver. However as this scale is self-reporting it has limited use in aphasic patients. Use of a proxy is not always reliable. This has limited use of the SIS in a clinical trial setting with no trials to date using it as a primary outcome measure[125 127]. There are other neurological impairment scales used in clinical trials: the Scandinavian Stroke Scale (SSS), the Canadian Neurological Scale (CNS) and the Orgogozo neurological scale. Each of these scales has appeared as a primary outcome measure in fewer than 2% of stroke trials[127].

The Glasgow Outcome Scale (GOS) and the SF-36 were developed and validated for use in other conditions and have been adopted for use in stroke. The GOS is routinely used for

the assessment of brain injury and the SF-36 was developed as a general measure of quality of life. The GOS was used in approximately 6.3% of acute stroke trials and was a primary outcome measure in 1.6%[127], whereas the SF-36 was not reported in any acute trial though it has been used in trials looking at post-stroke care[137].

As no single scale adequately encompasses post-stroke outcome, the use of a single scale is not desirable. One method to generate a meaningful assessment of post stroke outcome is the combination of multiple scales, deriving what is known as a global test statistic. A global test statistic allows the assessment of treatment effect across multiple scales simultaneously, improving power and providing evidence of treatment efficacy across multiple domains[125]. The NINDS rt-PA trial study group devised a global test encompassing the BI, the mRS, the GOS and the NIHSS[138]. The International Citicoline Trial in acute Stroke (ICTUS) trial used a combination of the BI, the mRS and the NIHSS. Analysis of endpoints will be discussed further in section 1.7.2.

1.5.1.4. What is recommended?

As an ideal outcome measure is likely never to exist, the choice of outcome measure is entirely dependent upon what is expected from the intervention. The European Medicines Agency (EMA) Points to Consider document (2001) suggests that the primary efficacy endpoint should be defined according to the expected effect from the study drug. The document suggests that if a functional outcome measure such as the mRS is used as the primary efficacy variable, a second primary efficacy variable such as the NIHSS should look at neurological deficit in order to demonstrate efficacy across another domain[139].

The European Stroke Organisation working group on outcomes for acute trials support the mRS at 3 months or later as their preferred outcome, with the interview being

conducted by a certified rater[140]. They propose that the use of a second primary outcome measure showing a similar level of efficacy is unnecessary: if a second measure is used it should simply need to agree with the mRS in the direction of effect.

1.5.2. Prognostic factors affecting outcome

Several factors are predictive of functional outcome after stroke. Initial stroke severity as measured by the NIHSS is the most important of these[141-142]. Another extremely influential factor for stroke outcome is age; it is well documented that advancing age is associated with poorer functional outcome.[143] This may be mostly due to comorbidities. Older people are more likely to have other risk factors associated with stroke[63 144]. The Optimising the Analysis of Stroke Trials (OAST) collaboration found a significant relationship between baseline severity and functional outcome in 30 trials studied, and a relationship between age and functional outcome in 29 out of 30 trials studied.

Several other factors influence functional outcome. Many of these factors are pre-existing risk factors for stroke and some are pre-existing medical conditions. Diabetes and ischaemic stroke often arise together. Admission glucose level is significantly associated with poor functional outcome after ischaemic stroke[145-150]. Hypertension is a risk factor for stroke and is more prevalent in the elderly[151] so both increasing age and high blood pressure at admission have an additive effect on outcome[152]. The relationship between functional outcome and baseline blood pressure has been widely discussed[153-155]. The TAIST and IST trials found that both high and low blood pressure were independent predictors of poor functional outcomes after stroke[156-157]. Atrial Fibrillation (AF) is more prominent in older patients and is an independent predictor of severe stroke and poor outcome, particularly in the elderly[76 158-159]. These trends have been observed in both the Copenhagen stroke study and the NINDS rt-PA trial[160-161].

1.6. Clinical trials in acute ischaemic stroke

With the exception of the NINDS rt-PA stroke trial leading to the licensing of rt-PA and ECASS III leading to extension of the license, there have been limited advances in the development of new therapies for acute ischaemic stroke. The development of new therapies is difficult, costly and time consuming. Statistical analysis of trials will be discussed in section 1.7.

1.6.1. Trial Design

Though the ultimate aim is to demonstrate efficacy and confirm acceptable tolerability and safety, this is usually achieved in stages. Trials leading up to acceptance of a treatment are arbitrarily defined as phase 1, 2 or 3. Post marketing studies are labelled as phase 4. A Phase I trial, usually conducted in a small group of perhaps healthy people, evaluates safety of the therapy, observing safe dose ranges and identifying any possible side effects. Phase II studies explore proof of concept, extending the size of the group and giving a preliminary indication therapeutic effect, although these trials are usually not sufficiently powered to determine definite efficacy. In Phase III trials, the drug is given to a large group of patients to confirm efficacy, confirm side effects and perhaps compare the treatment to other commonly used approaches. Often, regulatory authorities demand two phase 3 trials, each positive at a statistical threshold of $p<0.05$, to give a joint type I error rate of 0.0025.

Well designed randomised control trials (RCTs) are the best way to evaluate a new treatment; treatment allocation should be randomly allocated and blinded to both patient and clinician. There are many considerations that need to be made in the design

of clinical trials to ensure the results are valid and robust. Considerations that need to be made for a phase III trial include[162]:

- Dose selection based on the phase I/II studies
- Time window for initiation of drug
- Patient selection
- Outcome measure: Single primary endpoint or use of a global assessment
- Severity of stroke population to be studied
- Length of follow up
- Use of surrogate endpoints to support efficacy
- Pre-specification of analysis plan

If the intervention is invasive and it is considered unethical to recruit placebo patients to the study, a single-arm is sometimes conducted. Efficacy is analysed by comparing each patient's outcome to their condition at baseline, before the intervention. Outcome measures used differ from standard efficacy measures used in RCTs. An example of this is the MERCI trial[106] where rate of recanalisation is the primary efficacy measure.

Cluster randomised trials are trials in which groups of individuals are randomised at group level to receive different interventions. Cluster trials are becoming increasingly popular in epidemiological research. The units of randomisation are often trial-specific such as study centre, city, country etc. This method of trial design reduces statistical efficiency but there are also several attractive features of this design such as reduced risk of experimental contamination and increased administrative efficiency.

Observational studies are another alternative to RCTs. The objective in observational studies is to include all incident cases. These studies may observe the effect of an

intervention as standard of care and compare it to effects in patients who did not receive the intervention but were otherwise treated similarly. One benefit of observational studies is that potential participants will not have the concerns associated with treatment allocation. The aim is to observe the effect of the intervention given as standard of care, without blinded or randomised recruitment. Observational studies may overestimate treatment effect but this is a topic of much debate[163]. In stroke research observational studies have more often been used for investigations into rehabilitation[164].

1.6.2. Thrombolysis trials

In the late 20th and early 21st century there has been a large increase in the number of clinical trials conducted in acute stroke, many of these assessing thrombolytic agents[165]. According to a review by Wardlaw et al (2009) there have been 26 trials investigating thrombolytic agents in patients with acute ischaemic stroke, including a total of 7152 patients[95]. Thrombolytic agents investigated in these trials were rt-PA, streptokinase, urokinase, recombinant pro-urokinase and desmoteplase. The Meta analysis performed in this review found that treatment with any thrombolytic therapy reduced the proportion of death or dependency odds ratio (OR) 0.81 95% Confidence Interval (CI) (0.73-0.90) but increased the risk of SICH OR 3.49 95% CI (2.81-4.33).

The only licensed treatment for i.v. thrombolysis is rt-PA treated within 4.5h post stroke. Two placebo controlled trials have shown efficacy; NINDS part II giving treatment within 3h[96] and European Cooperative Acute Stroke Study (ECASS) III, giving treatment between 3 and 4.5h post stroke[98]. Six individual trials showed no efficacy of rt-PA based on the primary outcome of interest; NINDS part I[96], ECASS I and II[166-167], ATLANTIS part A[168] and part B[169] and the Echoplanar Imaging Thrombolytic Evaluation Trial (EPITHET)[170].

Conversely the NINDS part I study showed efficacy when looking at day 90 mRS compared to the primary endpoint; 4 point improvement in NIHSS at 24h[96]. Lees et al[171] performed a pooled analysis including all of the rt-PA trial data from the ECASS, ATLANTIS, EPITHET and NINDS trials, concluding that patients selected using CT could benefit from rt-PA up to 4.5h. They found insufficient evidence of efficacy for treatment given post 4.5h. Table 1-1 shows the odds of favourable outcome based on dichotomising mRS 0-2 vs. 3-6 from each of the significant studies mentioned above.

The third International Stroke Trial (IST-3) study group recently completed a large, international, multicentre trial investigating the use of IV rt-PA in a wider range of patients than clinical guidelines allow[172]. The current European Union (EU) guidelines only allow treatment for patients up to 80 years of age, with no real evidence from current thrombolysis trials to substantiate this as only a small number of patients >80 were recruited to the original studies[95 173-174]. Many clinicians disregarded this restriction, instead following ESO clinical guidelines[175-176]. US clinicians faced no such restriction[94]. The IST results were non-significant when analysed based on the primary efficacy endpoint of OHS 0-2 at 6 months. A significant effect of rt-PA was found when looking at the secondary endpoint (OHS 0-1 at 6 months). Due to the difference in protocol from other thrombolysis trials, comparison to other studies is difficult and inadvisable[177]. Regarding age, the findings of this trial suggest that treatment with rt-PA in the >80 age group was at least as effective as in the younger age group, substantiating previous findings from a non-randomised analysis using VISTA and comparing SITS registry data to VISTA[63 178].

Table 1-1: Results from each of the iv thrombolysis trials showing significant functional outcome (mRS 0-1) after treatment with rt-PA along with the pooled analysis by Lees et al[171]. * denotes adjusted analysis

Trial (Treatment window)	OR (95% CI) for favourable outcome mRS at 3 months
NINDS II (0-3h)	1.7 (1.1-2.6) p=0.019
ECASS III* (3-4.5h)	1.42 (1.02-1.98) p=0.04
NINDS I (0-3h)	2.3 (1.4-3.6) <0.001
Pooled analysis* (0-1.5h)	2.55(1.44-4.52) p=0.0013
Pooled analysis* (1.5-3h)	1.64(1.12-2.40) p=0.0116
Pooled analysis* (3-4.5h)	1.34(1.06-1.68) p=0.0135

The Safe Implementation of Thrombolysis in Stroke-Monitoring Study (SITS-MOST) was a Europe wide multi-centre observational study investigating whether the use of IV thrombolysis within 3h of ischaemic stroke was as safe in clinical practice, outside of a randomised trial[179]. The SITS-MOST study looked only at thrombolysed patients and compared them to the corresponding treated patients from the pooled thrombolysis trials treated within 3h of onset[180]. Complete recovery as measured by mRS (0-2) at 3 months was observed in 38.9% of patients in SITS-MOST compared to 42.3% in the pooled RCTs[179-180] . Retrospectively, an adjusted analysis was performed on the SITS-MOST data and found that favourable outcome at 3 months was 50.4% in SITS-MOST compared to 50.1% in RCTs, supporting the conclusions from RCTs and pooled analyses that IV rt-PA within 3h is safe in routine clinical use[181].

1.6.3. Neuroprotectant trials

In the 20th century there were at least 178 clinical trials in acute stroke, 88 of these involving neuroprotectant agents alone[165]. The approval of thrombolytic therapy for

acute stroke prompted further interest in the development of neuroprotective therapies in stroke. Despite the promising results from thrombolysis trials, this trend has not followed in the investigation of neuroprotection strategies. Trials of neuroprotectant agents conducted in the late 20th century such as Glycine antagonist in neuroprotection (GAIN) Americas[182] and International[183], Lubeluzole[184], Selfotel[185], Aptiganel[186], Tirilazad[187] Intravenous Magnesium Efficacy in Stroke (IMAGES)[188], and Enlimomab[189] all showed no efficacy and the reasons for this have been widely considered[190-192].

Studies conducted more recently such as Repinotan modified randomised exposure controlled trial (mRECT)[193]and the NXY-059 (SAINT I and II)[194-196] studies have also failed to show efficacy. Citicoline had shown efficacy in a pooled analysis of individual patient data from clinical trials[197] OR = 1.33 (95% CI, 1.10 to 1.62; P=0.0034) though only one trial within this meta-analysis showed efficacy on its primary outcome measures[198-199]. A randomised, multi-centre, double blinded, sequential placebo controlled study in citicoline in the treatment of acute ischaemic stroke was recently completed, known as the ICTUS trial. The aim of this trial was to confirm efficacy of citicoline in a large randomised trial[200]. The trial was discontinued after enrolment of 2298 patients showing no efficacy on the primary global endpoint.

Table 1-2 below lists all the trials mentioned above along with each of their primary end points, significance levels for treatment effect and whether the trial reached completion or was discontinued.

Table 1-2 Primary endpoint, significance levels and results of the failed neuroprotectant agents mentioned above

Trial	Primary efficacy endpoint	P-Value	Completed or discontinued
GAIN Americas[182]	3 month BI, trichotomised 0-55, 60-90, 95-100	0.79	Complete, no benefit
GAIN International[183]	3 month BI, trichotomised 0-55, 60-90, 95-100	0.8	Complete, no benefit
Lubelozole[184]	3 month BI, trichotomised 0-70, 70-100, dead/vegetative	0.162	Complete, no benefit
selfotel[185]	3 month BI, dichotomised ≥ 60	0.490	Discontinued due to mortality imbalances, no benefit
Aptiganel[186]	3 month mRS	0.44-low dose 0.12-high dose	Discontinued due to mortality imbalances, no benefit
Tirilazad[187]	3 month GOS and BI, dichotomised	>0.05 both outcomes	Discontinuted, no benefit
Enlimomab[189]	3 month mRS	0.04 favouring placebo	Completed, no benefit
IMAGES[188]	Global endpoint common odds of death/ disability	0.59	Completed, no benefit
mRECT[193]	3 month BI, dichotomised ≥ 85	0.149	Complete, no benefit
SAINT I[194]	3 month mRS (0 to 5) ordinal	0.038	Completed, benefit shown on primary outcome. Second study proposed
SAINT II[196]	3 month mRS ordinal	0.33	Completed, no benefit
ICTUS[200]	Global test BI, mRS and NIHSS	0.364	Discontinued, no benefit

1.6.4. Databases and registries

Due to the data monitoring practices implemented in clinical trials, the trials have the opportunity to provide a rich, reliable source of complete patient data to aid in the planning of future trials. The data for many of these trials are kept by pharmaceutical companies or stored in academic archives for years after trial publication; some trials such as NINDS and the original IST trial have made the data available to the public in an anonymised form after a number of years post trial[201-202].

The importance of the data contained within a trial resource is often underestimated. Combining several trials into a single resource or database allows new hypotheses to be tested without the cost associated with research and development of a new intervention[203]. The Virtual International Stroke Trials Archive (VISTA) is a large international database collating data contained in 29 RCTs in acute stroke, containing reliable data on over 28,000 patients. This database alone has resulted in over 40 publications many of which have implications for the management of stroke patients in the acute setting and the design of future trials[203-204].

Stroke databases or registries are established to gather information for scientific research, to acquire epidemiological data to inform healthcare planning and to evaluate and improve the quality of care[205]. Studies based on such registries have led to advances in medicine and enhanced the understanding of the history of disease[206]. Several of these prospective registries have been implemented in different countries worldwide such as; the registry of the Canadian Stroke Network[206], the Performance, Effectiveness and Cost of Treatment episodes in stroke (PERFECT) study in Finland[205], the South London stroke registry in the UK[207], the Harvard cooperative stroke registry[8] and the German Stroke databank[208]. Each of these registries is centred on one geographic region. The

opportunity for further research comes from multinational registries such as the stroke data bank[209].

One of the most widely used international registries is The multinational Safe Implementation of Thrombolysis in Stroke-International Stroke Thrombolysis Register (SITS-ISTR). SITS-ISTR is an international, web-based, interactive registry of stroke patients treated with thrombolysis, including those that did not meet the criteria for SITS-MOST, as discussed in section 1.6.2. SITS-ISTR is an ongoing registry and enables participating centres to compare treatment results and encourages collaborative research[176 179]. Since the inception of the registry a total of 29 publications have been accepted utilising the data, many of which investigate the safety of thrombolytic therapy and make recommendations on its use[210].

1.6.5. Inclusion criteria in stroke trials

It is widely accepted that a rigorously designed randomised control trial is necessary to evaluate the efficacy of new therapies in acute stroke; however study design and recruitment rate can adversely affect the cost and duration of such studies[165 211]. The efficiency with which subjects are recruited into trials is largely influenced by the stringency of the entry criteria. A meta-analysis performed by Elkins et al found that trial entry and organisational criteria considered in their analysis accounted for approximately 40% of the variability in subject recruitment efficiency across the included studies[211].

Several trials in the past have chosen inclusion criteria to facilitate the inclusion of as wide a rage of patients as possible, excluding only those with very good and very weak prognoses. Although it is desirable to show efficacy in all patient subgroups by including as wide a range of patient groups as possible, the nature of stroke itself may have led to

this being a failed strategy. The ECASS III trial showed the benefit of a more stringent approach to patient inclusion in comparison to other extended time window trials in thrombolysis such as ATLANTIS. The NINDS study group reassessed their data by restricting their patient group to follow the more restrictive ECASS III criteria and found that restricting inclusion further did not alter the treatment effect of rt-PA less than 3 hours[212]. It is a matter of debate that the heterogeneity of stroke makes it unlikely that any one intervention will show efficacy in all patients, suggesting the inclusion criteria should suit the intervention[191].

The inclusion criteria applied to a trial are entirely dependent on the intervention being studied. For interventions such as thrombolysis, the time window for patient enrolment remains one of the most important factors for inclusion. Once the mechanism of action of the drug is known, enrolling patients as soon as possible after stroke onset can maximise the chances of detecting treatment benefit[162].

An appropriate patient population also needs to be selected to avoid inadvertently causing harm, for example thrombolysis trials such as ECASS III[98] excluded ages < 18 and >80. Another consideration in the patient population is the baseline severity (NIHSS): mild strokes <6 and severe strokes >25 are often excluded as patients with moderate deficit increase the likelihood of detecting a clinical benefit[162].

A very important consideration that has been brought into trial eligibility is the use of imaging to differentiate subtypes of stroke[191]. The Stroke Therapy Academic Industry Roundtable (STAIR) group support the development and implementation of techniques for patient selection based on imaging-based penumbral identification. The main issue of using imaging for routine use in selecting patients is the limited imaging capability of

some trial centres and the variability in methodology, processing and interpretation of images across centres, potentially limiting applicability of study results in routine clinical practice[213].

1.6.6. Improvements in stroke trial design what have we learned?

Following the disappointing success of neuroprotectant trials, there has been a great deal of discussion and collaboration regarding trial design and implementation of clinical trials in acute stroke. The STAIR committee recommend considering a selective approach to patient inclusion taking into consideration the mechanism of the drug, as well as proper management of medical complications and standardised management of likely side effects of the drug. Choice of an optimal end point measure and the time-point of this measurement are also very important considerations to be made[162 213].

Grotta discusses the need to better translate the conditions under which animal models have been successful into trials conducted in human studies, suggesting that trials which have adhered to this have been the only positive trials to date[214]. Standardisation of factors such as stroke severity, a shortened time to treatment, the combination of neuroprotection with reperfusion strategies, sufficient dose and potency should all be considered when selecting and testing a potential neuroprotective agent. Lees[215] suggests further considerations to be made are the selection of neuroprotective agent, improving patient selection and more optimal selection of endpoints. Many of these considerations are also discussed by Dorman et al[216].

1.7. Statistical analysis in stroke trials

Specific analytic techniques incorporated within this thesis will be discussed in more detail in later chapters. A brief discussion of statistical analysis in stroke trials is given below.

1.7.1. Baseline factors

For clinical trials in acute stroke substantial baseline data are collected. Many of these baseline factors have an influence on outcome as discussed in section 1.5.2. In a double blinded RCT key baseline factors such as age, baseline severity, diabetes etc should be balanced between groups. In contrast imbalances in baseline factors between groups would be expected when conducting an observational study.

Covariates play an important role in the analysis of clinical trials. For analysis of outcome by treatment group some influential baseline factors may be treated as covariates for adjustment or subgroup analysis performed to assess if treatment effect depends upon certain patient characteristics. The choice of covariates is an important consideration in trial design. The two main methods proposed are: adjusting for covariates that are imbalanced across treatment groups, adjusting for covariates correlated with outcome and covariates that follow both rules. As the decision on which covariates to include is required to be made at the design stage of a trial, reliable covariates need to be identified from relationships observed in previous studies. This makes accounting for unexpected imbalances in a clinical trial setting difficult as there is a pre-specified analysis plan already in place which cannot be changed.

Adjusting solely for imbalanced covariates can give an over-inflated estimate of the p-value so truly significant differences may be missed. This method also fails to adjust for

covariates associated with outcome such as age or baseline NIHSS if they are balanced between treatment groups[217]. As mentioned previously, this method is not desirable in clinical trials as it either requires a post hoc decision or the decision needs to be based on previous studies which may have a different distribution of patient demographics[218].

Senn[219] showed that non-significant covariate imbalance can still matter if the covariate is strongly related to outcome. Adjustment for these factors achieves a more precise estimate of treatment effect and a more valid measurement of significance[220].

Covariate adjusted analysis is more complicated than a simple unadjusted analysis and can be difficult for readers to understand, however adjustment using factors associated with outcome achieves more reliable results[218]. As it is now commonplace in stroke trials to perform adjusted analysis, steps should be taken to ensure correct variable selection is made. These selections should be based on both clinical knowledge and the wealth of data available in databases such as VISTA.

1.7.2. Outcome measures

Measures of functional outcome such as the mRS and the BI are non-parametric in nature and can pose analytical difficulties. Typically, most published stroke trials have dichotomised these scales into good/bad outcome and discarded the remaining information. There is some inconsistency in the dichotomisation of these scales: for example in the ECASS II and III studies mRS 0-1 was the primary endpoint[98 167] and in the ATLANTIS study mRS 0-2 was used as a secondary endpoint[169]. Table 1-2 above shows the variability in the dichotomisations and trichotomisations of the Barthel index in studies of different neuroprotectants. This inconsistency makes comparison between studies difficult without access to the raw data.

Over the past decade, analysis of completed trials and theoretical work has generated discussion into the analysis of end points in acute stroke trials[221]. While dichotomised scales make analysis of the endpoint easier to perform and interpret, it results in the loss of information. For example consider a treated patient achieving mRS day 90 of 3 and compare them to an equivalent placebo patient achieving mRS day 90 of 5. While a clinically important difference, it would not be detected in a dichotomised analysis id the mRS was dichotomised at 0-2. An alternative to dichotomised analysis is an ordinal approach such as proportional odds logistic regression and the Cochran Mantel Haenszel. These approaches have gained favour in recent years after the use of ordinal mRS in the analysis of the SAINT I and II trials[195]. The mRS grades 5 and 6 were combined in SAINT I but the full ordinal scale was used in SAINT II. The advantage of these tests is that they take into account the ordinal nature of the given scale rather than just cut at a specific point so results reflect all health state transitions in the analysis.

When considering scales such as the NIHSS which are semi-continuous the nonparametric nature of the data becomes problematic. In some cases the NIHSS is dichotomised into specific categories. For example, in the NINDS study, favourable outcome was defined as NIHSS ≤ 1. In the ECASS III study favourable outcome on the NIHSS was defined as ≤ 1 or >8 point improvement from baseline. This results in a loss of information, an alternative option used in several retrospective analyses[63 222] is to convert the NIHSS into an ordinal measure using pre-defined cut points. This results in the loss of less information than dichotomisation.

From previous trials there have been many lessons learned regarding the analysis of outcome measures in acute stroke, generating a great deal of discussion from academic collaborations. The Optimising the Analysis of Stroke Trials (OAST) collaboration have

suggested ongoing and future trials utilise the full ordinal nature of the data when considering the analysis of the primary outcome[223]. The ESO working group recommend the mRS as the primary outcome measure and discuss different analytic techniques[140]. The STAIR committee discuss the advantages and disadvantages of analysis techniques for all outcome measures[224].

1.7.3. Power and sample size

Sample size calculation is an essential component in the design of a clinical trial. The object of this calculation is to find a sample size adequate for the study to detect a real treatment effect as statistically significant[225]. As per study protocol; the required study sample size must be determined before recruitment starts[226]. Studies with too many or too few patients are economically and ethically unjustified. Key components in sample size calculation include the intended power, significance level, the expected event rate in the control group and the expected treatment effect[227].

In acute stroke the consistent failure of many RCT's is often attributed to sub-optimal trial design, with many acute stroke trials failing to report a sample size calculation at all. Trials that did report sample size calculations often assumed unrealistic event rates and intervention effects or used inappropriate primary outcomes, resulting in an underpowered study[227]. However this may be unfair as it is difficult to predict event rates or intervention effects beforehand for a new therapy. Analysis of the primary outcome plays an important role in the sample size calculation. It has been shown that trials designed to use an ordinal analysis will, on average, be smaller than using a dichotomous outcome[223].

The CONSORT statement[228] gives a checklist of essential items to be included in reports of RCT's. It suggests that in a trial report the authors should indicate how sample size was determined and if a power calculation was used, the outcome this was based upon. It also makes suggestions on how to discuss errors made in the calculations that have been realised retrospectively. Since its inception the reporting of sample size calculations in acute stroke trials has improved[227].

1.8. Aims

The aim of this work was to employ various analytic techniques to a vast range of historical data allowing us to suggest more optimal methods for consideration in the design of clinical trials in acute stroke. Chapter two discusses some of the methodology used throughout the thesis. Chapter three aims to explore some simple prognostic factors often considered for trial inclusion and their relationship with outcome. These include age, onset to treatment time (OTT), the interaction between the two and atrial fibrillation (AF). In chapter four, the merits of using a later baseline measurement of NIHSS for trial inclusion is examined. A hyper-acute measurement and 24 hour measurement are compared under simulated trial conditions and will be assessed using different outcome measures in terms of statistical power and required sample size. Chapter five aims to utilise a previously published prognostic score to investigate the selection of a subgroup of subjects in whom thrombolysis is safe and effective above 4.5 hours. Suitable boundaries of this score were found using a dataset with patients treated at < 3.5h but sampled to match the published outcome distribution of those treated > 4.5h. These prognostic boundaries were then validated on external randomised data with patients treated > 4.5h.

Chapter 6 considers the reduction in sample size achieved by utilising an earlier endpoint in exploratory stroke trials. This was tested on non-randomised historical data using a simulation approach and the results validated on RCT data. In chapter 7 we explored whether a combination of endpoints incorporating both early and late measurements of outcome could enhance statistical power. A simple incremental combination as well as the global test statistic were utilised here in simulations similar to those employed in chapter 6. In chapter 8 an evaluation of different propensity score matching techniques is given. This chapter aims to highlight the benefits and flaws of using a matched control group instead of a fully randomised control population. Finally chapter 9 aims to use historical data to provide reliable estimates of intracluster correlation coefficients (ICCs) for specific outcome measures for sample size calculations for future cluster randomised trials in acute stroke.

Chapter 2
Data and Methods

2.1. The VISTA database

2.1.1. What is VISTA

The VISTA database is a large collection of stroke data, including data from several pivotal trials in acute stroke. Eligibility criteria are applied to individual datasets before they are allowed entry into VISTA. These criteria have been put in place to facilitate the compatibility of data from multiple sources and include specific details regarding documentation of the study, relevant baseline and outcome assessments and monitoring procedures in place to validate the data. More detail regarding these criteria are outlined in the paper by Ali et al[204].

VISTA contains in excess of 28,000 patients with a general aim to promote research into stroke care and clinical trial design. The premise of using trial data to make inferences about future trials is not new; several studies have investigated aetiology, risk factors and subgroup differences in stroke outcome on individual trial datasets[146 214 229-235]. The benefit of utilising databases such as VISTA as a resource for novel analysis is the wealth of data contained within. As VISTA is comprised of trial data, baseline demographics, treatment definitions and outcome measures are carefully recorded. Time points are more reliable in this type of data due to individual trial data monitoring practices and pre-defined statistical protocols. This gives us a large and diverse study population for use in the analysis of different aspects related to stroke.

To maintain the anonymity of subjects within and the integrity of the VISTA database stringent guidelines have been put in place detailing handling of confidential patient information, ethics, representation and publication[236]. While VISTA aims to facilitate the planning of future randomised trials in stroke, it does not permit the re-analysis of treatment effect in any individual trial. This is primarily to encourage those who hold the data to contribute to VISTA without the risk of any original trial results being challenged. Individual trial meta-analysis is also not permitted.

A steering committee was created containing principal investigators from all contributed trials. Any project proposal applying for the use of VISTA is reviewed by all members of the steering committee, allowing them to accept or reject based on scientific merit, give comments or opinions and providing the opportunity for collaboration. This review system ensures that only proposals above a certain standard progress through to the analysis stage and avoids any conflicts or direct competition of projects within VISTA. A similar review system is also in place for any manuscripts to be submitted for publication with the steering committee members giving comments and opinions on the publication which must be addressed before submission to a journal.

Since its inception VISTA has led to a number of publications and has identified several areas for improvement in current practice. This success has led to the extension of VISTA to other forms of stroke research such as ICH, imaging and rehab[203]. For the purposes of this thesis VISTA refers to VISTA-acute only as none of the other arms of VISTA have been utilised.

Databases such as VISTA allow retrospective analysis on a larger cohort of patients than any individual trial dataset. The distribution of baseline and outcome variables contained

within VISTA has been described previously[237]. Topics previously studied and published within VISTA include the difference in outcomes between countries, adverse events post stroke[238-239], healthcare cost[240], the influence of different baseline factors on outcome[63 141 241-243], the use of home time as an outcome measure[244-245] and the effect of different attributes of medical history on outcome post stroke[246]. Some analyses have investigated imaging[247-249] but the imaging data within VISTA acute limited the scope of these analyses. With the creation of VISTA Imaging[203] individual scans are becoming more readily available for novel analysis.

2.1.2. Caveats associated with VISTA

While there are many benefits associated with the use of a resource such as VISTA there are some caveats that need to be considered.

Certain restrictions are applied to VISTA datasets before they are released for an individual project; these are mainly for the protection of those who contributed the data. The data are anonymised with respect to both individual patient and trial. VISTA does not sanction re-analysis of any individual trial that will test treatment effects. Due to these restrictions, data compiled from the VISTA database are non-randomised between treatment groups. For example: if we wish to investigate the effect of thrombolysis on outcome we cannot use patients from trials in which thrombolysis was the drug of interest. This leaves us with patients who were treated with thrombolysis as standard of care in trials with other investigational drugs. A comparison of the baseline factors of this group with a non-thrombolysed group of VISTA patients would likely show the two groups to be unbalanced. This means that the data are not representative of a randomised trial and must be taken into consideration when discussing results of an analysis.

As VISTA contains a collection of clinical trial data, inclusion criteria have been applied to the patients before entry. Each trial has different inclusion criteria based on the intervention being tested. This may lead to certain subgroups of patients being under-represented within the database. For example in order to avoid treating patients in whom the stroke was very mild or very severe there are usually upper and lower boundaries of baseline NIHSS applied to the patients before inclusion into a trial. Similarly there is often an upper and lower age restriction in trials, for example the ECASS I and II trials had a lower age limit of 18 and an upper age limit of 80 years[98 166].

Finally, caution must be taken when making recommendations based on historical data, with the development of new treatments and the expansion of stroke units data would be expected to evolve over time. However, VISTA acute contains data from patients entering into a trial during the years 1989 to 2006. With data readily available on thrombolysis given as standard of care as well as thrombolysis trials, VISTA is considered to represent current trends in acute stroke. These caveats must be taken into consideration when making inferences following an analysis.

2.1.3. What do we use it for?

Various VISTA data compilations have been used for the projects contained in this thesis. Each project was approved by the VISTA steering committee. Each dataset contains a unique selection of patients combined for the project purpose, data descriptions are given within the relevant chapter. Given the restrictions placed upon VISTA projects, the data given were non-randomised and so not representative of a true trial population. In some cases, where possible, methods were validated on external RCT data.

VISTA data has been used here to explore trial design in acute stroke, in particular to investigate how inclusion criteria and outcomes influence sample size. This has been achieved using various analytical approaches including regression, simulation and propensity score matching methods.

2.2. Simple statistical techniques

2.2.1. Displaying data graphically

With the inherent non-parametric nature associated with stroke data it is important to ensure the data are displayed correctly. When looking at continuous or discrete variables such as age and NIHSS within stroke data, the distribution is generally skewed to favour higher ages and lower values of NIHSS. As the data are generally non-normal in nature they are best displayed using either histograms or box plots. In medical literature dot plots are often used in place of these when there is a small sample size. Due to the large sample given in VISTA datasets these were not utilised here. Figure 2-1 shows a histogram and a box plot for NIHSS at day 90 taken from a sample of VISTA data. As can be seen in the plots the data are positively skewed i.e. the majority of the data are centred on lower values and tail off toward the higher values. Data displayed in this way are easy to interpret and the non normal nature of the data is clear.

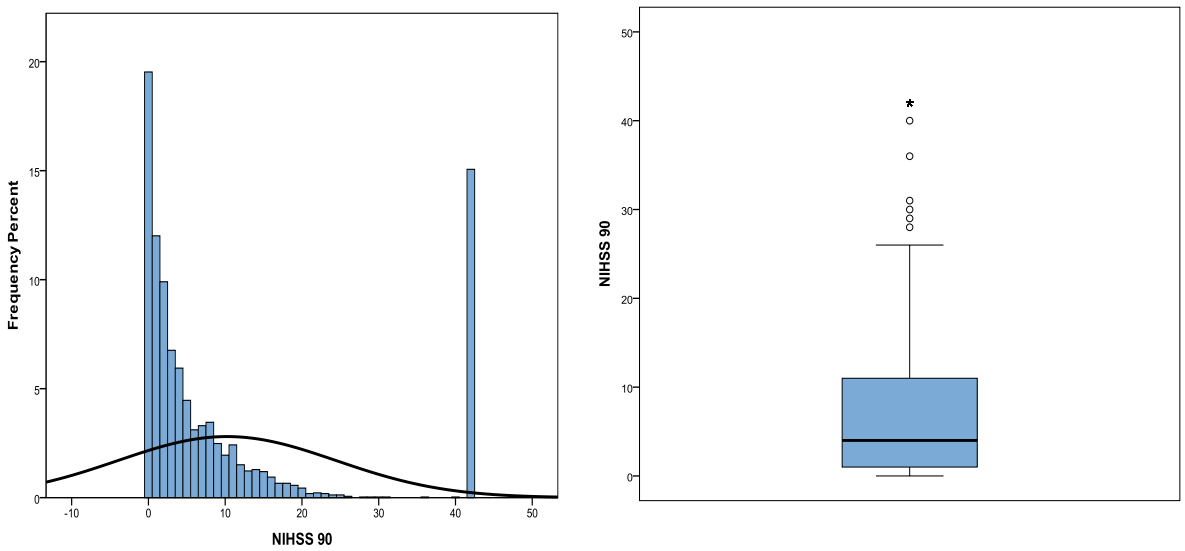

Figure 2-1: Histogram and box plot illustrating the distribution of NIHSS at day 90

For ordinal data such as the mRS these plots are uninformative and can lead to wrongful interpretations of the data. A method of displaying ordinal data such as the mRS which have a select number of categories is to use bar charts. A method predominantly used in stroke research is a horizontal bar chart for each treatment group with vertical lines in between categories for each grade of the mRS. This is informally known as a 'Grotta bar' and example is given in Figure 2-2. The numbers given in each box represent the percentage of subjects from each treatment group within each mRS grade.

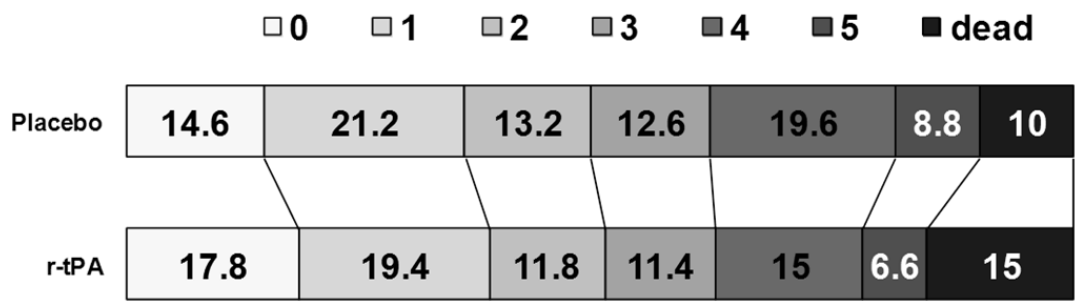

Figure 2-2: Bar chart or 'Grotta bar' illustrating the distribution of mRS between treatment groups

The graphical representation of data gives a good introduction to any analysis and its importance is often underrated.

2.2.2. Tests of association

2.2.2.1. Correlation coefficients and coefficient of determination (R^2)

Correlation coefficients are used to measure the strength of the association between two continuous variables. Correlation coefficients range from -1 to 1 where a perfect linear relationship with a positive slope would give a correlation coefficient of 1. If both variables have a high value the correlation coefficient is positive. A perfect linear relationship with a negative slope would give a correlation coefficient of -1. If one variable has a high value whilst the other has a low value the correlation coefficient is negative. A correlation coefficient of 0 indicates no linear relationship between the two variables. When calculating a correlation coefficient in statistical packages a p-value is given, this p-value assesses the hypotheses:

H_0: Population correlation coefficient (r) = 0

H_1: Population correlation coefficient (r) ≠ 0

There are at least two types of correlation coefficient that can be calculated. The Pearson correlation coefficient[250] is used under the assumption that the variables are jointly normally distributed so may be suitable for continuous variables such as age and weight. When considering discrete or ordinal variables, particularly those with a limited number of groups such as the mRS the Pearson correlation coefficient may underestimate the value of the association. In such cases a non-parametric measure can be used called the Spearman-Rank correlation coefficient. This assigns a rank to instances of each variable and gives a better indication of the relationship between the two variables when the relationship is non-linear[251].

SAS gives the option to calculate a partial correlation, adjusting the correlation coefficient for other covariates. This attempts to determine the degree of association between two variables that would exist if all of the influences from the adjusting variables were removed. Given this; the partial correlation coefficient represents the correlation between two variables after removing the common variance with other predictors from both variables [252].

The coefficient of determination, commonly referred to as R^2 is a value between 0 and 1 and, in general, is often used to indicate the goodness of fit of a regression model. It quantifies the proportion of variation in the outcome that can be explained by the regression equation[250]. When looking only at the relationship between two variables R^2 can be crudely calculated by simply squaring the correlation coefficient. As the absolute value of the correlation increases, R^2 increases regardless of the sign of the correlation. When considering other variables in the equation this becomes slightly more complex although squaring a partial correlation coefficient does give a crude estimate of R^2. When conducting a linear regression with a continuous outcome most statistical packages will

give the adjusted R^2 value representing the percentage of variation in the outcome explained by the predictors.

When looking at logistic regressions predicting ordinal and binary outcomes there is no alternative to the ordinary R^2 so something called a pseudo R^2 must be calculated[253-254]. These are called pseudo R^2 as they look like ordinary R^2. When calculated within SAS these measurements can be interpreted in the same manner[255].

2.2.3. Comparison between groups

In many of the projects contained within this thesis the initial stage of the project involved performing univariate analyses, investigating whether differences exist in a variable between groups. Different tests are used for continuous and categorical variables. SAS version 9.2/9.3 and R were used for all statistical analyses.

2.2.3.1. Tests for continuous variables

When considering a continuous variable first the nature of the variable must be taken into consideration. If the variable is normally distributed within each group, for example age, then a 2-sample t-test[256] can be performed. This assesses the difference in the population mean between the two groups. Using the example of age as the continuous variable and rt-PA as the grouping variable we obtain a p-value assessing the hypothesis:

H_0: In the wider population there is no difference in mean age between treatment groups

H_1: In the wider population there is a difference in mean age between treatment groups

If the p-value is lower than the accepted significance level of 0.05 then a significant difference in age exists between the two treatment groups. A 95% confidence interval is also given quantifying this difference. Other than normality there are two more

assumptions associated with this test, the observations are independent and the two groups have equal variances.

If the variable is not normally distributed within one or both groups then either a transformation could be considered such as the log or square root transformation or a non-parametric test can be performed. Due to the non-parametric nature of stroke data transformation of the data generally makes very little difference. In this case a Mann Whitney U[257] test can be performed to compare the distribution of the two groups. This test does not assume normality, allowing for the non-parametric nature of data such as NIHSS score. This test works by ranking the observations from low to high, disregarding their group. This gives a p-value assessing the hypothesis:

H_0: In the wider population, the underlying distribution of NIHSS is the same in both treatment groups

H_1: In the wider population, the underlying distribution of NIHSS is different in both treatment groups

If the p-value is lower than the accepted significance level of 0.05 then a significant difference in NIHSS exists between the two groups. From this test a 95% confidence interval can be generated quantifying the difference in median NIHSS between the two groups. The assumptions required for this test are that the data are at least ordinal and the observations are independent.

2.2.3.2. Tests for categorical variables

When comparing a categorical variable between two groups there are two tests that can be used. If the categorical variable is binary, a test of two proportions can be used[258]. The

test of two proportions compares the proportion of 'events' in each group, for example history of hypertension the 'event' would be 'Yes'. With P_1 denoting the proportion of events in the treatment group and P_2 denotes the proportion of events in the control group; the test gives a p-value assessing the hypothesis:

H_0: In the wider population, $P_1=P_2$

H_1: In the wider population, $P_1 \neq P_2$

If the p-value is lower than the accepted significance level of 0.05 then a significant difference exists between the proportions of those with history of hypertension between the two groups.

More widely used, particularly in medical literature, is the pearson chi-square (χ^2)[258] test of association. This test is used to assess the relationship between two categorical variables each with 2 or more groups. The χ^2 test treats each variable as nominal i.e. there is no order to the categories and gives a p-value assessing the hypothesis:

H_0: In the wider population, there is no evidence of an association between the two variables

H_1: In the wider population, there is evidence of an association between the two variables

This test assumes that the expected cell counts have a value of 5 or higher, if this assumption is not met Fishers exact test can be used.

2.3. Analysis of outcome measures

2.3.1. Analysing ordinal and binary outcome measures

The primary outcome in most clinical trials in acute stroke is the mRS. As mentioned above there is some debate as to how this should be analysed so we will discuss treating it as both a binary and an ordinal measure.

When outcome is treated as a binary measure, usually denoted good outcome=1 and poor outcome=0, binary logistic regression can be used to analyse outcome. When treated as an ordinal measure scales such as the mRS can be analysed using ordinal logistic regression. The significance can then be assessed using the Cochran Mantel Haenszel (CMH) test. Logistic regression allows us to quantify the predictive value of an explanatory variable on a response variable. Using regression methods allows for adjusted analysis, accounting for other covariates.

2.3.1.1. Binary logistic regression

Binary responses are often used in stroke research by dichotomising the outcome scale to create an outcome measure with two categories. Such a response is analysed using the binary logistic regression model[259]. This calculates the probability of an event Y=1, given the values of the predictors X. The model is in the form:

$$Prob\{Y = 1|X\} = [1 + e^{-X\beta}]^{-1}$$

Where $X\beta$is the weighted sum of the predictors; given below.

$$X\beta = \beta_0 + X_1\beta_1 + X_2\beta_2 + \cdots . X_k\beta_k$$

The regression parameters β are estimated using the maximum likelihood method and the log odds of an event occurring are given by the inverse of the logistic function P.

$$x = \log\left[\frac{P}{(1-P)}\right]$$

This can be explained as the log(odds that outcome $Y = 1$ occurs) also known as $logit\{Y = 1\}$. The model assumes that, for every predictor X_j,

$$logit\{Y = 1|X\} = \beta_j X_j + C$$

Assuming all other predictors are held constant, the parameter β_j is the change in the log odds per unit change in X_j. The odds ratio (OR) giving the odds that $Y = 1$ when X_j is increased by one unit can then be calculated by e^{β_j}.

2.3.1.2. Ordinal logistic regression with the proportional odds model

In stroke research the proportional odds model has gained popularity in recent years for analysis of outcomes such as the mRS[260]. The benefit of ordinal response models is that they do not assume equal spacing between levels of outcome (i.e. outcome can be 0,1,2 or 0,100,1000 the difference between each level is irrelevant). The ordinal models rank the values of outcome and order these ranks. For the purposes of this thesis we will consider only the proportional odds model[259].

For a response variable Y with levels 0,1,2,…k: the proportional odds model is given as follows:

$$P[Y \geq m|X] = \frac{1}{1 + e^{-(\propto_m + X\beta)}}$$

For a fixed m (where $m = 1,2,..,k$), the model becomes similar to a binary logistic model for the event $Y \geq m$. Using a common vector of regression coefficients β connecting the probabilities for varying m the proportional odds model allows for parsimonious modelling of the distribution of Y. For any predictor j the odds ratio for $Y \geq m$ is given by e^{β_j}, whatever the cut-off m. This single odds ratio is assumed to apply to all events.

One of the main drawbacks of using this model in stroke research is the ordinality assumption is often violated. This is a basic assumption of all commonly used ordinal regression models. It states that the response variable behaves in an ordinal fashion with respect to all predictors. This can also be explained as the effect of each predictor being consistent across all levels of outcome[259]. Due to the nature of the mRS this assumption is often not met. There is some debate in the literature as to the importance of the hypothesis test for this assumption[224 260-261].

2.3.1.3. Cochran Mantel Haenszel (CMH) test

In order to get a more conservative estimate of significance given the violation of the ordinality assumption; the CMH test[262] was used for ordinal outcomes. The CMH test measures the strength of the association between treatment and outcome, after stratifying on the observed covariates. This accounts for confounders in the analysis. The significance level reported is the probability that the response is independent of the explanatory variable, after adjusting for control variables. The hypothesis being tested is:

H_0: After adjusting for covariates the response is independent of treatment

H_1: After adjusting for covariates the response is dependent on treatment

2.4. Simulation techniques

Simulation techniques were used throughout this thesis as a method to estimate power and required sample size. A multiple re-sampling approach was applied with slightly different methods employed depending on the given analysis; specific details are given within each chapter. Here a general overview of the method is given.

For each simulation a relevant dataset or 'trial' needed to be generated, any baseline restriction desired e.g. baseline NIHSS restricted 4-20 was applied beforehand. Each simulation used a different seed for sampling; this allows future replication of results and ensures the random samples taken for each simulation are different. A selected number of patients (n) were sampled, with replacement, from all of the available control data to generate a placebo group and similarly form all available treated patients to generate a treatment group. This gives a simulated trial for analysis. Using R statistical programming language, a specified number of trials (in general between 1,000 and 10,000 were simulated from the given VISTA dataset.

For each of these simulated trials analysis was performed using both ordinal and binary outcome measures as endpoints. Logistic regression was performed to assess the effect of (simulated) treatment on outcome adjusted for the baseline used for inclusion in the trial and age, either binary or ordinal logistic regression were used depending on the outcome measure. The significance level for the simulated treatment effect along with OR for treatment and the distribution of mRS within each group (treatment and placebo) were recorded. Unadjusted analysis was also performed on each trial for exploratory purposes.

Power was obtained by calculating the proportion of times the treatment effect was found significant over the 10,000 trials. This can be expressed as 1-β where β is the probability of a type II error (false negative). This method is outlined in the flowchart given in Figure 2-3.

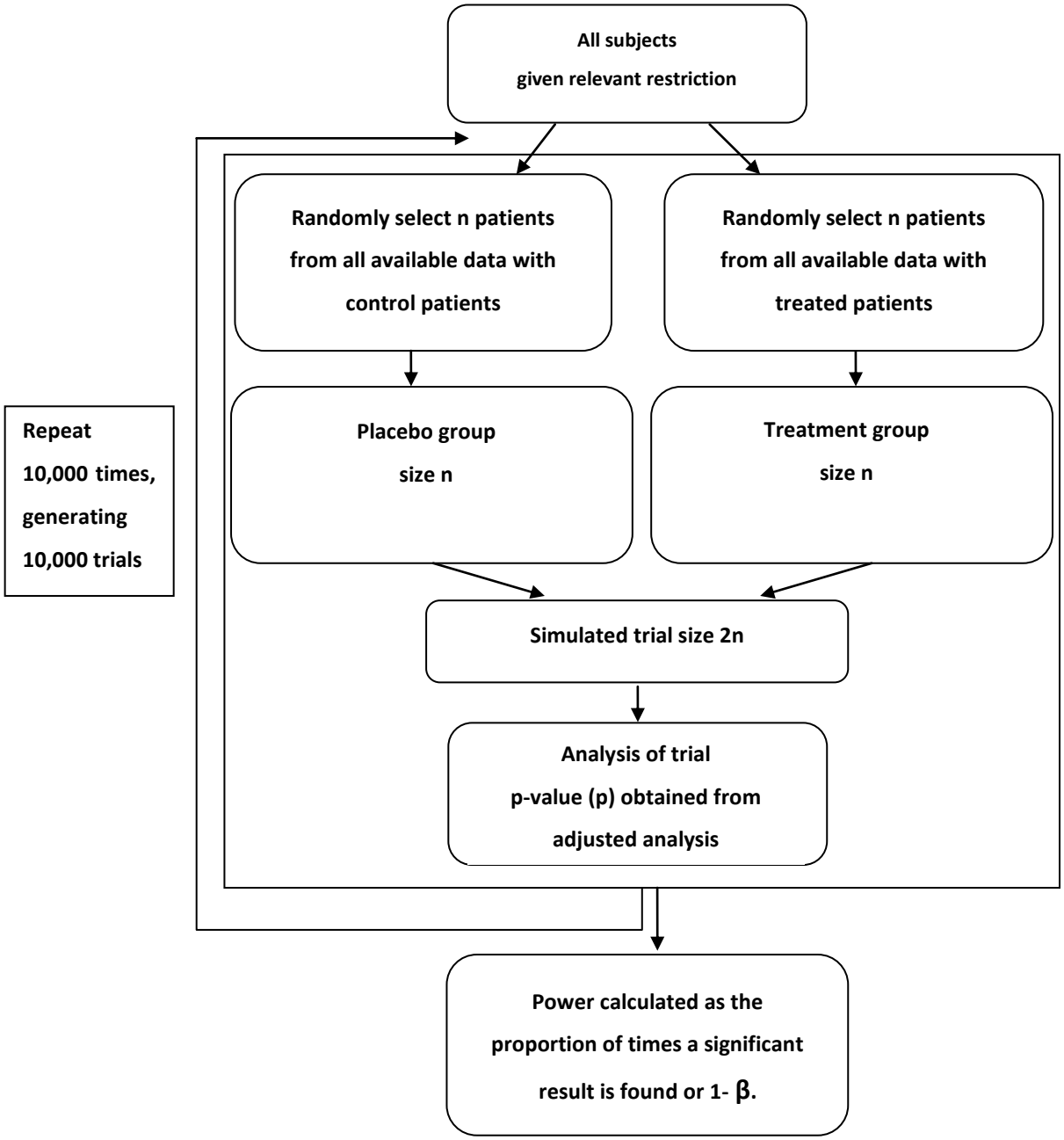

Figure 2-3: Outline of simulation method

In order to look at the relationship between power and sample size an iterative approach was taken. The sample size (n) was gradually reduced from a starting point (say 1000 per

treatment group) in decrements of 10 and the simulations described above were performed for every given sample size. The sample size was reduced until power settled to 80%, and this allowed estimation of the required sample size under all data restrictions.

2.5. Matching based on propensity scores

2.5.1. Propensity scores

An important problem in causal inference, particularly when using non-randomised data, is how to estimate treatment effects. Observational studies often provide the only available information to evaluate treatment effect. It is often the case that there are a large number of control patients in the study population with a small number of treated patients. Matching from a different population based on covariates is a method often used to account for this imbalance[263-265]. Matching allows us to generate a control group similar to the treatment group based on baseline characteristics.

Because matching exactly on the values of all covariates individually is difficult, particularly continuous covariates, alternative methods need to be pursued. Propensity score matching has gained popularity in recent years[266], matching based on all covariates to substitute for the absence of experimental controls.

The propensity score was introduced by Rosenbaum and Rubin[267] and is defined as the probability of receiving treatment conditional on observed pre-treatment covariates[268]. Let the binary variable Y denote treatment assignment taking the value 0 or 1, where 1 denotes treatment. Let X represent a vector of pre-treatment covariates. The propensity score $e(X)$ for a subject is defined as the conditional probability of that subject being treated given their covariates, X: $e(X) = P(Y = 1|X)$. This reduces the dimensionality of

the matching process when several covariates are involved by naturally weighting all covariates and yielding an unbiased one-dimensional estimate for matching.

Matching based on the propensity score balances the distributions of the observed covariates in the treated and control groups. Here the estimated propensity score is obtained by fitting a logistic regression model considering different combinations of pre-treatment variables. Among subjects with a given propensity score, the covariate distribution X is on average the same among the treated and control populations. Packages such as SAS and R have in built functions for propensity score matching.

2.5.2. Methods of propensity score matching

For the purposes of this thesis, propensity score matching was performed using the matchIt package[269] available within the R statistical programming language. There are several methods of propensity score matching which involve matching each treated unit to a suitable control unit. The methods of matching included in this package are exact, nearest neighbour, optimal, full and genetic matching. These methods are described in some detail below.

2.5.2.1. Exact matching

This method matching on exact values across all covariates, pairing each treated subject with one control for whom all values are identical. As mentioned above, implementing this method can be difficult when there are several covariates, limiting the number of possible matches, often resulting in no match being found.

2.5.2.2. Nearest Neighbour Matching

The desired number of matches, *r* can be defined here matching the r closest controls to each treatment subject. This is also known as 'greedy' matching, where the closest match for each subject is found from all unmatched controls. This is performed without trying to minimise the global distance measure.

2.5.2.3. Optimal matching

Contrary to 'greedy' matching the optimal matching method matches the samples by minimising the distance across all matched pairs. While it is more computationally intensive, optimal matching is felt to be better at minimizing the global distance measure. It tends to provide better balance when there is a limited sample of control patients for matching.

2.5.2.4. Full matching

Rather than matching a single control to each case the full matching method creates matched 'sets'. Each set consists of either one case unit with one or more matched control units or one control unit and one or more matched case units. This allows for a larger matched group than the other methods which have a fixed case to control ratio. Full matching minimises the weighted average of the estimated distance measure between each treated subject and each control subject within each subclass.

2.5.2.5. Genetic matching

This matching method uses a genetic search algorithm to apply weights to each covariate so that optimal balance can be achieved after matching is completed[269]. Genetic matching maximises the balance of observed baseline covariates across the matched treated and

control groups. Genetic matching has been shown to have more appealing properties than other matching methods, with better performance in terms of efficiency. However this increase in efficiency comes at the expense of computational time.

Chapter 3
Simple factors for consideration in trial inclusion

3.1. Introduction

With the inherent variability in stroke patients[270], several factors need to be considered for inclusion or exclusion when designing a clinical trial. Many of these factors strongly depend on the action of the drug under study. With neuroprotectant agents such as thrombolysis it is important to exclude the patients in whom the potential for harm could outweigh the benefit. For example patients in the ECASS III trial of thrombolysis[271] had to be treated within 3 to 4.5h of symptom onset and were excluded if they had a baseline NIHSS score > 25, they were older than 80 or younger than 18, had either previous stroke or diabetes or oral anticoagulant treatment was given prior to stroke. It is important to consider different prognostic factors for trial inclusion and how they may affect outcome.

The relationships between age, history of diabetes, previous stroke and the NIHSS at baseline and outcome have previously been studied within the VISTA database[63 178 241 272]. Within this chapter we investigated the interaction between onset time to treatment (OTT) and age and their combined effect on outcome and the relationship between Atrial Fibrillation (AF) and outcome.

3.2. Onset time to treatment and age

3.2.1. Background

Therapeutic benefit from thrombolysis in acute ischaemic stroke declines sharply as stroke onset to treatment delay (OTT) increases. The time window for safe initiation of thrombolysis closes around 4.5 hours from stroke onset[171]. Fears that elderly patients derive less benefit, encounter greater risk, or have a shorter therapeutic time window have limited the use of thrombolysis in this group of patients[273].

Investigations utilising both the VISTA database and the SITS registry have indicated that benefit from thrombolysis for acute ischaemic stroke does not deteriorate with advancing age[63 178 273], but the association with OTT in the elderly had not been examined due to absence of information on treatment delay in controls. The third international stroke trial (IST 3)[274], which has recently been published in combination with an updated meta-analysis of summary data, described the association of therapeutic effects with treatment delay (≤3 versus 3-6 hours) using surrogate information on OTT[275]. The surrogate information used was time to randomisation. This provided an estimate of OTT comparable between treated subjects and open control[275].

We applied a similar approach among VISTA patients with reliable surrogate information on OTT to assess outcomes among thrombolysis treated patients versus untreated controls across a range of OTT according to age.

3.2.2. Methods

3.2.2.1. Data source and patients

We collected demographics, clinical data, and functional outcome measures from neuroprotectant trials in ischaemic stroke conducted from 1998 to 2008 and held within VISTA[237]. Data were anonymised in relation to patients and trials. We excluded trials that had tested effects of thrombolysis or of any drug now known to influence outcome after stroke, and patients who lacked relevant baseline or outcome information: baseline National Institutes of Health Stroke Score (NIHSS), age, thrombolysis administration as standard of care, and modified Rankin Scale (mRS) at day 90. Only non-thrombolysed patients (control) having a recorded OTT of the investigational drug as well as thrombolysed patients (treatment) having a recorded OTT of thrombolysis ≥1h and ≤3.5h were included.

3.2.2.2. Statistical analysis

A nonrandomised adjusted comparison of outcomes was conducted, differentiating between patients aged ≤80 and >80 years as well as between patients who received thrombolysis versus patients who did not. Analyses were adjusted for age and baseline NIHSS, which has previously been justified in detail[141 241]. As this was a subset of the data used by Mishra et al[63] the results were expected to follow the same trend.

Our primary outcome measure was the mRS at 90 days, analysed across the whole distribution of scores with the use of the Cochran Mantel Haenszel (CMH) test. Odds ratios (OR) and 95% confidence intervals (CI) were computed using ordinal logistic regression. Differences in rate of mortality were assessed using binary logistic regression. Unadjusted baseline comparisons were conducted using the 2 sample t-test, the Mann

Whitney U test, or the chi square test depending on the distribution and nature of the data.

By treating OTT as continuous in a multiple logistic regression model, we established whether the effect of thrombolysis varies across OTT. The interaction between OTT and age along with the interaction between thrombolysis and age were investigated. Insignificant interactions were removed and the analysis repeated.

Looking at the data independently for those above and below 80 years of age is sub-optimal. The three-way interactions were repeated on the full data, treating age as a dichotomised variable (≤ 80 vs. > 80). This allowed us to see if the interaction between OTT and thrombolysis varied across the two sub groups of age. Finally the three-way interaction of age (continuous) variable, OTT and thrombolysis was modelled to investigate if age as a continuous variable influenced the interaction between OTT and thrombolysis.

Plots were generated to illustrate the effect of thrombolysis on full scale mRS at day 90 as well as mortality at day 90 across the entire range of OTT. At each time-point the OR and 95% CI for treatment effect was calculated using logistic regression adjusting for age, baseline NIHSS and the interaction between treatment and time. These plots were generated for different age ranges: ≤ 80 and > 80. Similar plots were generated for the entire population, adjusting for age as a continuous variable and baseline NIHSS.

Analyses were undertaken using SAS 9.2 and R version 2.10.0.

3.2.3. Results

3.2.3.1. Data and preliminary analysis

We collected data from 3603 ischemic stroke patients of whom 2341 (65.0%) were thrombolysed within 3.5h of stroke onset. 597 (16.6%) patients were aged >80 of whom 352 (60.0%) were thrombolysed. Detailed baseline characteristics are given in Table 3-1 with p-values for between group comparisons included.

Table 3-1: Baseline Demographics of patients split between age sub-populations studied.

	Control	Thrombolysis	P –value for difference	All
Sex (male) %				
Aged ≤80	58.3%	59.9%	**0.4068**	59.4%
Aged >80	39.6%	41.8%	**0.5958**	40.9%
Age (years) mean (SD)				
Aged ≤80	67.0 (10.1)	65.2 (11.5)	**<.0001**	65.8 (11.0)
Aged >80	85.0 (3.6)	84.8 (3.5)	**0.5472**	84.9 (3.5)
NIHSS baseline median (IQR)				
Aged ≤80	11 (7	13 (8)	**<.0001**	13 (8)
Aged >80	14 (10)	15 (9)	**0.0643**	14 (9)
Diabetes %				
Aged ≤80	23.7%	19.2%	**0.0036**	20.7%
Aged >80	18.4%	17.9%	**0.8652**	18.2%
Glucose baseline (mg/dl) mean (SD)				
Aged ≤80	137.9 (56.7)	132.2 (51.4)	**0.0084**	134.1 (53.3)
Aged >80	134.9 (48.4)	130.9 (46.3)	**0.3421**	132.5 (47.2)
Hypertension %				
Aged ≤80	74.9%	67.9%	**0.0002**	70.1%
Aged >80	78.2%	81.9%	**0.2967**	80.5%
SBP baseline (mmHg) mean (SD)				
Aged ≤80	154.8 (27.1)	153.8 (24.8)	**0.3581**	154.2 (25.7)
Aged >80	158.9 (24.6)	158.9 (27.1)	**0.9898**	158.9 (26.0)
DBP baseline (mmHg) mean (SD)				
Aged ≤80	84.8 (15.9)	82.8 (16.0)	**0.0021**	83.5 (16.0)
Aged >80	81.8 (16.2)	78.9 (17.1)	**0.0549**	80.2 (16.8)

3.2.3.2. Multiple logistic regression Analysis

Irrespective of age, day 90 outcomes were more favourable among thrombolysed patients than comparators. Adjusted for baseline NIHSS and age, the OR for improved outcome with thrombolysis measured by the mRS was 1.31 in patients aged ≤80 (95% CI 1.15-1.51, p<0.001; shown in Figure 3-1) and 1.46 in >80 year old patients (95% CI 1.08-1.97, p=0.001; illustrated in Figure 3-2). This is a subset of the analysis performed by Mishra et al[63], results given here are for illustrative purposes only.

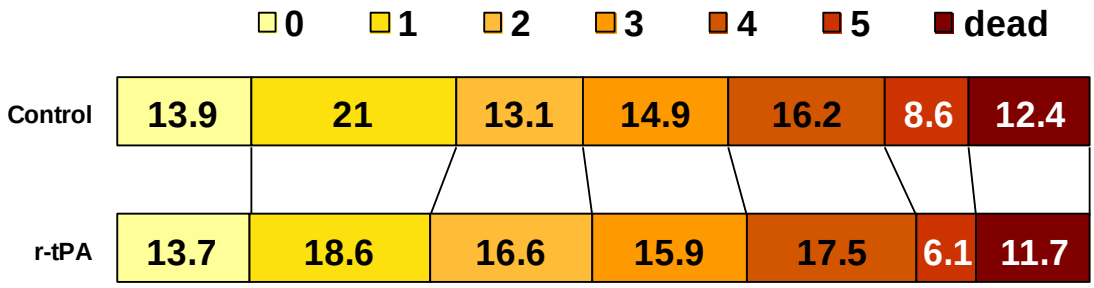

Figure 3-1: Distribution of mRS day 90 scores for each treatment group for those ≤ 80 years old

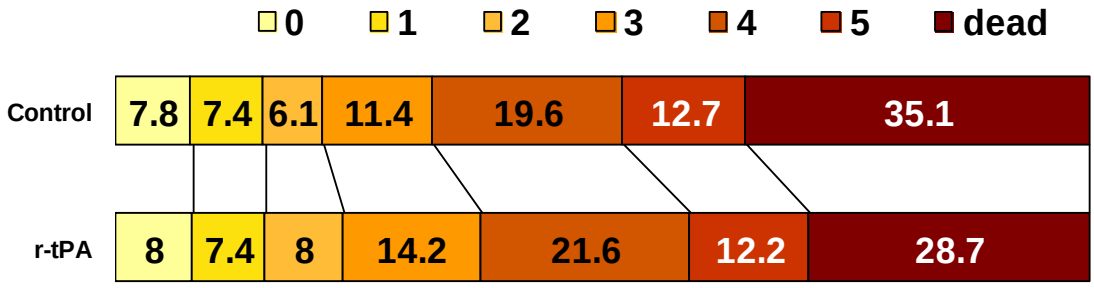

Figure 3-2: Distribution of mRS day 90 scores for each treatment group for those > 80 years old

When OTT was added into the ordinal logistic regression analysis, the interactions between OTT and other variables needed to be considered. We looked at the interaction between OTT and both age and thrombolysis as well as the interaction between age and

thrombolysis. The corresponding p-values are given in Table 3-2. Further details of the fitted models are given in Appendix A.

Table 3-2: Significance levels for all interactions in the ordinal logistic regression model. Outcome full scale mRS day 90

	Interaction OTT and thrombolysis	Interaction OTT and age	Interaction age and thrombolysis	Interaction OTT and thrombolysis after removing all other non-significant interactions	Interaction OTT, age and thrombolysis after removing all non-significant interactions
All Data	0.0194	0.7098	0.0755	0.0159	Age continuous = 0.5292 Age dichotomised= 0.2871
≤80	0.0076	0.1376	0.5231	0.0130	
>80	0.4499	0.6874	0.0851	0.4499	

The interactions with age and both thrombolysis and OTT were non-significant in all models and were removed. The models were then re-fitted to investigate the interaction between OTT and thrombolysis after adjustment for age and baseline NIHSS. This was repeated for mortality and results are shown in Table 3-3, the same pattern was observed. Further details of the fitted models are given in Appendix A.

Table 3-3: Significance levels for all interactions in the logistic regression model. Outcome mortality day 90

	Interaction OTT and thrombolysis	Interaction OTT and age	Interaction age and thrombolysis	Interaction OTT and thrombolysis after removing all other non-significant interactions	Interaction OTT, age and thrombolysis after removing all non-significant interactions
All Data	0.0070	0.7998	0.3010	0.0060	Age continuous = 0.0833 Age dichotomised= 0.2431
≤80	0.0049	0.1471	0.9856	0.0073	
>80	0.2118	0.7475	0.1508	0.2667	

The relationship between mRS90 and treatment with thrombolysis over time for the

entire range of OTT is shown in Figure 3-3 (A and B) for each age restriction, OR show the odds of favourable outcome. This was repeated for mortality and is shown in Figure 3-3 (C and D) for each age group, OR represent ting the odds of mortality by day 90.

Though in patients aged <80 thrombolysis showed decreasing favourable outcome with longer OTT (interaction mRS: p=0.0130; interaction mortality: p=0.0073), among patients aged >80 CIs were wider and there was no significant interaction (interaction mRS: p=0.4650; interaction mortality: p=0.2677).

The primary interest lies in whether the interaction between OTT and thrombolysis varies across age subgroups ≤ 80 and > 80. The three-way interactions for dichotomised age, OTT and thrombolysis gave significance levels of 0.2871 and 0.2431 for mRS day 90 and mortality day 90 respectively. This interaction also lacked significance when age was treated as a continuous variable (Table 3-2 and Table 3-3).

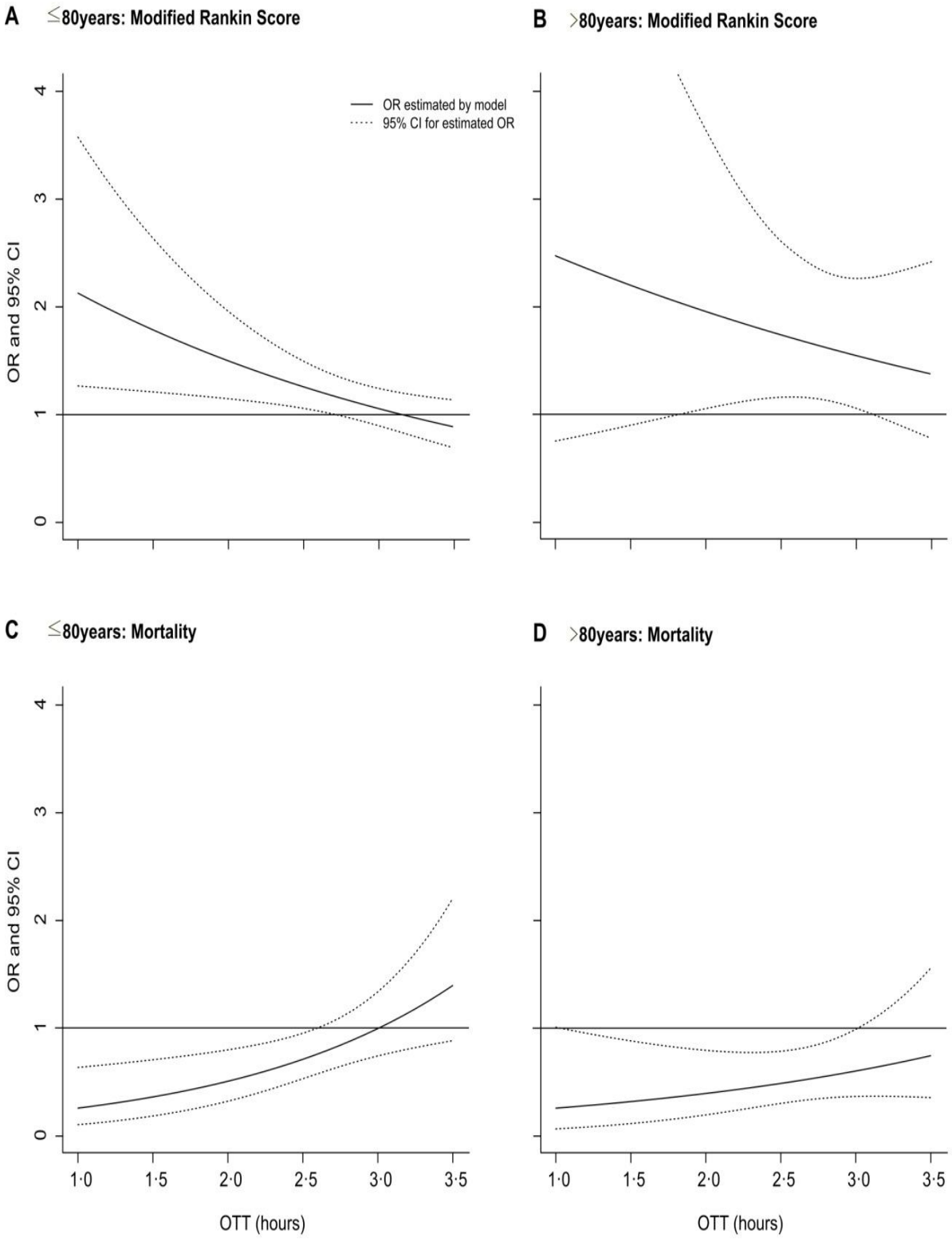

Figure 3-3: Relation of stroke onset to start of treatment (OTT) with treatment effect assessed by day 90 modified Rankin Score (A, B) and by day 90 mortality (C, D) after adjustment for baseline NIHSS and for age in patients aged ≤80 (A, C) and >80 (B, D).

The relationship between mRS90 and treatment with thrombolysis over time for all data is shown in Figure 3-4 (left), OR show the odds of favourable outcome after adjustment for age and baseline NIHSS. This was repeated for mortality and is shown in Figure 3-4 (right), OR representing the odds of mortality by day 90.

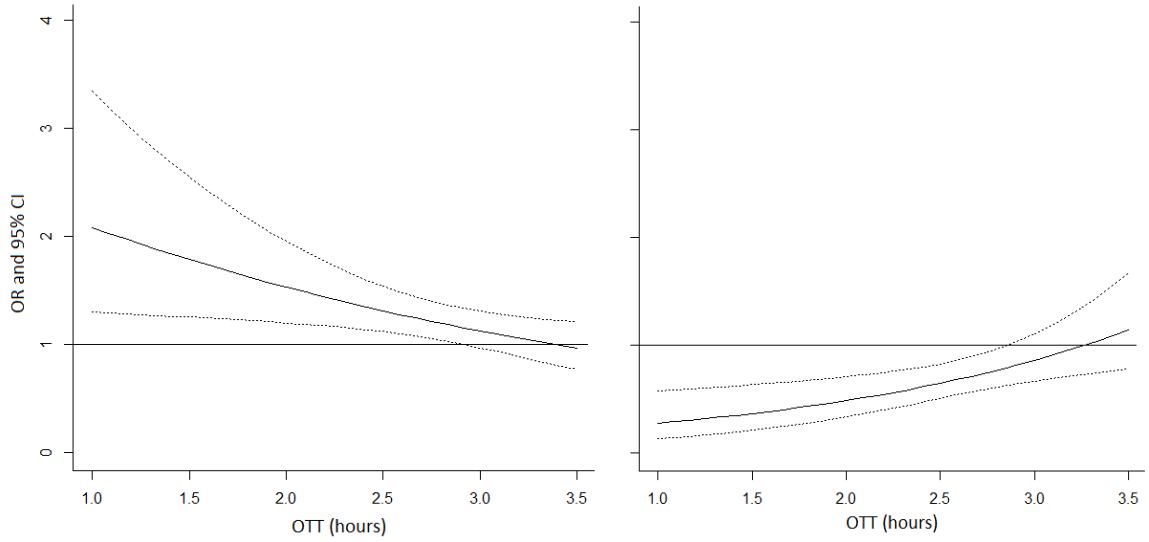

Figure 3-4: Relation of stroke onset to start of treatment (OTT) for all data with treatment effect assessed by day 90 modified Rankin Score (left) and by day 90 mortality (right) after adjustment for baseline NIHSS and age.

3.2.4. Discussion

We investigated the interaction between OTT and thrombolysis and the relationship to outcome in subgroups of patients aged ≤80 and >80. While in the elderly a significant interaction of OTT with thrombolysis treatment benefit was not seen, the estimate of the relation is reassuringly comparable for the elderly to that confirmed in younger patients.

IST-3 was planned to show the effectiveness of thrombolysis treatment in very elderly stroke patients[276]. The IST-3 investigators state that they lacked power to examine the relation between onset to treatment time and the effect of thrombolysis. With over 3000

patients, this may be due to heterogeneity of their sample rather than lack of numbers[172]. In an updated meta-analysis of summary data a significant treatment effect for thrombolysis is demonstrated in very elderly patients with initiation of treatment (OTT) within 3 hours after stroke onset, with no effect evident in data grouped across 3-6 hours[275]. Our analysis has only been possible by using the same compromise for estimation of OTT in control patients as the updated Cochrane meta-analysis[275]; however, our data provide powerful additional evidence that IST-3 could not offer.

The present analysis of non-randomised registry data, whether or not derived from trials, inevitably incorporates selection bias for thrombolytic treatment and other confounders, which have been previously discussed in detail[241].

Evidence from multiple sources is consistent. From this analysis we support the conclusion that thrombolysis is effective and safe in very elderly stroke patients until at least 3.5h after stroke onset.

3.3. Atrial Fibrillation

3.3.1. Introduction

One in three patients treated by thrombolysis within 3h of ischaemic stroke onset achieves significant benefit[277], raising the question whether a factor such as comorbidity would identify patients in whom treatment would confer no measurable advantage. Age, diabetes and prior stroke appear unlikely to influence treatment response but a comorbidity such as atrial fibrillation (AF) may be a candidate[178].

Atrial fibrillation (AF) is an independent risk factor for stroke and increases its incidence nearly fivefold[76]. The attributable risk of stroke for AF rises with age, from 1.5% for those

aged 50-59 years to 23.5% for those aged 80-89 years[76]. There are theoretical reasons for questioning whether thrombolysis will have a different effect among patients with AF versus sinus rhythm.

On the one hand, patients with AF may have old emboli which after reaching intracranial vessels are unlikely to dissolve with thrombolysis[278]; on the other hand, they might have soft emboli formed at ruptured sites of membranous thrombi and that are at least in theory amenable to thrombolysis, whereas cholesterol emboli from carotid lesions, occlusion of small vessels from hypertensive disease and other non-embolic causes of stroke should be less responsive to treatment. In association with the factors that may influence clot dissolution, the timing of restoration of blood flow and the abruptness could also have an influence on the risk of haemorrhagic transformation, with or without thrombolysis.

Thus, the question arises whether stroke patients with AF gain significant benefit from thrombolysis?

The few studies that have been conducted on this subject showed that stroke patients with AF more frequently had poor outcome after treatment with thrombolysis when compared to those without AF[278-281]. However, they did not compare with a control group which has not been given thrombolysis. Among the randomized controlled trials examining efficacy of thrombolysis in stroke patients, results for the AF subgroup are described only for the ECASS III and NINDS trials. In the NINDS trial, disregarding treatment groups, AF was significantly associated with worse global outcome (OR for more favourable outcome 0.57, 95% CI 0.38-0.86) but there was no significant interaction with treatment and further data were not reported[282]. In the ECASS III trial, there was a

non significant trend among the subgroup of patients with AF (Thrombolysis n=53, Control n=55) that again favoured placebo over thrombolysis treatment (OR 0.68, 95% CI 0.30-1.55)[271]. A small retrospective study compared stroke patients with AF who were (n=22) or were not (n=44) treated with thrombolysis. Here, by contrast, there were significantly improved odds of favourable outcome for patients with AF when treated with thrombolysis (OR 2.67, 95% CI 1.06-6.74)[283].

3.3.2. Methods

3.3.2.1. Data source and patients

We gathered demographics, clinical data and functional outcome measures from neuroprotection trials in ischaemic stroke conducted in the period 1998 to 2008. We obtained our data, anonymised in relation to patients and trials, from the Virtual International Stroke Trials Archive (VISTA)[204]. We excluded trials that had tested effects of thrombolysis or of any drug now known to influence outcome after stroke. We excluded patients who lacked relevant baseline and outcome information: baseline National Institutes of Health Stroke Score (NIHSS), age, history of AF, thrombolysis administration as standard of care, occurrence of adverse events (AE) as well as serious adverse events (SAE), and modified Rankin Scale (mRS) day 90. Death was recorded as mRS grade 6. Symptomatic recurrent strokes and symptomatic ICHs (SICH) were defined as any stroke/ICH with neurologic deterioration, as indicated by an NIHSS at 24 hours that was higher by 4 points or more than the value at baseline, or any stroke/ICH leading to death. Patients receiving oral anticoagulation at stroke onset or up to 3 days post-stroke were excluded.

3.3.2.2. Statistical analysis

A nonrandomized adjusted comparison of outcomes was conducted, differentiating between patients with and without history of AF and between patients who received thrombolysis versus patients who did not. Analysis was adjusted for age and baseline NIHSS which has previously been justified in detail[241 284]. As VISTA is anonymised for trial source, the precise definition of AF could not be established, though few acute trials define this in any case.

Our primary outcome measure was the distribution of mRS at 90 days using the full scale [241]. For comparison with prior trials and reports, dichotomized outcomes at 90 days (mRS 0-1, mRS 0-2, NIHSS 0-1 and mortality) were reported.

Unadjusted baseline comparisons were conducted using the 2 sample t-test, the Mann Whitney U test, the 2 proportions test or the chi squared test depending on the distribution and nature of the data. Correlations (r) were calculated using the Spearman rank correlation coefficient, which takes into account the ordinal nature of the mRS. The coefficient of determination (r^2) was calculated from these correlations.

Reported odds ratios (OR) express the odds of improved outcome in association with thrombolysis treatment or presence of AF respectively, adjusted for specified covariates. OR's and 95% confidence intervals (95% CI) of ordinal outcome measures were computed using ordinal logistic regression and p-values computed using the Cochran Mantel Haenszel test. Dichotomised outcome measures were assessed using binary logistic regression.

The significance of the interactions between AF and Thrombolysis; AF, age and thrombolysis; and age on mRS at day 90 were investigated. Plots of the effect of AF and

effect of thrombolysis (AF patients only) on full scale mRS day 90 against age were constructed to investigate the effect over age.

Analyses were undertaken using SAS 9.2, and R version 2.10.0.

3.3.3. Results

3.3.3.1. Data and preliminary analysis

We collated data from 7091 ischemic stroke patients of whom 3027 (42.7%) had been thrombolysed within 3h of stroke onset. 1631 (23.0%) patients had known history of AF at time of hospital admission of whom 639 (39.2%) were thrombolysed. Patients with AF were much older (mean 7.5 years) and had a higher baseline NIHSS (median 2 points). Age accounted for 8% ($r=0.28$) and baseline NIHSS for 24% ($r=0.49$) of the variation in mRS at day 90 (both $p<0.001$) and were included in all models (together r^2 is approximately 29%). Detailed baseline characteristics are given in table 1 including p-values for between treatment group comparisons and between AF group comparisons.

Table 3-4: Baseline characteristics for patients included in analysis. The difference between treatment groups was investigated for AF and Non-AF patients separately; p-values are given in the right hand column. Similarly the difference between AF and Non-AF patients was investigated for each treatment group separately; p-values given underneath each variable.

	Control	Thrombolysis	P-Value for difference
Sex (male); %			
AF	47.6%	47.3%	**0.986**
Non AF	55.3%	58.2%	**0.038**
P-Value	***<0.001***	***<0.001***	
Age (years); mean (SD)			
AF	73.9 (9.7)	74.2 (9.5)/639	**0.552**
Non AF	67.1 (12.2	65.7 (12.5)/2388	**<0.001**
P-Value	***<0.001***	***<0.001***	
NIHSS baseline; median (IQR)			
AF	14 (9)	15 (9)	**<0.001**
Non AF	11 (8)	13 (9)	**<0.001**
P	**<0.001**	**<0.001**	
Diabetes; %			
AF	21.3%	18.5%	**0.167**
Non AF	23.8%	17.8%	**<0.001**
P-Value	***0.092***	***0.729***	
Glucose baseline (mmol/l); mean (SD)			
AF	7.7 (2.9)	7.3 (2.6)/	**0.017**
Non AF	7.8 (3.6)	7.5 (3.1)	**0.001**
P-Value	***0.428***	***0.407***	
Hypertension; %			
AF	75.5%	71.7%	**0.094**
Non AF	67.5%	61.6%	**<0.001**
P-Value	***<0.001***	***<0.001***	
SBP baseline (mmHg); mean (SD)			
AF	154.6 (26.4)	154.6 (26.6)	**0.978**
Non AF	156.5 (26.9)	153.8 (25.4)	**<0.001**
P-Value	***0.069***	***0.509***	
DBP baseline (mmHg); mean (SD)			
AF	84.9 (16.7)	82.0 (17.9)	**0.002**
Non AF	84.7 (15.8)	82.7 (15.7)	**<0.001**
P-Value	***0.827***	***0.383***	
Myocardial Infarction; %			
AF	10.6%	13.9%	0.053
Non AF	11.5%	12.6%	0.248
P-Value	***0.457***	***0.366***	

3.3.3.2. Analysis

Irrespective of AF status, the mRS scores at day 90 had a more favourable distribution among thrombolysed patients than comparators as illustrated in Figure 3-5 and Figure

3-6. Adjusted for baseline NIHSS and age, the OR for improved outcome with thrombolysis was 1.44 in AF patients (95% CI 1.12-1.73, p<0.001) and 1.53 in non-AF patients (95% CI 1.39-1.69, p<0.001).

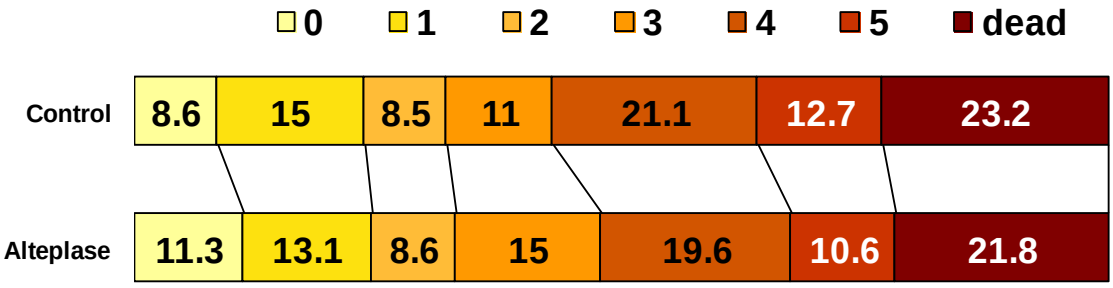

Figure 3-5: Distribution of mRS day 90 for those with history of AF

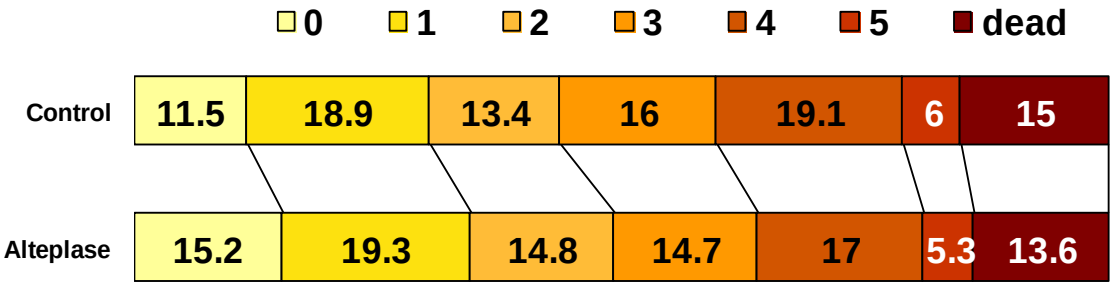

Figure 3-6: Distribution of mRS day 90 for those without history of AF

Presence of AF had no influence on mRS at day 90 after adjusting for baseline NIHSS and age (OR 0.93, 95% CI 0.84-1.03, p= 0.409).

Analyses with dichotomised outcomes are presented in Table 3-5. These are presented for each individual outcome as the OR for good outcome given history of AF within in each treatment group (given within the rows) and corresponding p-value. The OR for treatment effect for AF only patients is also given with the corresponding p-value in the final two columns of the table.

Table 3-5: Dichotomised outcomes

	Control	Thrombolysis	OR (95% CI)	P-Value
mRS 0-1 at 90 days; n/N (%)				
AF	234/992 (23.6%)	156/639 (24.4%)	**1.42 (1.08-1.85)**	**0.011**
Non AF	936/3072 (30.5%)	826/2388 (34.6%)		
OR (95% CI)	1.26 (1.04-1.52)	0.94 (0.75-1.82)		
P-Value	**0.020**	**0.615**		
mRS 0-2 at 90 days; n/N (%)				
AF	318/992 (32.1%)	211/639 (33.0%)	**1.43 (1.11-1.83)**	**0.005**
Non AF	1348/3072 (43.9%)	1179/2388 (49.4%)		
OR (95% CI)	**1.07 (0.89-1.28)**	**0.81 (0.66-1.00)**		
P-Value	**0.474**	**0.050**		
NIHSS 0-1 at 90 days; n/N (%)				
AF	234/758 (30.9%)	163/491 (33.2%)	**1.39 (1.07-1.82)**	**0.016**
Non AF	867/2598 (33.4%)	801/2051 (39.1%)		
OR (95% CI)	**1.36 (1.12-1.66)**	**0.99 (0.79-1.24)**		
P	**0.002**	**0.922**		
Mortality at 90 days; n/N (%)				
AF	230/992 (23.2%)	139/639 (21.8%)	**0.78 (0.60-1.01)**	**0.055**
Non AF	461/3072 (15.0%)	325/2388 (13.6%)		
OR (95% CI)	**1.08 (0.89-1.31)**	**1.07 (0.83-1.37)**		
P-Value	**0.443**	**0.607**		

When analysing outcome on the full scale mRS day 90 no interaction was found between treatment with thrombolysis and age among those with history of AF (p=0.671). Figure 3-7 (A) shows how the treatment effect of thrombolysis changes as age increases (adjusted for baseline NIHSS) for those with history of AF.

The presence of AF was possibly associated with more favourable outcome among younger patients and less favourable above approximately 75 years, after adjusting for baseline NIHSS and treatment with thrombolysis (illustrated in Figure 3-7 (B)). The interaction between AF and age had borderline significance (p=0.010).

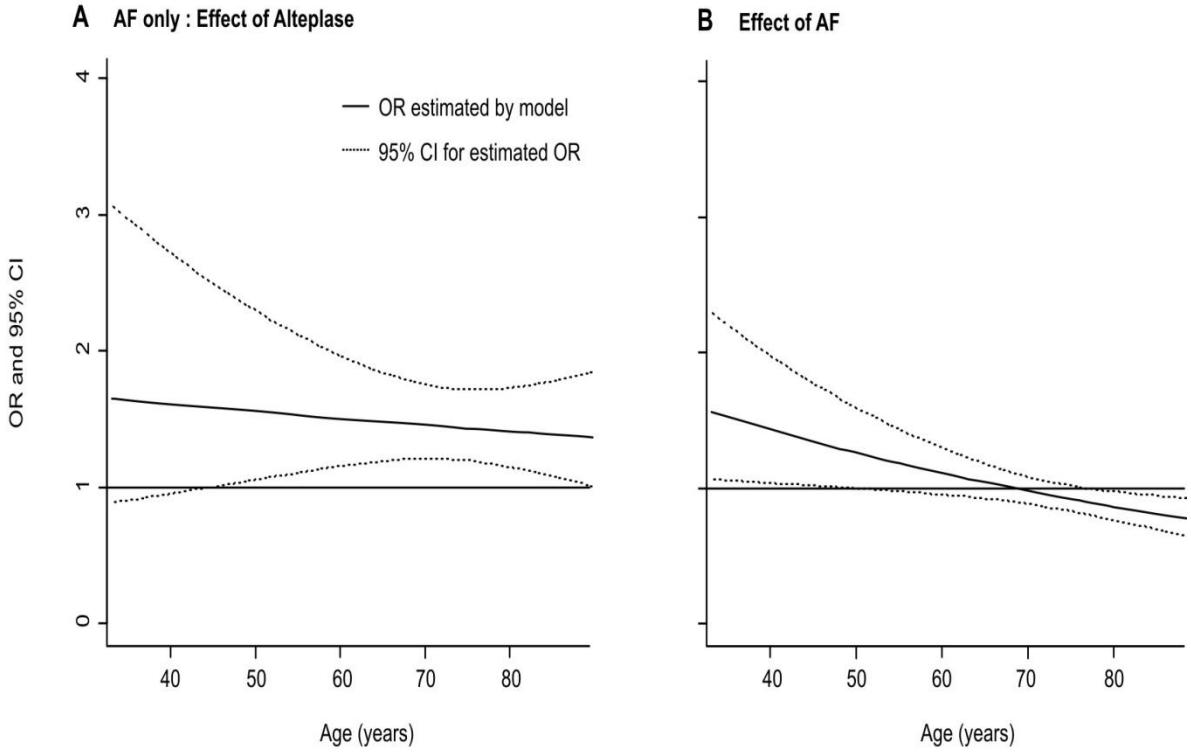

Figure 3-7: Odds of more favourable outcome across the full range of 90 day mRS for each year of age. This is presented (A) in patients with AF, contrasting outcomes in patients treated with thrombolysis versus not and (B) in all patients, contrasting outcomes for patients having AF versus normal rhythm.

3.3.4. Discussion

Despite being older and more severely affected, and thus at risk of poorer outcome than non-AF patients, stroke patients with AF had better outcome at 90 days when treated with thrombolysis than their untreated peers. No interaction between AF and thrombolysis treatment on outcome was found. Our results were supported by the range of secondary dichotomised measures that we examined.

However, it has to be stressed that this postulation just is valid for the group of AF patients as a whole. When looking at the interaction of age and AF on stroke outcome across the entire age range, our data indicate that AF seems to have a beneficial effect in young and a negative effect in elderly patients. This is in line with prior studies showing that age has an independent and additive effect on the hypercoagulable or prothombotic

state in AF[285-286]. Nevertheless, it remains unclear why in the present sample AF seemed to have a beneficial effect in younger stroke patients.

It remains important to acknowledge that an analysis of non-randomised registry data, whether or not derived from trials, inevitably incorporates selection bias for thrombolytic treatment and other confounders. These have been previously discussed[241].

It is important to note that many subjects who present with medical history of AF at stroke onset may currently be on anticoagulants such as Warfarin. Current treatment with anticoagulants can increase the risk of a secondary bleed and in turn may influence the decision to thrombolyse. This could lead to the non-thrombolysis group within the AF population being those who are potentially at higher risk of complications. Further work investigating the relationship between anticoagulant therapy and thrombolysis within the AF population is currently being conducted within VISTA.

3.4. Conclusion

In summary, these data lend support to the use of thrombolysis across all age groups of stroke patients with history of atrial fibrillation within approximately 3.5h after stroke onset. AF appears to be a marker of high age and baseline NIHSS rather than an independent risk factor for poor stroke outcome, and has no discernible impact on the effect of thrombolysis.

Chapter 4
Effect on study sample size of an extended time window for initiation of neuro-restorative therapy

4.1. Background

Functional outcome after acute stroke, such as modified Rankin Scale score after 90 days, is highly variable[129-130]. This reflects diversity in location and extent of initial brain damage, variation in age and comorbidities that influence potential for recovery, imprecision of the tools available to assess clinical status, and the inherently unpredictable clinical course in the hyperacute period of stroke. The first few hours after an ischaemic insult are characterized by substantial rates of complete recovery, but a few patients progress to coma or death. Certain measures of residual disability at 90 days such as Barthel Index reflect this as a bimodal distribution. Prognostic indicators of outcome available in the early hours after a stroke lack precision and reliability.

In the absence of perfect predictions of outcome, acute stroke trials will have inadvertently included patients who have little prospect to benefit from investigational therapy. Moreover, the reliance on statistically less powerful dichotomisation analyses in

many instances further compromises trial power, compared to use of continuous variables.

This combination of imprecise patient selection and weakened endpoint analysis inflates the sample sizes for stroke trials assessing early interventions. This renders proof-of-concept, or "early phase" trials unattractively expensive, or unreliable if underpowered. In turn, this limits the opportunities to explore novel treatments and increases the risk that small but useful therapeutic advances are missed.

Thrombolytic or neuroprotective agents administered to reduce the impact of ischaemic stroke must be given soon after the acute episode to prevent irreversible neuronal loss. However, a restorative treatment modality that can enhance recovery of surviving neurons may have applicability even if initiated after more severely compromised cells have already suffered irreversible damage or death. In contrast to neuroprotective strategies, restorative strategies may offer potential benefit if initiated later during the ischaemic cascade, for example 24-48 hrs after stroke onset. We aimed to investigate whether deferred selection of patients and outcome analysis that is adjusted using later, more reliable, estimates of prognostic factors would enhance trial efficiency, where the mechanism of treatment offers biological support for this strategy.

4.2. Data source and patients

We used the following individual patient data from the Virtual International Stroke Trials Archive (VISTA) for patients with ischaemic stroke enrolled in trials conducted between 1998 and 2008 [204]: National Institutes of Health Stroke Scale (NIHSS) score within 6h of stroke onset and at 24h; modified Rankin Scale score recorded at 90 days; age; exposure to iv thrombolysis (or not) as standard of care. We accepted only patients who had no

investigational drug administered that has a confirmed influence on stroke outcome (i.e. non-significant trials testing neuroprotectants and only placebo patients from thrombolysis trials). We excluded patients missing baseline NIHSS, age or mRS day 90 unless due to death, which was recorded as mRS grade 6 and NIHSS 42.

4.3. Outcome measures

The key efficacy endpoint of interest was the full range of the modified Rankin scale at 90 days.

4.4. Methods

4.4.1. Statistical analysis

We examined the relation between initial stroke severity and final outcome, contrasting predictions from NIHSS measured at <6h from stroke onset with those from the NIHSS collected after 24h. We sought to determine whether NIHSS scores from either time point provided more robust associations with 90 day mRS. We did so by examining the coefficient of determination (r^2) from basic correlations of NIHSS with mRS. The Spearman rank correlation coefficient was used to account for the ordinal nature of the mRS. Two different restrictions of NIHSS were used, 4-20 and 7-20. This allows investigation into a more restricted range for inclusion.

Ordinal analysis of the mRS to investigate the relationship between each NIHSS time point and mRS at 90 days was performed using logistic regression with the proportional odds model, significance was measured using the Cochran Mantel Haenzsel test. Analysis was adjusted for age and treatment with thrombolysis.

This analysis was initially performed on the entire data including thrombolysed patients. However as the 24h NIHSS will already be subject to the initial tPA treatment effect, further analyses were performed investigating the two groups separately. Note that trial centres may have applied additional selection criteria within the first 3-4.5h from stroke onset before offering thrombolytic therapy.

4.4.2. Simulation method

The goal was to simulate a treatment effect using VISTA data that was approximately half that typically observed due to thrombolytic therapy administered within 91-180 minutes of stroke onset. The distribution of mRS for this treatment time is shown in Figure 4-1. This treatment effect shows an improvement in mRS 0-2 of 10.8% with treatment and a decrease of 9.8% in mRS 3-6. Half of this is an increase of approximately 5.4% in mRS 0-2 and a decrease of 4.9% in mRS 3-6.

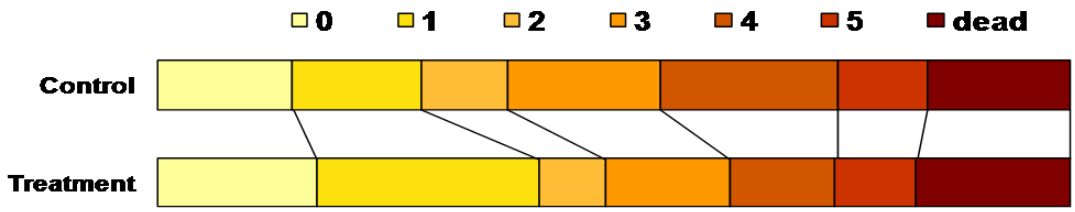

Figure 4-1: Distribution of mRS for treatment with tPA 91-180 minutes

We undertook simulations in a similar manner to those performed by Young et al[234 287] to model the sample size required to detect an improvement in overall mRS distribution equivalent to a 5% absolute difference in proportion achieving mRS 0-2 versus 3-6, setting power at 80% and assuming adjustment for entry age and NIHSS. Baseline NIHSS criteria of 4-20 and 7-20 were each investigated.

Using R statistical programming language, 10,000 mock trials were simulated for each NIHSS entry range. For each trial, a selected number of patients (2*n) were sampled from all eligible data, without replacement. This generated control and treatment groups of size n with similar baseline distributions for age and NIHSS at entry. For the treatment group a treatment effect was generated. To avoid biasing the adjusted analyses, the treatment effect was simulated using 90 day mRS rather than baseline covariates. The method for generating a treatment effect is outlined in Figure 4-2 and described below.

For each subject in the treatment group t_i (i=1,....,n) another subject m_i was matched with the probability of having 1 point lower mRS at 90 days (the probability selected here was 30%). The new treatment subject with simulated outcome is composed of the baseline variables from the subject t_i and outcome variables from the matched subject m_i, illustrated in Figure 4-2.

For each simulated trial, analysis was performed using ordinal methods to take into account the full scale of the mRS. Logistic regression using the proportional odds model was performed to assess the effect of (simulated) treatment on outcome adjusted for entry NIHSS used for inclusion and age. The significance level for the simulated treatment effect along with OR for treatment and the distribution of mRS within each group (treatment and placebo) were recorded.

Within each set of simulations, the sample size was gradually reduced until power settled to 80%, thus allowing estimation of the required sample size under each data restriction. Patients who had received acute thrombolytic treatment as standard care were considered independently. This treatment would likely influence the NIHSS score over the first 24h.

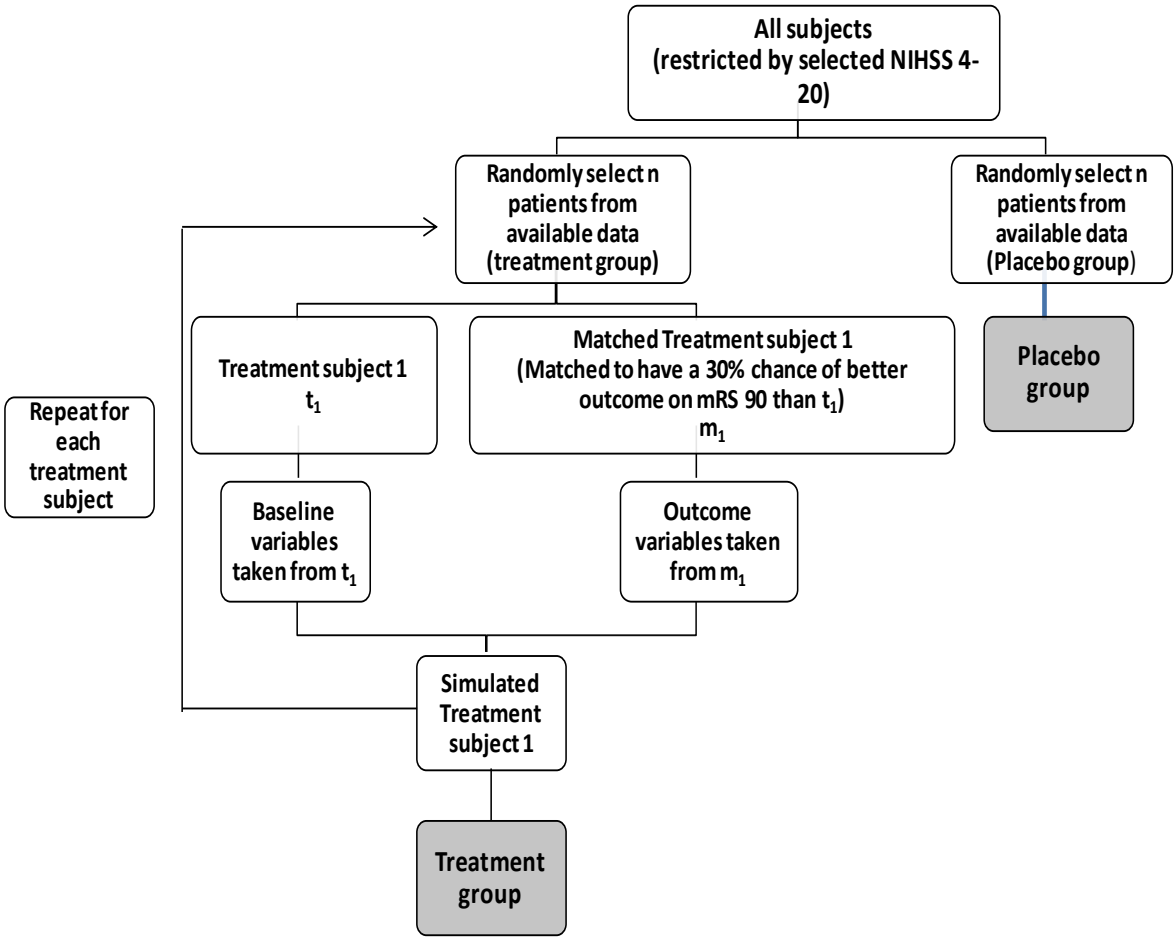

Figure 4-2: Flowchart outlining the method for generating a treatment effect within each of the simulated trials.

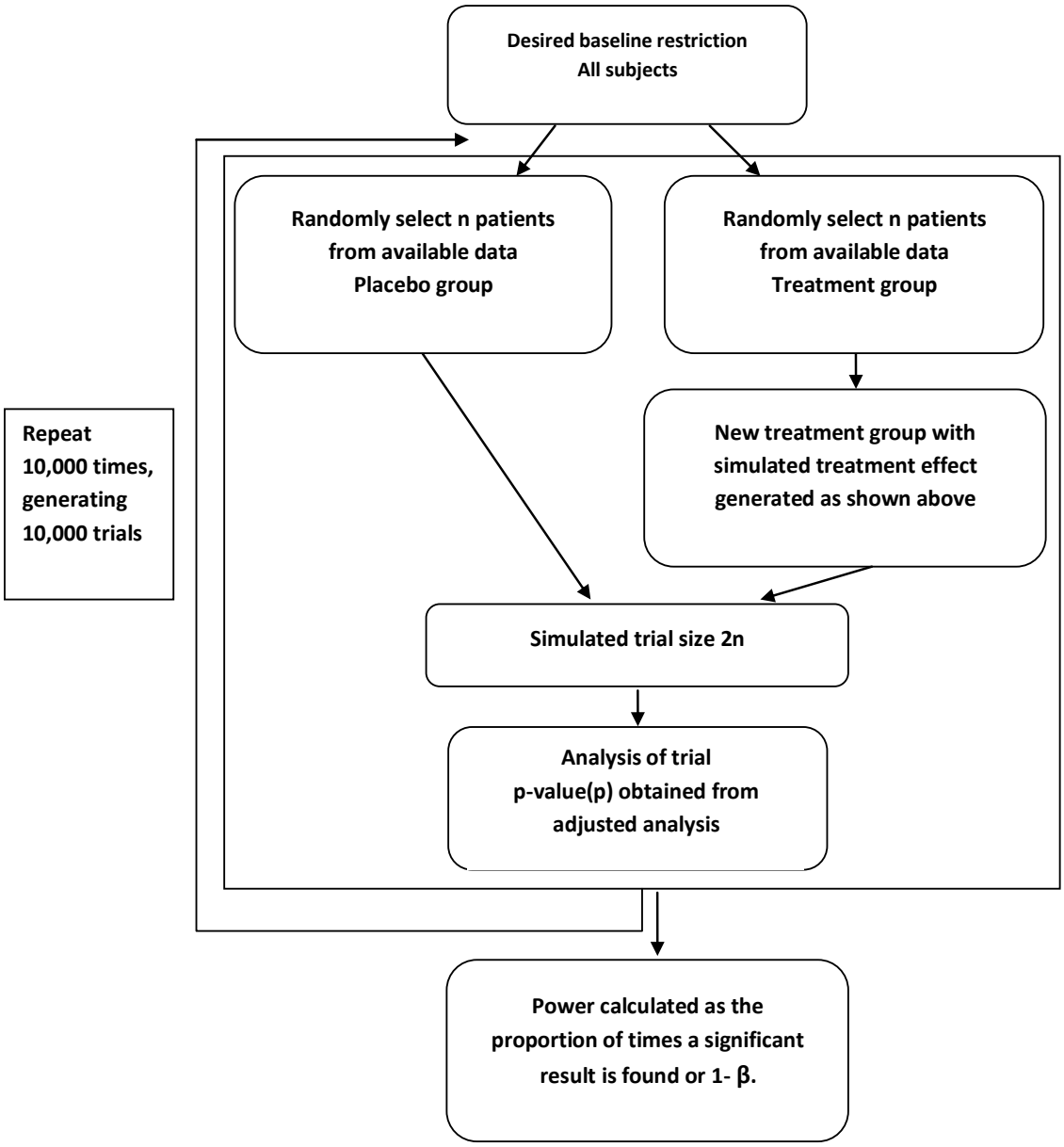

Figure 4-3: Flowchart outlining the method of trial simulation

4.5. Results

4.5.1. Preliminary analysis

VISTA provided data on 7636 patients who met our criteria, with mean age 68 ± 12; <6h median NIHSS 12 (IQR 8-17); 24h NIHSS 9 (5-15); 3316 (43%) had been treated with thrombolysis. Limiting the sample with each NIHSS restriction yields the altered samples given in

Table 4-1 below. Note here that the sample contains both thrombolysed and non-thrombolysed patients.

Table 4-1: Descriptive statistics of each restriction. N denotes number of observations, N* denotes number of missing observations.

Descriptive statistics for labelled NIHSS measurement							
NIHSS restriction	**N**	**N***	**Std dev**	**Mean**	**Median**	**Max**	**Min**
<6h NIH 4-20	6841	0	4.3	11.9	11.0	20	4
24h NIH 4-20	4436	0	4.7	10.3	10.0	20.0	4.0
<6h NIH 7-20	6137	0	4.0	12.6	12	20	7
24h NIH 7-20	3200	0	4.0	12.4	12.0	20.0	7.0

While using 24h NIHSS as opposed to NIHSS measured in the hyperacute phase (<6h) reduces the available sample it was postulated using this time-point would be more highly associated with outcome. This was investigated in Table 4-2 below.

Table 4-2: Ordinal analysis result assessing the predictive ability of labelled baseline on mRS at 90 days adjusted for age and treatment with thrombolysis

Baseline restriction of data	Correlation (R^2) Baseline measure with mRS	OR for NIH measure adjusted for age and rtPA	Upper 95% CI for OR	Lower 95% CI for OR
<6h NIH 4-20	0.44(0.19)	1.23	1.22	1.25
24h NIH 4-20	0.58(0.34)	1.31	1.29	1.33
<6h NIH 7-20	0.40(0.16)	1.21	1.20	1.24
24h NIH 7-20	0.48(0.23)	1.27	1.25	1.30

In Table 4-2 it is clear that for both restrictions, 24h NIHSS is more highly correlated with outcome than the <6h hyperacute measurement. This suggests that 24h NIHSS is a more

predictive measurement to use for trial admission provided the treatment permits a later time window. When restricted within 4-20, 24h NIHSS correlated more closely than <6h with 90 day mRS. With r^2 values of 0.34 and 0.19 respectively, as shown in Table 4-2, 24h NIHSS explains 15% more of the absolute variation in 90 day mRS than <6h NIHSS. Narrowing the restriction of NIHSS reduces this difference to 7%.

The odds ratios given in this table suggest that 24h NIHSS is a more powerful predictor for mRS at day 90 for both restrictions. Each of the odds ratios reported represents the odds of tending toward worse outcome measured by mRS for every one point increase in NIHSS. These odds ratios suggest that 24h NIHSS restricted 4-20 is the most predictive of outcome at 90 days.

Looking at Figure 4-4 and Figure 4-5 it is clear that selection of patients based on the 24h measurement provides a subgroup within which only a small proportion of patients are asymptomatic at 90 days. This is presumably because patients with early resolution over the period between 6 and 24 hours have been censored.

The distributions of day 90 mRS for each time point are given in Figure 4-4 and Figure 4-5. These are shown for the restrictions 4-20 and 7-20 respectively. It is clear that selection of patients based on the 24h measurement provides a subgroup of patients within which there is only a small proportion who are asymptomatic at 90 days, presumably because patients with early resolution over the period 6-24 hours have been censored.

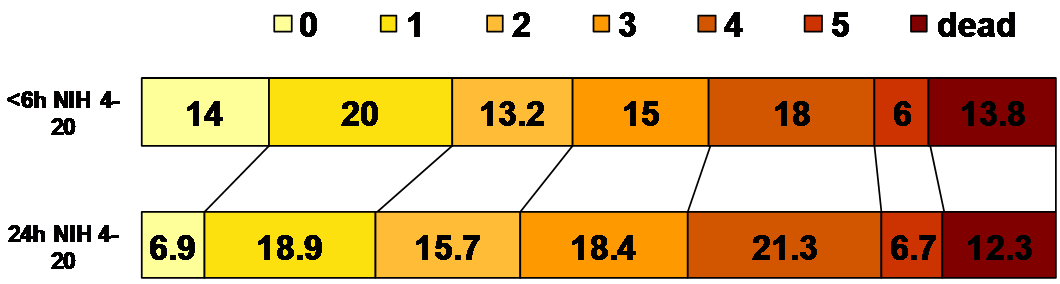

Figure 4-4: Distribution of mRS day 90 for 4-20 NIHSS restriction <6h and 24h NIHSS measurements

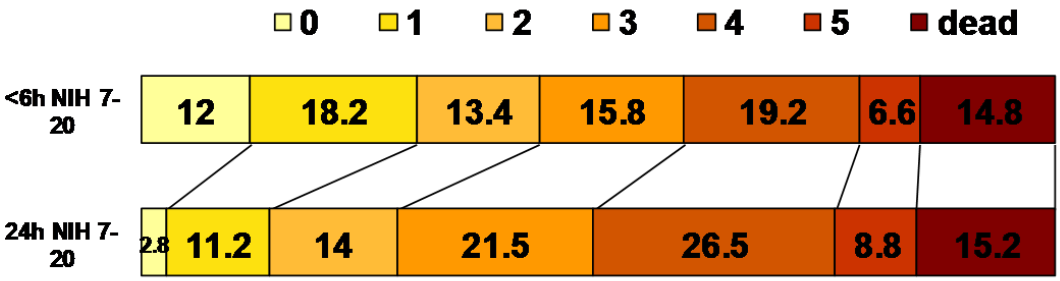

Figure 4-5: Distribution of mRS day 90 for 7-20 NIHSS restriction <6h and 24h NIHSS measurements

The data described above did not exclude those treated with thrombolysis. Results separated by treatment group are shown in Table 4-3 below.

Table 4-3: Ordinal analysis results assessing the predictive ability of labelled baseline on mRS at 90 days adjusted for age. Thrombolysed and non-thrombolysed analysed separately

Baseline restriction of data	Correlation (R^2) Baseline measure with mRS	OR for NIH measure adjusted for age	Upper 95% CI for OR	Lower 95% CI for OR
<6h NIH 4-20 Non_Throm	0.48(0.23)	1.25	1.23	1.27
<6h NIH 4-20 Throm	0.43(0.18)	1.21	1.19	1.24
24h NIH 4-20 Non_Throm	0.60(0.36)	1.31	1.29	1.34
24h NIH 4-20 Throm	0.57(0.32)	1.30	1.27	1.33
<6h NIH 7-20 Non_Throm	0.44(0.19)	1.23	1.21	1.26
<6h NIH 7-20 Throm	0.40(0.16)	1.21	1.19	1.23
24h NIH 7-20 Non_Throm	0.50(0.25)	1.29	1.26	1.32
24h NIH 7-20 Throm	0.47(0.22)	1.28	1.24	1.06

Comparing the R^2 values and odds ratios given in Table 4-3 the NIHSS measure taken in the hyperacute period is less associated with outcome at 90 days in thrombolysed subjects compared to non-thrombolysed. However when the time window is extended to 24h the relationship with outcome matches for thrombolysed and non-thrombolysed patients.

4.5.2. Simulations

4.5.2.1. Non-Thrombolysed only

Simulations to assess necessary sample size determined that given a restriction of 4-20 a 43% reduction was possible if selection was delayed until 24h (Table 4-4).

Table 4-4: Simulation results 80% power

Simulations to obtain 80% power based on adjusted analysis non thromblysed patients only *Adjusted for reported baseline and age*							
Baseline restriction (Inclusion criteria)	Trial size (assuming equal groups)	Power (%)	Adjusted OR for treatment effect	Lower 95% CI for OR	Upper 95% CI for OR	Increase mRS 0-2 (%)	Decrease mRS 3-6 (%)
<6h NIHSS 4-20	1280	80	1.319	1.311	1.327	6.1	5.1
24h NIHSS 4-20	730	80	1.462	1.451	1.475	5.7	5.6
<6h NIHSS 7-20	1200	80	1.344	1.335	1.353	5.2	5.3
24h NIHSS 7-20	600	80	1.528	1.513	1.542	7.0	7.0

Based on generating the same treatment effect size: the sample reduces from 1280 (640 per group) with selection <6h to 730 (365 per group) with selection at 24h, for a 5% absolute difference in proportion achieving mRS 0-2 and 80% power, illustrated in Figure 4-6 and Figure 4-7.

Given a restriction of 7-20 the reduction is increased to 50%. The simulated treatment effects are illustrated in Figure 4-8 and Figure 4-9.

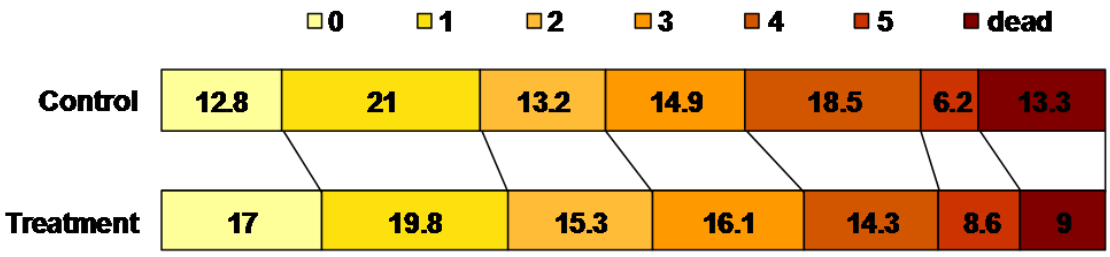

Figure 4-6: Distribution of mRS day 90 for simulations using <6h NIHSS 4-20

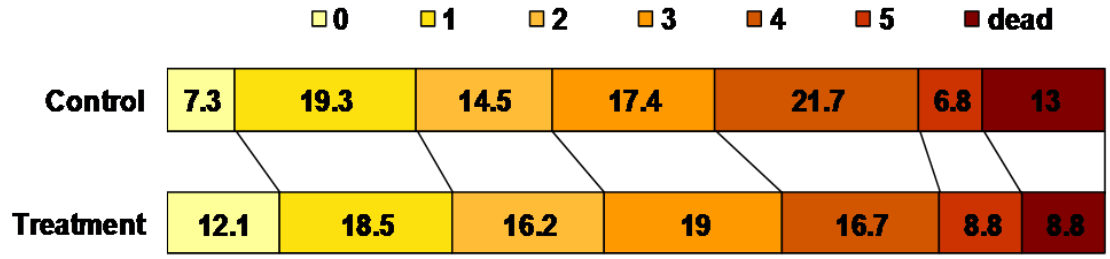

Figure 4-7: Distribution of mRS day 90 for simulations using 24 h NIHSS 4-20

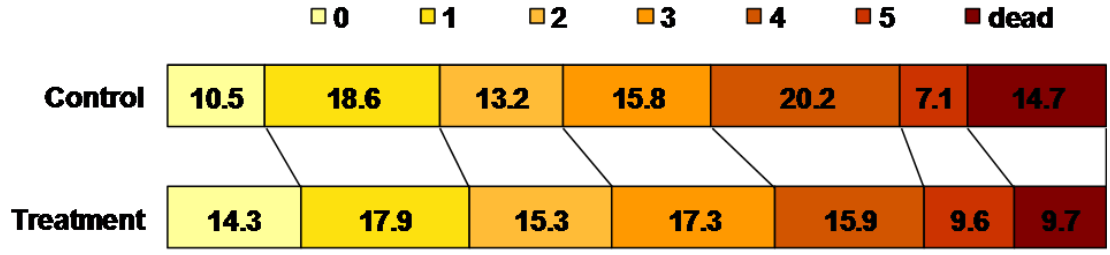

Figure 4-8: Distribution of mRS day 90 for simulations using <6h NIHSS 7-20

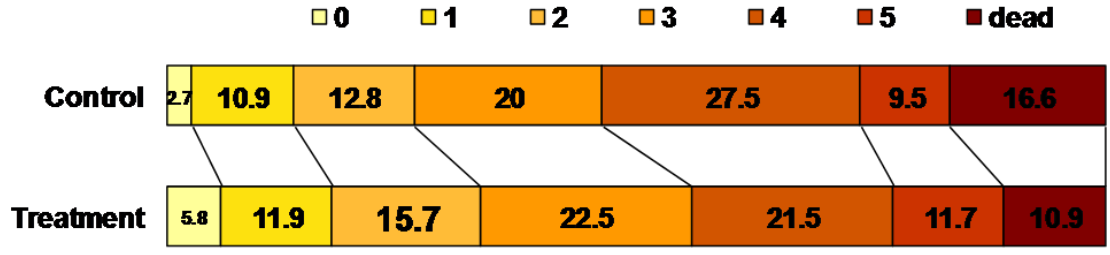

Figure 4-9: Distribution of mRS day 90 for simulations using 24 h NIHSS 7-20

Table 4-5 shows the results to obtain 90% power; these follow the same trend as above.

Table 4-5: Simulation results 90% power

Simulations to obtain 90% power based on adjusted analysis non thromblysed patients only *Adjusted for reported baseline and age*							
Baseline restriction (Inclusion criteria)	Trial size (assuming equal groups)	Power (%)	Adjusted OR for treatment effect	Lower 95% CI for OR	Upper 95% CI for OR	Increase mRS 0-2 (%)	Decrease mRS 3-6 (%)
<6h NIHSS 4-20	1650	90	1.332	1.325	1.339	6.1	5.1
24h NIHSS 4-20	930	90	1.460	1.449	1.470	5.7	5.6
<6h NIHSS 7-20	1600	90	1.340	1.333	1.347	5.2	5.3
24h NIHSS 7-20	820	90	1.521	1.509	1.533	7.0	7.0

If using NIHSS 4-20 as inclusion criterion, approximately 44% fewer subjects would be needed to obtain 90% power if the baseline measure was collected at 24h NIHSS instead of <6h. If using NIHSS 7-20 as inclusion criterion, this reduction increases to 49%.

4.5.2.2. Thrombolysed only

Among thrombolysed patients, the advantage in delayed selection was more apparent. Selection at <6h required a sample of 1720 (860 per group) whereas selection based on 24h NIHSS required 660 (330 per group). Results are given in Table 4-6 below.

Table 4-6: Thrombolysed only simulation results to obtain 80% power

Simulations to obtain 80% power based on adjusted analysis thrombolysed patients only *Adjusted for reported baseline and age*							
Baseline restriction (Inclusion criteria)	Trial size (assuming equal groups)	Power (%)	Adjusted OR for treatment effect	Lower 95% CI for OR	Upper 95% CI for OR	Increase mRS 0-2 (%)	Decrease mRS 3-6 (%)
<6h NIHSS 4-20	1720	80	1.271	1.264	1.278	4.8	4.9
24h NIHSS 4-20	660	80	1.493	1.480	1.506	6.7	6.8
<6h NIHSS 7-20	1500	80	1.291	1.284	1.298	5.2	5.2
24h NIHSS 7-20	600	80	1.538	1.524	1.551	8.1	7.8

4.6. Discussion

NIHSS recorded at 24h has a stronger relationship than hyperacute recordings with outcome at 90 days. The later baseline assessment explains 15% more of the absolute variability in outcome (an 80% relative advantage) among patients with baseline NIHSS scores of 4-20. This is supported by the adjusted odds ratios for initial NIHSS against final outcome. In contrast to earlier selection, 'enrolment' at 24h on the basis of an NIHSS range 4-20 delivers a population that includes few patients who become asymptomatic or die by 90 days: a higher proportion remain informative for trial purposes.

Our simulations suggest that if delayed selection can be supported on biological grounds for the treatment modality under test, then sample size reductions of 43% may be viable. Furthermore, the data suggest that the main treatment effect of thrombolysis may largely occur within the first few hours after treatment and the required sample may be further reduced if thrombolysed cases form part of this population. Indeed, exclusion of thrombolysed patients would be counterproductive to such a trial, both in terms of the

lost sample and longer enrolment period and because these patients are at least as informative as the non-thrombolysed group.

Overall, the analysis presented suggests that trial selection and analysis based on a delayed measurement of NIHSS selects a population in whom untreated outcome is more predictable and may allow trial samples to be reduced by 35-40%. We found statistical value in extending the window towards 24 hours.

Extending the time window for patient selection provides a measurement which has a stronger more predictive relationship with outcome. This subsequently allows a larger population to benefit from treatment. Such extension of the time window must be balanced against any anticipated decay in biological effect of the treatment with increasing delay from stroke onset. However, with recent interest in neurorestorative treatments, as discussed by Steven Cramer[288], any strategy that can limit the cost of proof of concept trials is desirable. Public access data show two such trials currently in phase II study; ISIS[289] and extended release Niacin (Niaspan©)[290].

Our simulation approach derives strength from the large number of samples taken (10,000) and from the large, independent dataset used as source. As the treatment effect here was simulated, we cannot make definite conclusions about the sample size, just recommendations. Ideally this would be tested on a population in whom there is a known treatment given after 24h with a comparative baseline population; however this appears impossible to validate due to the lack of applicable data available from current stroke trials. Since the adjusted and unadjusted analyses predicted comparable sample sizes, it appears that the statistical power advantage derives mainly from improved selection rather than improved covariate adjustment.

While trial selection and analysis based on a delayed measurement of NIHSS may allow trial samples to be reduced 35-40%, such an extension of the time window must be balanced against any anticipated decay in biological effect of the treatment with increasing delay from stroke onset.

Chapter 5
Selection for Treatment Based on a Prognostic Score

5.1. Background

Treatment of patients with acute cerebral ischemia using intravenous thrombolysis is safe and effective if initiated within 3h[96] or 4.5h[98] of stroke onset. There is evidence from RCT and pooled analysis of individual patient data to suggest that the conditions of use may be conservative and that additional patients treated within 6h may benefit[171]. However, there is also evidence from the pooled data analysis to suggest that other patients may be harmed by initiation of treatment within OTT range 4.5-6h[171]. Absolute mortality is 4.8% higher with thrombolysis treatment than placebo among patients treated at OTT within 4.5-6 h (p=0.05), and favourable outcome is 1.8% better (P=Non-significant). Thus, at these later OTT, it is unknown whether only a few patients are harmed and few benefit or whether a much larger group of patients benefit but harm offsets this benefit in a commensurately greater number.

There is considerable interest in exploring enhanced methods of patient selection, in order to identify subgroups of patients who may derive benefit at later OTT without suffering harm. Imaging approaches using mismatch of perfusion and diffusion weighted magnetic resonance scans, or CT perfusion scans, are under study. So far, these have not

led to changes in routine clinical practice though trials are still underway or in planning stages.

An alternative approach that has not been adequately explored, but which extends the principles underlying marketing authorisation for intravenous thrombolysis use in stroke in Europe, is to select patients for treatment based on their inherent prognosis. Thus, since we are aware that symptomatic intracerebral haemorrhage (SICH) carries with it an elevated risk of death or severe disability, and since we know that exposure to intravenous thrombolysis enhances the odds of SICH by around 4-fold, we aim to avoid treating patients who have a high absolute risk of SICH *without thrombolysis.* We do so because a four-fold increase in odds of a common event leads to a much higher absolute increase in risk than a fourfold increase in odds of an uncommon event. For example, if the background rate of SICH is 1% and the odds ratio for SICH with thrombolysis treatment is 4, then the absolute SICH rate may be 4%, an absolute increase in risk of 3%. Alternatively, if the background rate of SICH were 2% then an odds ratio of 4 would elevate the absolute risk under thrombolysis to 7%, which is an absolute increase of 5%.

These remain relatively small risks, but once we start dealing with common events such as death or disability, these influences of untreated prognosis can have a more profound effect on net treatment outcomes. Thus, a population with an expected untreated mortality of 20%, exposed to a drug that increases odds of death by 50% (odds ratio 1.5), is likely to show mortality of 27% with treatment (an absolute excess of 7%). If we can separate that population into two groups, one with expected mortality of 10% and one with expected mortality of 30%, then the absolute excess of deaths in these two groups should be under 5% in the former and 9% in the latter. Couple that with matching effects

on independence, and we may be able to demonstrate a net beneficial effect that is not outweighed by the absolute excess of harm.

We hypothesised that we could generate patient selection criteria for thrombolysis using a prognostic score based on simple clinical measures, developing optimal selection criteria using cohorts from an observational dataset of treated and untreated patients, and that we could validate our criteria within a separate existing dataset of pooled RCT data.

Due to the absence of available data on patients treated with thrombolysis after 4.5h, it was necessary to disregard OTT for the development and initial testing of our strategy. In light of the assumptions we had to invoke in generating our selection criteria, we planned prospective validation among patients treated in the 4.5-6h time window from the RCT pooled dataset. Our objective was to demonstrate a net benefit from treatment after offsetting any harm, with intravenous thrombolysis initiated 4.5-6h from stroke onset using simple clinical measures.

5.2. Methods

5.2.1. Data source and patients

For the development phase we gathered demographics, clinical data and functional outcome measures from neuroprotection trials in ischaemic stroke conducted in the period 1998 to 2007. We obtained our data, anonymised in relation to patients and trials, from the Virtual International Stroke Trials Archive (VISTA)[204]. We excluded trials that had tested effects of thrombolysis or of any drug now known to influence outcome after stroke. We retained patients who received intravenous thrombolysis as standard care, at the time of analysis information on the time of administration of thrombolysis was

unavailable but within the data was known to be <3h. Trials included within VISTA stipulated only basic imaging requirements for patient eligibility and did not routinely record detailed imaging parameters. These patients formed our "VISTA thrombolysis group", and were complemented by VISTA patients who were managed without thrombolysis, "VISTA controls". Finally, we excluded patients who lacked relevant baseline and outcome information: enrolment National Institutes of Health Stroke Scale score (NIHSS), age, modified Rankin score (mRS) day 90 and NIHSS day 90. Death was recorded as mRS grade 6.

For the validation phase, we used the pooled individual patient data from the published randomised trials of thrombolysis[98 166-171 291]. These data have been described previously[171] and trials included patients based on a defined onset time of stroke <6h and a CT scan to exclude haemorrhage.

It is important when performing such analyses that separate data are used for the development stage and the validation stage, first to avoid data mining and second to ensure the methods are applicable externally.

5.2.2. Outcome measures

Our primary outcome measure was mRS, analysed as an ordinal scale, with dichotomisation at mRS 0-1 v 2-6, and mortality (0-5 vs. 6) as secondary endpoints. Symptomatic haemorrhage, defined as parenchymal haemorrhage type 2 (PH2) was also a secondary outcome[166]. Due to lack of data on parenchymal haemorrhage within VISTA, PH2 was only used as a secondary outcome within the validation analysis.

5.2.3. Protocol

We conducted our study according to a pre-specified analysis plan shared with the VISTA steering committee and RCT investigators. We completed the development phase and declared our results including prognostic thresholds before accessing the RCT data for validation. As we were working with existing anonymised data, we did not require ethical review under institutional rules.

5.2.4. Exploratory analysis

Rather than developing a new prognostic score we chose to use a score which had been previously published and validated on different datasets, ensuring a robust and reliable analysis. Weimar et al[143] had developed a score to describe probability of functional independence after ischaemic stroke, using age and baseline NIHSS score. This prognostic score has been validated by Köenig et al on data from VISTA[141] and can be summarised in the formula Score=145-0.46*age-2.5*NIHSS. The authors limited the score to these two variables after rigorous analysis showing no other assessed factor to be an independent predictor of outcome alongside age and baseline NIHSS. Trials included within VISTA stipulated only basic imaging requirements for patient eligibility and did not routinely record detailed imaging parameters: we could not incorporate imaging data to the prognostic score. We chose this validated score for its simplicity, practicality and relevance to our data.

We ranked patients in the VISTA thrombolysis versus control dataset according to prognostic score, and generated ROC curves for benefit on each cut-point of mRS by selecting cut-offs across this range. After exploratory work confirming expected outcomes according to case mix distributions and treatment effects in patients treated

with thrombolysis within 3 hours OTT versus control, we estimated likely effects in a population treated between 4.5 and 6.0 hours OTT as follows.

5.2.5. Development of prognostic score thresholds

We began with the known distribution of mRS outcomes for a thrombolysed versus control population treated 4.5-6h after stroke onset as shown in Figure 5-1, taken from the recent published analysis of pooled RCT data[171]. As onset time of thrombolysis was not recorded in VISTA though required to be within 3h of stroke onset, we needed to generate a development population which would be comparable to the RCT data for later-treated patients[171]. We populated a training dataset from VISTA to match the known distribution of outcomes in the 4.5-6h treated group: on the basis of their 90-day mRS, 500 VISTA control patients and 500 VISTA thrombolysis patients were selected at random to supply the correct number of patients within each outcome category to match those of the pooled RCT data[171]. We did not permit prognostic factors of these patients to influence their selection. This generated a 'VISTA trial population' of 1000 patients who had outcomes that almost exactly matched the known outcomes of late-treated RCT patients, differing appropriately between thrombolysis and control groups. This sample was used for all further first stage analyses.

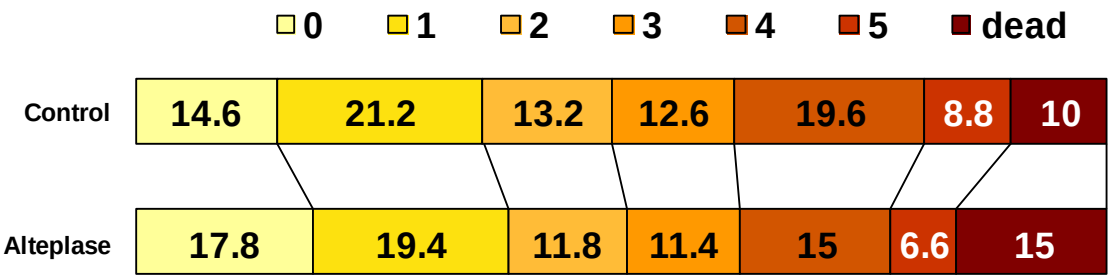

Figure 5-1: Distribution of mRS scores from pooling analysis by Lees et al[171] for patients treated above 4.5h.

We ranked patients in the VISTA trial population according to their prognostic score. We then excluded patients with the worst prognosis and at the other end of the range also excluded patients with the best prognosis, effectively applying a prognostic score "window", within which we retained the patients in our treatment sample. We hoped to maximise the size of this retained sample, since statistical power is partly driven by sample size; but also to maximise the treatment effect size through exclusion of "minimal responders", since statistical power is also driven by the extent of the effect. We assumed that the optimal window of prognostic scores for identifying our study sample would represent a compromise between those that delivered a large sample with modest average treatment effect versus criteria delivering a small sample deriving a large treatment effect.

We used an iterative approach to select the optimal prognostic score window, first to identify a patient sample that had a treatment benefit identifiable at a statistical threshold of $p<0.05$ and then a more restricted dataset with benefit detectable at $p<0.01$.

Using logistic regression with the proportional odds model, adjusting for age and baseline NIHSS, we generated odds ratios for more favourable mRS between thrombolysed and control groups for each selected (better predicted outcome) or excluded (poorer

outcome) population, with 95% confidence intervals. Significance was assessed by Cochran Mantel Haenszel (CMH) test.

Focussing on limits that appeared likely to select a group of patients with moderate prognosis in whom the treatment effect of thrombolysis may be optimal, we then sought to optimise the sample. Our strategy was to keep the selection criteria as wide as possible, since this would maximise the population benefit and the sample size, thereby delivering both statistical power and clinical relevance; but we also sought to ensure that our final sample was likely to show a statistically significant net benefit, and for this we had to limit inclusion to patients in whom odds of net benefit would be larger. We recognised that our training dataset would likely differ from the RCT population in certain respects and that any choice of prognostic score limits may be only preliminary. We therefore base our final choice of prognostic score limits on those that maximise the sample whilst still delivering a significant net treatment benefit at $p<0.01$, ie the 99% confidence interval for the odds ratio estimated by ordinal regression does not encompass 1.0.

5.2.6. Validation procedure

We supplied the chosen prognostic score thresholds to an independent statistician who undertook CMH test, ordinal logistic regression and dichotomised analysis of the individual patient data from the pooled RCT for patients treated 4.5-6h from stroke onset.

Results are expressed as odds ratios and 95% CI for more favourable mRS distribution under thrombolysis versus control, with p values from CMH test. Secondary dichotomised outcomes (mRS 0-1, mortality and PH2 rate) were analysed as previously described [171].

We ran exploratory analyses among 3-4.5h and 0-3h (0-90' and 91-180' combined) groups in a similar manner, recognising that here the small sample sizes undermined power.

5.3. Results VISTA stage

5.3.1. Prognostic score validity

Table 5-1 to Table 5-4 display the cross tabulations of prognostic score against mRS category for various VISTA populations, before selection according to the prognostic score. Figure 1 shows the ROC curves for identifying good prognosis according to each of the possible dichotomisations of mRS (0 versus 1-6 through 0-5 versus 6).

Table 5-1: Descriptive statistics of prognostic score applied to VISTA data for entire dataset, control and thrombolysis

Group	N	Mean	Median	Minimum	Maximum	Lower 95% CL for Mean	Upper 95% CL for Mean
All data	8692	79.97	82.00	14.00	122.00	79.62	80.31
Control	6166	80.53	83.00	14.00	122.00	80.11	80.94
Thrombolysis	2526	78.59	79.00	20.00	118.00	78.00	79.19

Table 5-2:Descriptive statistics of prognostic score by mRS grade for the full dataset

mRS grade	N	Mean	Median	Lower 95% CL for Mean	Upper 95% CL for Mean
0	979	92.16	94.00	91.41	92.92
1	1475	90.37	93.00	89.75	90.99
2	1106	87.11	89.00	86.35	87.87
3	1246	80.90	82.00	80.15	81.66
4	1656	75.18	76.00	74.54	75.81
5	687	66.73	66.00	65.71	67.75
6	1543	67.44	66.00	66.70	68.18

Table 5-3: Descriptive statistics of prognostic score by mRS grade for patients not treated with thrombolysis

mRS grade	N	Mean	Median	Lower 95% CL for Mean	Upper 95% CL for Mean
0	664	93.64	95.00	92.76	94.51
1	1039	91.74	93.00	91.02	92.46
2	735	88.58	90.00	87.68	89.48
3	855	82.18	84.00	81.28	83.08
4	1191	76.06	77.00	75.30	76.83
5	514	66.48	66.00	65.28	67.68
6	1168	67.56	67.00	66.69	68.42

Table 5-4: Descriptive statistics of prognostic score by mRS grade for patients treated with thrombolysis

mRS grade	N	Mean	Median	Lower 95% CL for Mean	Upper 95% CL for Mean
0	315	89.06	91.00	87.65	90.48
1	436	87.11	88.00	85.94	88.28
2	371	84.19	86.00	82.83	85.55
3	391	78.11	79.00	76.76	79.47
4	465	72.90	73.00	71.79	74.01
5	173	67.45	67.00	65.50	69.40
6	375	67.06	66.00	65.64	68.48

These simple descriptive statistics show an association between prognostic score and outcome, with thrombolysis treatment influencing this association. Within both treated and untreated groups, the mRS at day 90 increases as the prognostic score at baseline decreases. In Table 5-3 and Table 5-4 it is evident that patients treated with thrombolysis achieve lower mRS scores despite lower prognostic scores at baseline, highlighting the

treatment effect. For example in the placebo group the average prognostic score among patients with a 90 day mRS of 2 is 88.6 whereas the average prognostic score among treated patients attaining a 90 day mRS of 2 is 85.6.

The ROC curves in Figure 5-2 demonstrate that the prognostic score has a comparable association with each mRS grade, underlining the proportionality of the relationship in the overall population.

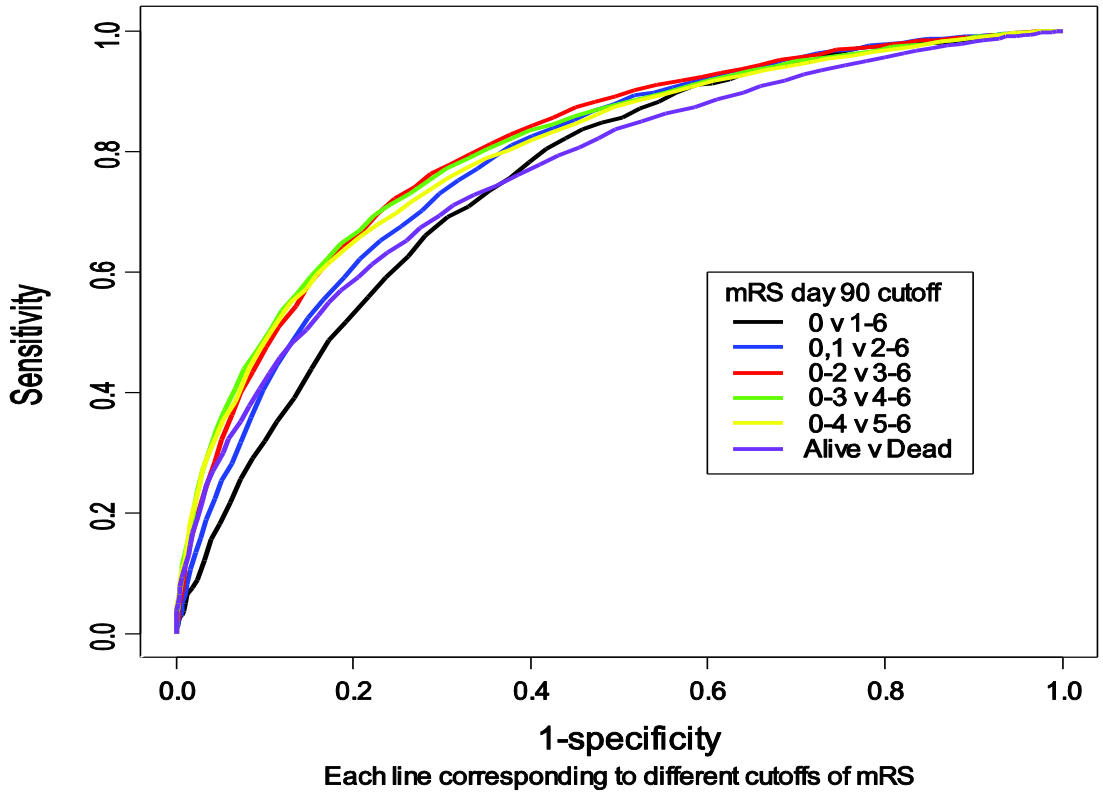

Figure 5-2: ROC curves showing the sensitivity and specificity corresponding to prognostic score as a test for outcome measured by dichotomised mRS day 90

5.3.2. Choice of selection criteria based on prognostic score

From the VISTA dataset we sampled 1000 patients for our trial population to match the outcome distributions in the RCT data[171]. Their baseline characteristics were, control: Age 69 ± 13, baseline NIHSS 10 IQR 7, 16; Thrombolysis: Age 68 ± 13, baseline NIHSS 13 IQR 9, 18.

Table 5-5 shows the most promising boundaries of prognostic score for selection of a population that appeared to benefit from thrombolysis. This was based on a significance level of 0.01, with 99% confidence intervals calculated. Figure 5-3 shows the distribution of day 90 mRS for the VISTA sample contained within this group, these boundaries show net benefit for treatment with thrombolysis (OR=1.41, p=0.008) but still retains a large proportion of the original sample.

Figure 5-4 and Figure 5-5 illustrate the distribution of mRS for those above and below the prognostic interval respectively. These figures illustrate the ability of the prognostic boundaries to exclude those who are predominantly asymptomatic at 90 days and those who are severely symptomatic at 90 days.

Table 5-5: Finding boundaries of prognostic score above and below which thrombolysis should not be given, with 99% confidence limits and significance level of 0.01

Lower Prog Cut	Upper Prog cut	Total in group	OR for more favourable mRS	Lower 99% CI	Upper 99% CI	CMH	N given Thrombolysis	N Controls
53	98	801	1.388	1.00	1.926	0.0219	421	380
54	98	790	1.383	0.994	1.925	0.0226	416	374
54	97	774	1.374	0.985	1.919	0.0413	409	365
55	97	767	1.382	0.988	1.933	0.0413	405	362
55	96	746	1.429	1.017	2.008	0.0143	395	351
56	96	738	1.427	1.014	2.008	0.0100	390	348
56	95	714	1.412	1.000	1.998	0.0080	380	334

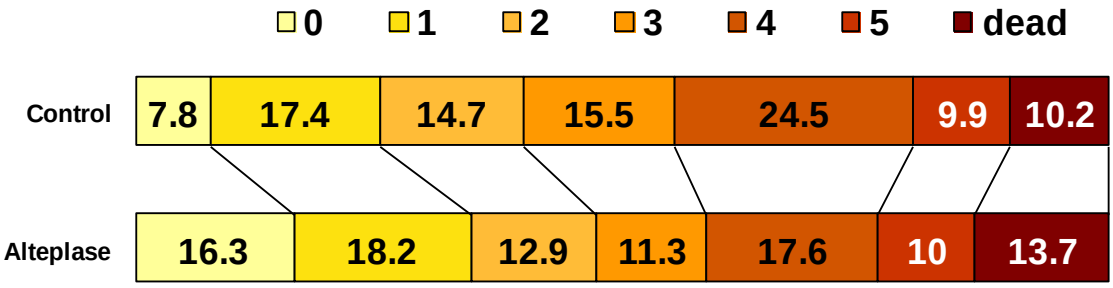

Figure 5-3: Distribution of mRS for those between prognostic cut-off of 56 and 95 N=714 control=334 Thrombolysis=380, OR=1.412, 99% CI = (1.000, 1.998)

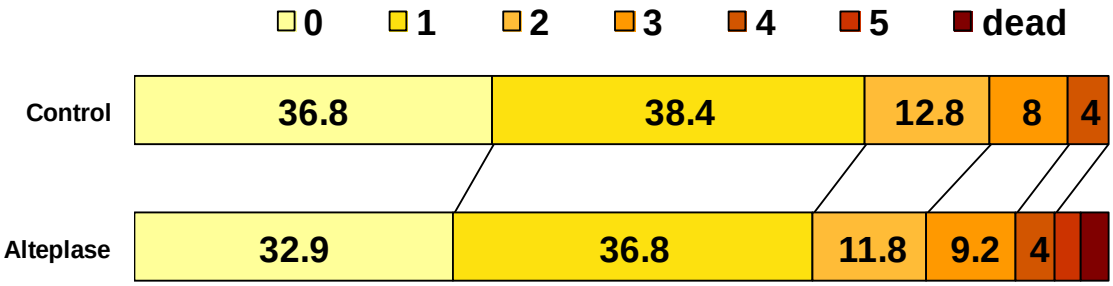

Figure 5-4: Distribution of mRS for those above cut of 95 N=201 control=125 Thrombolysis=76, OR=0.961, 99% CI = (0.470, 1.964)

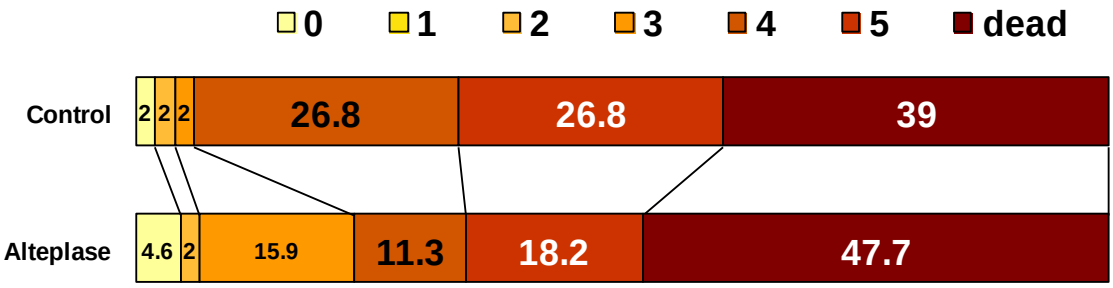

Figure 5-5: Distribution of mRS for those below cut of 56 N=85 control=41 Thrombolysis=44, OR=0.954, 99% CI = (0.334, 2.726)

Odds of favourable outcome were investigated for secondary outcomes and are given in Table 5-6. Significance was assessed at a level of 0.01 with 99% CI's calculated. Dichotomised mRS at both 0-1 and 0-2 show increased odds of favourable outcome given treatment with thrombolysis.

Table 5-6: Binary regressions for different mRS dichotomisations. Data between chosen boundaries of 56 and 95, looking at 99% CI

mRS cut	OR for rtPA	Lower 99% CI	Upper 99% CI	CMH significance	% excess or deficit(-) of thrombolysis group
0-1 vs. 2-6	1.977	1.238	3.155	0.0036	9.33
0-2 vs. 3-6	1.743	1.129	2.691	0.0262	7.55
Mortality	0.707	0.374	1.336	0.2209	-3.5

This was then repeated to find wider prognostic boundaries showing significance at the 5% level. These boundaries were found to be prognostic scores of 47 to 104. The distribution of the mRS for those contained within these boundaries is given in

Figure 5-6. This interval is more inclusive than the first calculated as it contains around 90% of the sampled population. These boundaries showed a net benefit of thrombolysis treatment, giving odds of favourable outcome OR=1.264 with 95% CI (1.003, 1.593), CMH p-value=0.024. Figure 5-7 and Figure 5-8 show the distribution of mRS for those above and below the interval respectively. This shows a similar trend to the initial boundaries, excluding those who are asymptomatic and severely symptomatic at 90 days.

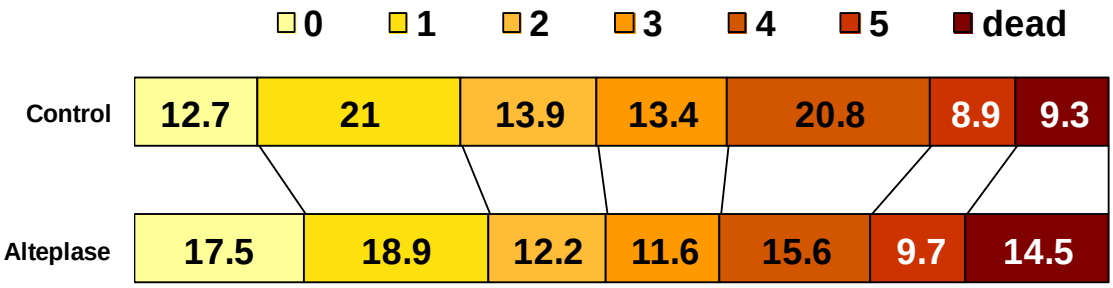

Figure 5-6: Distribution of mRS for those between prognostic cut-off of 47 and 104 N=937 control=462 Thrombolysis=475, OR=1.264, 95% CI = (1.003, 1.593)

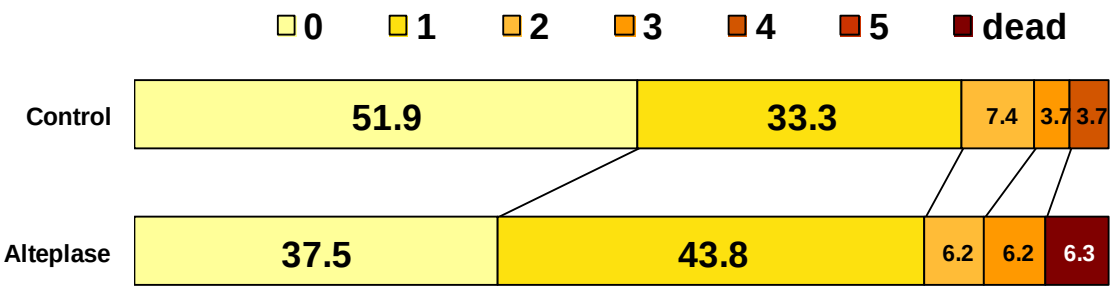

Figure 5-7: Distribution of mRS for those above cut of 104 N=43 control=27 Thrombolysis=16, OR=0.782, 95% CI = (0.190, 3.212)

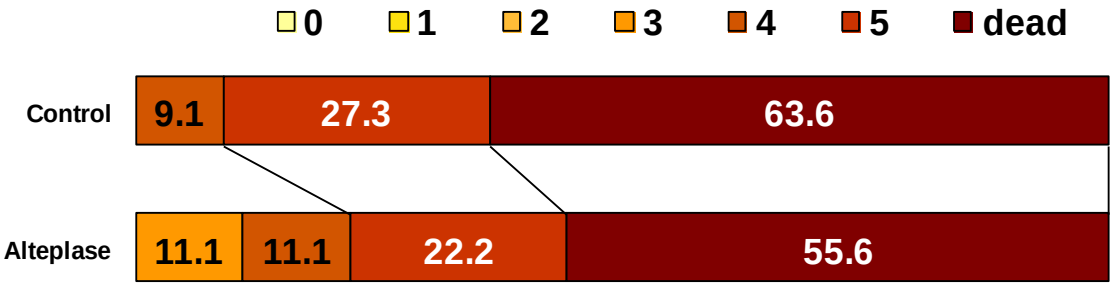

Figure 5-8: Distribution of mRS for those below cut of 47 N=20 control=11 Thrombolysis=9, OR=1.384, 95% CI=(0.199, 9.645)

Table 5-7: Binary regressions for each cut of mRS. Data between chosen boundaries of 47 and 104, looking at 95% CI

mRS cut	OR for rtPA	Lower 95% CI	Upper 95% CI	CMH significance	% excess or deficit(-) of thrombolysis group
0-1 vs. 2-6	1.671	1.223	2.284	0.007	2.65
0-2 vs. 3-6	1.536	1.133	2.083	0.054	1.01
0-5 vs. 6	0.643	0.415	0.995	0.136	-5.22

From this exploratory work we identified the prognostic score range of 56-95 inclusive as delivering a population who may derive treatment benefit, significant at $p<0.01$ and representing over 70% of the available patients. Wider limits of 47-104 inclusive retained approximately 90% of the population and appeared to offer benefit significant at $p=0.05$.

5.4. Conclusions from preliminary analysis

From preliminary analysis performed on the VISTA data, it would appear that adequate initial boundaries of prognostic score to apply to the pooling dataset would be 56 and 95. This encapsulates around 85% of the sample selected to match those outside the 4.5h time window in the pooling dataset. Secondary boundaries to apply to this dataset are 47 and 104.

However, as this analysis has been performed on data for which the time window for thrombolysis treatment has not been given and is known to be below 4.5 h when applying the prognostic scores to the real data these boundaries may narrow further.

5.5. Results of validation analysis

The pooled RCT data consisted of 3670 patients, of whom 1120 were treated between 4.5 and 6h with thrombolysis or placebo, patient demographics as previously described [171].

When applied to the 1120 patients in the pooled RCT 4.5-6h dataset, score limits of 56-95 retained 711 patients (64%) and gave OR for improved mRS distribution of 1.13, 95% CI 0.87-1.47, CMH p=0.89. The distribution of mRS scores for this population is given in Figure 5-9. More patients treated after 4.5h who fulfilled the score limits achieved mRS 0-1 (OR 1.44, 1.02-2.05, p=0.04) than in the overall population (OR 1.15, 0.88-1.51, p=0.30). However PH2 bleeds showed a trend to increase further, from OR 7.67, 2.99-19.7, p <0.0001 to OR 15.6, 3.7-65.8, p=0.0002. The OR for elevated mortality observed in the overall population was not limited by applying the selection score, the confidence interval simply widening: from 1.58 (1.07-2.33) p=0.02 the OR for mortality became 1.56, (1.01-2.40), p=0.04.

The wider prognostic boundaries of 47-104 gave ordinal OR 1.13 (0.90-1.41, CMH p=0.40) in 988 patients (88%) The distribution of mRS scores for this population is given in Figure 5-10.

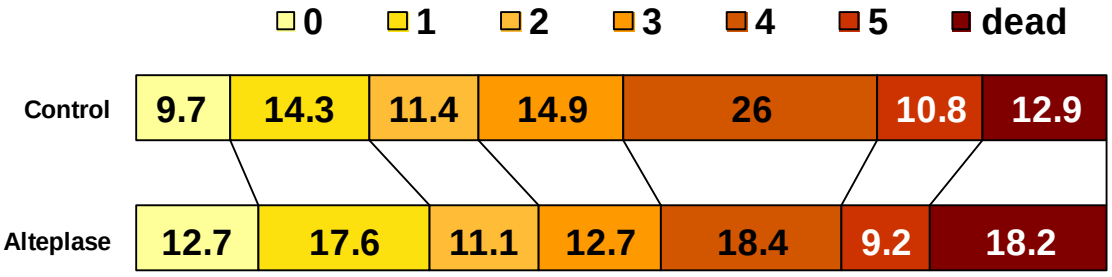

Figure 5-9: mRS distribution in pooled data with prognostic cut-points of 56-95 applied.

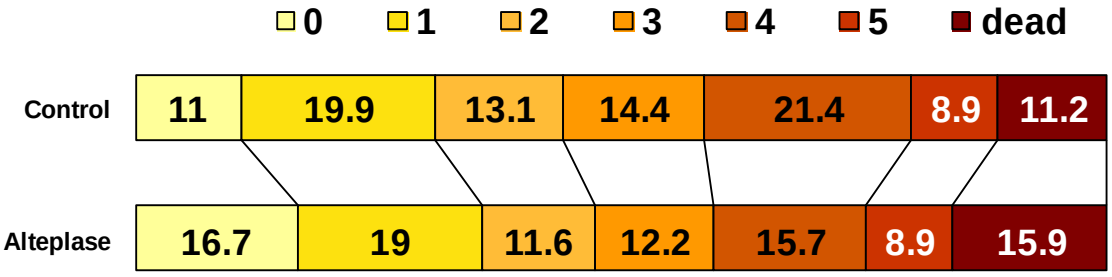

Figure 5-10: mRS distribution in pooled data with prognostic cut 47-104 applied.

When applied to the 1620 patients in the pooled RCT 3-4.5h dataset, among whom there is known benefit[171], score limits of 56-95 retained 1013 patients (63%) and gave OR for improved mRS distribution of only 1.05, 95% CI 0.84-1.30, CMH p=0.27. Odds of achieving mRS 0-1 decreased as the boundaries were applied and there was no advantageous effect on mortality or PH2 bleeds, shown in Table 5-8.

When applied to the 930 patients in the pooled RCT 0-3h dataset, score limits of 56-95 retained 624 patients (67%) and similarly offered no advantage in terms of any outcome as given in Table 5-8.

Table 5-8: Odds ratios, 95% CIs and p-values from logistic regressions. Investigating achieving mRS 0-1, mortality and the occurrence of PH2 bleeds, all analysis adjusted for baseline NIHSS and age on admission. Score I represents the initial prognostic cut points of 56-95 and score II represents the secondary cut points of 47-104.

Group (N)	OR for ordinal analysis (Proportional Odds) (95% CI)	P-value from CMH test	OR (95% CI) for achieving mRS 0-1	P-value mRS 0-1	OR (95% CI) mortality	P-value mortality	OR (95% CI) PH2 Bleeds	P-value PH2 bleeds
4.5-6h all (1118)	1.05(0.85-1.29)	0.39	1.15(0.88-1.51)	0.299	1.58(1.07-2.33)	0.02	7.67(2.99-19.66)	<0.0001
4.5-6h Score I (711)	1.13(0.87-1.47)	0.89	1.44(1.02-2.05)	0.04	1.56(1.01-2.40)	0.04	15.6(3.7-65.8)	0.0002
4.5-6h Score II (988)	1.13(0.90-1.41)	0.40	1.27(0.96,1.70)	0.097	1.54(1.04-2.28)	0.03	6.71(2.60-17.32)	<0.0001
3-4.5h all (1620)	1.15(0.97-1.37)	0.022	1.30(1.04-1.63)	0.024	1.18(0.84-1.66)	0.34	3.82(1.87-7.78)	0.0002
3-4.5h Score I (1013)	1.05(0.84-1.30)	0.27	1.23(0.92-1.65)	0.16	1.24(0.85-1.79)	0.26	3.52(1.64-7.55)	0.0012
3-4.5h Score II (1423)	1.12(0.93-1.35)	0.04	1.27(0.997,1.62)	0.053	1.22(0.86-1.72)	0.26	3.82(1.81-7.54)	0.0003
0-3h all (930)	1.35(1.07-1.70)	0.11	1.81(1.34-2.46)	0.0001	1.02(0.70-1.47)	0.93	7.37(2.19-14.84)	0.001
0-3h Score I (624)	1.23(0.93-1.62)	0.12	1.56(1.09-2.23)	0.016	1.05(0.68-1.64)	0.82	6.37(1.84-21.99)	0.0034
0-3h Score II (819)	1.37(1.07-1.75)	0.10	1.77(1.28,2.43)	0.0005	0.96(0.65-1.43)	0.85	6.42(1.87-22.05)	0.0032

5.6. Discussion

Selection of patients for treatment initiation between 4.5 and 6h based on simple clinical measures developed from an observational dataset failed to deliver a population in whom the thrombolysis effect would be both safe and effective.

We had postulated that by concentrating treatment among patients with low absolute risk of adverse outcomes, the adverse consequences of delayed treatment initiation may be limited sufficiently to uncover a net benefit. When we sought to validate this approach, we found that while favourable outcome by mRS 0-1 was improved, the risk of PH2 bleeds was also greatly exaggerated, and that there was no net benefit. This mirrors the effects that are observed in the unselected 4.5-6h dataset[171]. It is possible that our approach correctly identifies patients with neither a fixed deficit nor almost certain recovery, but that restoration of perfusion in these remaining patients simply carries substantial risk through bleeding. This undermines the use of "clinical judgement" to choose patients for delayed treatment unless the patient prefers to accept a higher risk of mortality to try for functional independence, because we are not able to improve the risk/benefit ratio for treatment within this time window by use of these simple selection criteria. Indeed, our approach also failed validation in the earlier time windows. While we recognise that imaging could be an instructive tool, our approach here does not inform the debate on use of more sophisticated imaging methods for patient selection. However, at this stage, perfusion-diffusion mismatch imaging has not developed sufficiently to be incorporated into routine practice for this purpose[292].

Our selection of prognostic score boundaries derives strength from the large, independent sample used to generate them, and from the clinically relevant size of the

selected population, around two thirds of patients who had been treated within this time window in the RCT. Weaknesses include the non-randomised nature of our VISTA treatment groups, the restriction of our prognostic score to only two variables, age and NIHSS, rather than including imaging parameters or other clinical variables, an absence of reliable data on PH2 bleeds among controls, and possibly most crucial, the absence of real data from treatment administered beyond 4.5h. While adding further variables such as blood glucose and blood pressure into the prognostic score could give a more precise answer these explain only a limited proportion of the variability and were previously discounted for the prognostic score by Weimar et al."

It is likely that most VISTA thrombolysis patients were treated within 3h of stroke onset. The natural history of NIHSS scores is for them to fall over the first hours after stroke onset. By using data from patients examined within 3h of stroke onset to generate a simulated population 4.5-6 h from stroke onset, without allowing for this average improvement, we will have slightly inflated our estimate of the baseline severity.

While our analysis based on ordinal outcomes failed to deliver a population in whom treatment >4.5h was safe and effective, analysis based on net benefit (mRS 0-1) showed significance. The analysis of trial data according to net benefit has proponents and opponents, the latter arguing that it may conceal useful treatment effects among subpopulations. Unless we can prospectively select patients for these sub-populations, the ordinal approach to interpretation may remain optimal as it better reflects the true outcome of clinical practice.

Chapter 6
Day 7 NIHSS is a Sensitive Outcome Measure for Exploratory Clinical Trials in Acute Stroke

6.1. Background

Acute stroke trials typically record outcomes after at least 90 days[293]. Prolonged follow-up increases the costs and duration of trials, and risks weakening conclusions due to patient attrition. To enhance efficiency of exploratory trials, there may be benefit from reliance on earlier assessment of outcomes, e.g. after 7 or 30 days. This may also minimise confounding due to unrelated adverse events. With established efficacy of recombinant tissue plasminogen activator (thrombolysis) on mRS at 90 days we have an opportunity to explore alternative approaches[171 291 294].

The NINDS thrombolysis trial group proposed 24h change in National Institutes of Health Stroke Scale (NIHSS) as endpoint for part 1 of their trial but abandoned this for part 2; they later concluded that early NIHSS measures may be more sensitive than 90 day measures[230]. This reversal of choice for the NINDS trials, the conflicting rank order of power in subsequent endpoint analysis work and a general acceptance that sustained

functional benefit should be demonstrated in pivotal trials may together account for the limited implementation of their suggestion.

Nevertheless, if earlier outcome measures are more sensitive to treatment effects than 90 day mRS, they should be implemented in exploratory trials. We investigated the sensitivity of 7 day NIHSS score and 30 day mRS to the established treatment effect of thrombolysis on 90day mRS, to inform their validity as endpoints in clinical trials in acute stroke.

While simulated trials based on registry data can generate useful hypotheses, non-randomised data are open to several sources of bias. Any conclusions should be validated on an external dataset in which treatment with thrombolysis has been randomised. If validated, application of an earlier endpoint such as 7 day NIHSS into exploratory stroke trials could potentially reduce cost and increase power. This approach may be of particular benefit with treatments that show efficacy rapidly; such as reperfusion strategies. We sought to validate our hypothesis using data from published randomised trials of thrombolysis in acute ischaemic stroke.[8]

6.2. Methods

6.2.1. VISTA Analysis

6.2.1.1. Data source and patients

Data were extracted from VISTA who met overarching selection criteria on data availability, namely: NIHSS recorded within 6h of stroke onset and at 24h; modified Rankin Scale score recorded at 90 days; age; exposure to iv thrombolysis (or not) as standard of care; no investigational drug administered that has a confirmed influence on

stroke outcome (ie source trials tested neuroprotectants, not thrombolysis); date of trial falls within period 1998-2008.

6.2.1.2. Statistical analysis

We undertook a controlled comparison of thrombolysis treated patients versus untreated controls from VISTA. The characteristics of this database have been described previously[237]. VISTA operating procedures preclude reanalysis of randomised controlled trials of thrombolysis for the thrombolysis effect. Instead we considered participants from trials of putative neuroprotectant agents. In these trials up to 50% of patients received thrombolysis as part of standard care alongside the trial drug. The decision to administer thrombolysis was not randomised, but comparison of thrombolysis treated patients with untreated controls offers an approximation of thrombolysis related treatment effects that have previously appeared to give a valid estimate. From VISTA we considered patients with complete data on receipt of thrombolysis, mRS at 30 and 90 days, and NIHSS score at 7 and 90 days. For patients known to have died during follow-up, a mRS score of 6, and NIHSS score of 42 were ascribed. Baseline patient characteristics were compared using Mann Whitney *U* tests for continuous variables and X^2 tests for categorical variables.

6.2.1.3. Simulation technique

Multiple re-sampling was performed to assess the sensitivity of each variation of the two outcome measures to the treatment effect of thrombolysis. From our VISTA dataset 10,000 random samples of patients were drawn at each of a series of sample sizes ranging from 1000 downwards in decrements of 10. Each sample consisted of equal numbers of thrombolysis treated and control patients. For every sample size, the percentage of trials yielding a statistically significant treatment effect was recorded. This percentage

approximates the statistical power of each outcome measure to detect thrombolysis treatment effects at that sample size.

In testing for the treatment effect of thrombolysis our primary analysis used ordinal logistic regression. This compares shifts across all categories of the mRS or NIHSS rather than arbitrarily privileging one health state as in dichotomised analysis[295]. For the purposes of this analysis NIHSS scores were grouped into the following categories: 0 (no measurable deficit), 1-4, 5-8, 9-12, 13-16, 17-20, 21-24, and ≥25[296].

The EMEA Points to Consider for stroke trial design advises that ordinal analysis may be supported by supplementary analysis of secondary outcome measures[297]. In this regard we undertook dichotomised analysis using binary logistic regression. The power of dichotomous endpoints depends on case-mix and the specific treatment effect studied[10]. To find the optimal dichotomised endpoint for each outcome measure a range of dichotomies were included in our simulations. For mRS outcome measures we used excellent (mRS 0-1 vs 2-6) and good (mRS 0-2 vs 3-6) functional outcomes; for NIHSS we used the dichotomies 0-1, 0-1 or improvement by 8 or more points from baseline, and 0-1 or improvement by 11 or more points from baseline. The most sensitive dichotomy for each outcome measure was used for comparisons.

In all analyses we adjusted for age and baseline NIHSS score. These are established as the most important predictors of stroke mortality and functional outcomes[298]. Sensitivity analyses performed on the total sample found adjustment for additional variables to have negligible influence on the results.

Simulations were performed using R version 2.12.1, all other analyses used PASW 18. A p-value <0.05 was considered significant.

6.2.2. Validation analysis

We included data from patients who had been enrolled and treated within 270 minutes of stroke onset, in any of 8 published randomised trials[171]. We compared thrombolysis treated patients against untreated controls using a multiple resampling approach known as bootstrapping. We repeatedly drew 1000 samples of unique patients, each time constraining the treated and untreated subsets (the simulated treatment groups) to be of equal size. A range of such sample sizes were tested. Each time we simulated a trial, we tested for a significant treatment effect based on outcome measures that included mRS day 90, 7 and 90 day NIHSS. The percentage of samples yielding significant results approximates the power of each endpoint at that sample size.

The validation study followed a predefined analysis plan and was conducted by an independent statistician. Missing data were handled using a last observation carried forward approach.

Our primary analysis included both ordinal and binary logistic regression. The ordinal approach compares shifts across all categories of the mRS or NIHSS. The binary approach arbitrarily privileges one health state (eg mRS 0-1 versus mRS 2-6) in a dichotomised analysis[295]. In all analyses we adjusted for age and baseline NIHSS score.

Simulations were performed in SAS 9.2. A p-value <0.05 was considered significant.

6.3. Results

6.3.1. VISTA Analysis

From VISTA we obtained data on 7886 patients, of who 4712 met the data requirements for inclusion. 1934 (41.0%) of our sample were treated with thrombolysis. The baseline

demographics of our sample are shown in Table 6-1. The groups were highly imbalanced at baseline: thrombolysis treated patients were younger and had more severe strokes.

Table 6-1: Baseline demographics from the VISTA data

	thrombolysis (n= 1934)	Control (n=2778)	P-value for difference
Age			
Mean (SD)	67.2 (12.8)	68.9 (12.2)	<0.0001
Median (IQR)	69.5 (18)	71 (17)	<0.0001
Male Sex	1094 (56.6%)	1485 (53.5%)	0.035
AF	440 (22.8%)	707 (25.4%)	0.034
Diabetes	357 (18.5%)	677 (24.4%)	<0.0001
Hypertension	1339 (69.2%)	2113 (76.1%)	<0.0001
CHF	150 (7.8%)	216 (7.9%)	0.98
IHD	432 (22.4%)	889 (32.0%)	<0.0001
MI	254 (13.1%)	322 (11.6%)	0.112
Prior Stroke	242 (12.5%)	623 (22.4%)	<0.0001
Baseline NIHSS			
Mean (SD)	13.6 (5.0)	11.7 (5.1)	<0.0001
Median (IQR)	13 (8)	10 (8)	<0.0001

The relationships between sample size and power for each of outcome measures, by ordinal and dichotomised analysis are shown in Figure 6-1 and Figure 6-2 respectively.

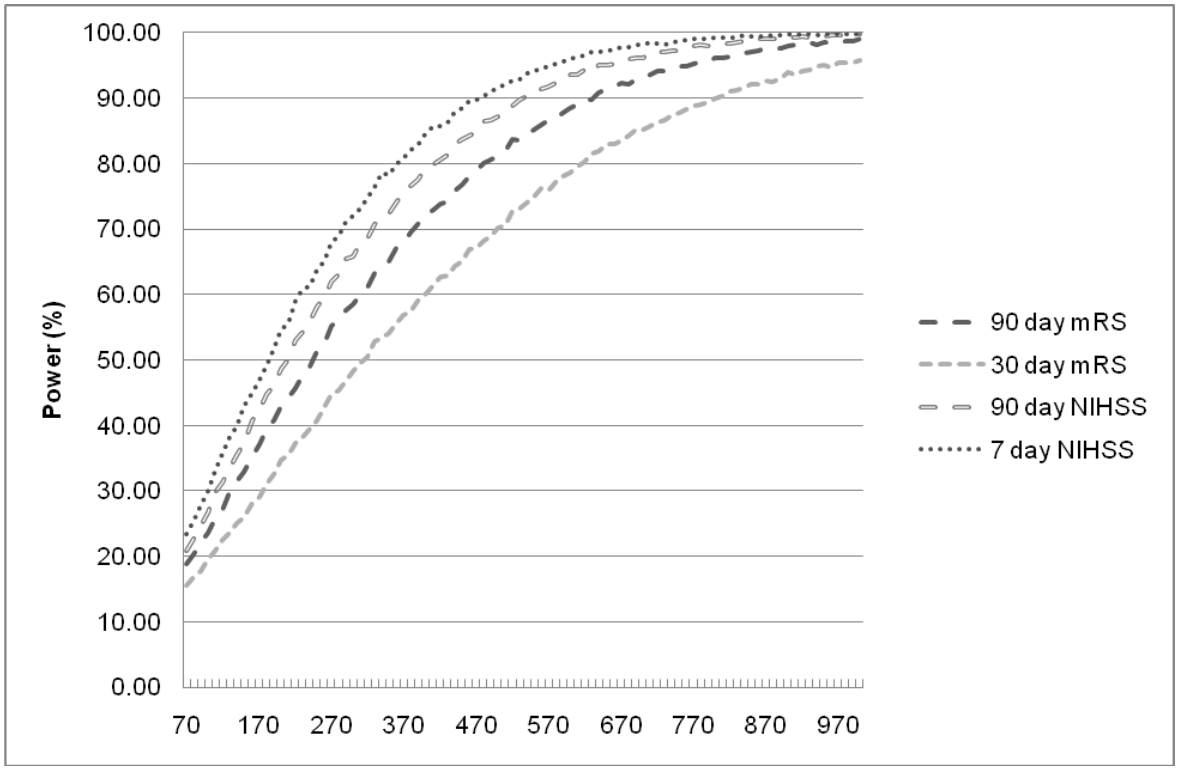

Figure 6-1: Line graph showing the relationship between sample size and statistical power for each of the 4 outcome measures using ordinal analysis.

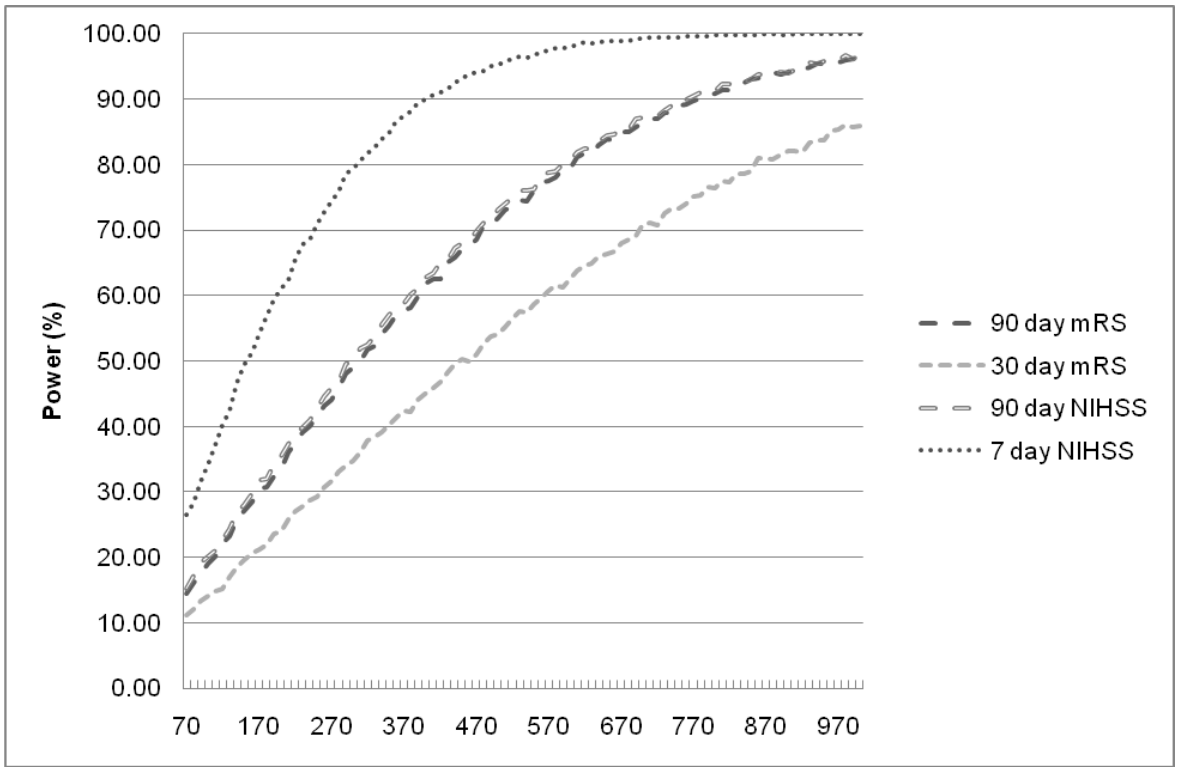

Figure 6-2: Line graph showing the relationship between sample size and statistical power for each of the four outcome measures using dichotomised analysis. For each outcome measure the most sensitive dichotomy was used.

Table 6-2 shows the minimum sample size required for each outcome measure to achieve statistical power greater than 80% and 90%. By both methods of analysis 7 day NIHSS is the most sensitive outcome measure and 30 day mRS the least. Between the two 90 day outcome measures, the NIHSS is modestly more sensitive than the mRS, though this difference is negligible on dichotomised analysis.

Table 6-2: Table showing the minimum sample size required to achieve statistical power of 80% for each outcome measure using ordinal and dichotomised methods. For each outcome measure the most sensitive dichotomy was used.

Endpoint	Ordinal	Dichotomised
80% Power		
90 day mRS	480	610
30 day mRS	620	860
90 day NIHSS	420	600
7 day NIHSS	370	310
90% Power		
90 day mRS	640	780
30 day mRS	800	>1000
90 day NIHSS	540	770
7 day NIHSS	480	410

Figure 6-3 shows the sensitivities of all of the 7 day NIHSS endpoints investigated. With the exception of the dichotomy 0-1, all methods of analysing 7 day NIHSS show high levels of sensitivity. The dichotomy: NIHSS score of 0-1 or an improvement by 8 or more points from baseline appears the most sensitive however the differences in sensitivity between the 3 most sensitive methods are modest.

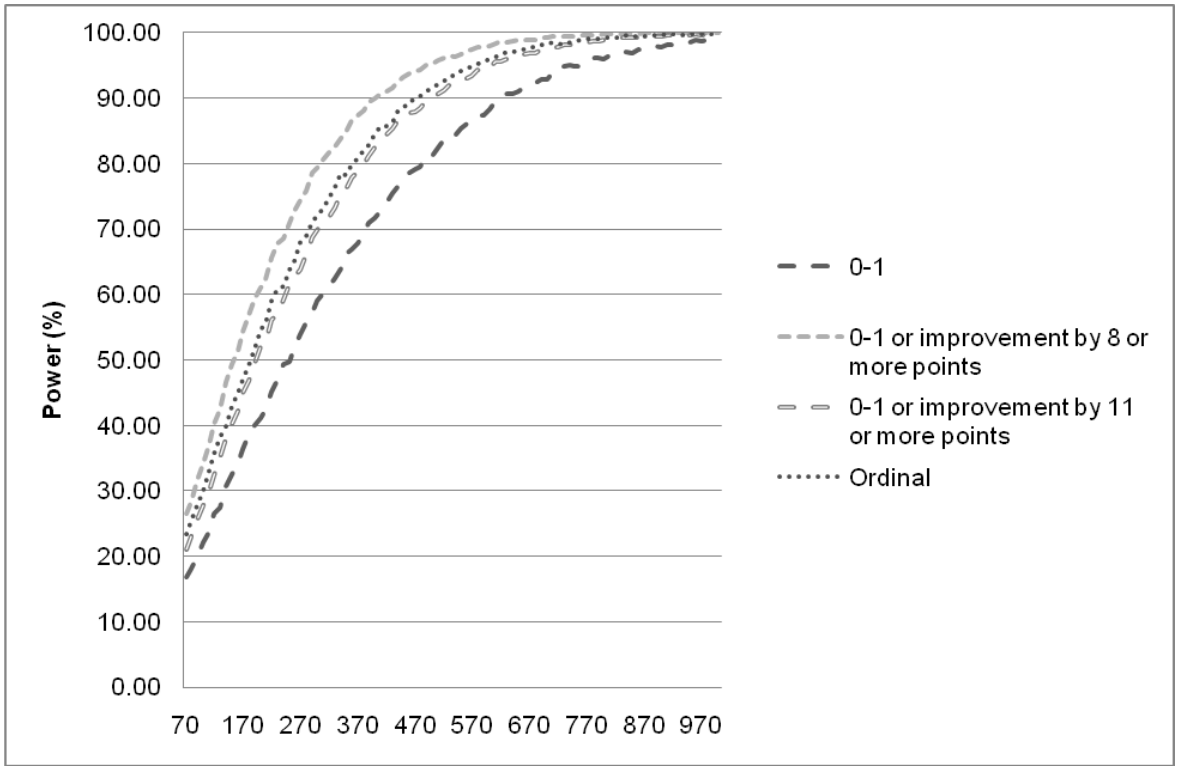

Figure 6-3: Line graph showing the relationship between sample size and statistical power for each of the 7 day NIHSS endpoints studied.

6.3.2. Validation Analysis

From the pooled thrombolysis trials, data on 2230 patients were available, of whom 2199 had a measure for each of the investigated outcomes. 1111 (49.8%) were treated with thrombolysis. The baseline demographics, previously published[171 229], were comparable between groups.

In this cohort of patients, dichotomised endpoints were more sensitive to thrombolysis treatment effect, Day 7 NIHSS requiring the lowest sample size per group. This was supported by results from ordinal analysis. The distributions of power and sample size for each (dichotomised) outcome measure are shown in Figure 6-4.

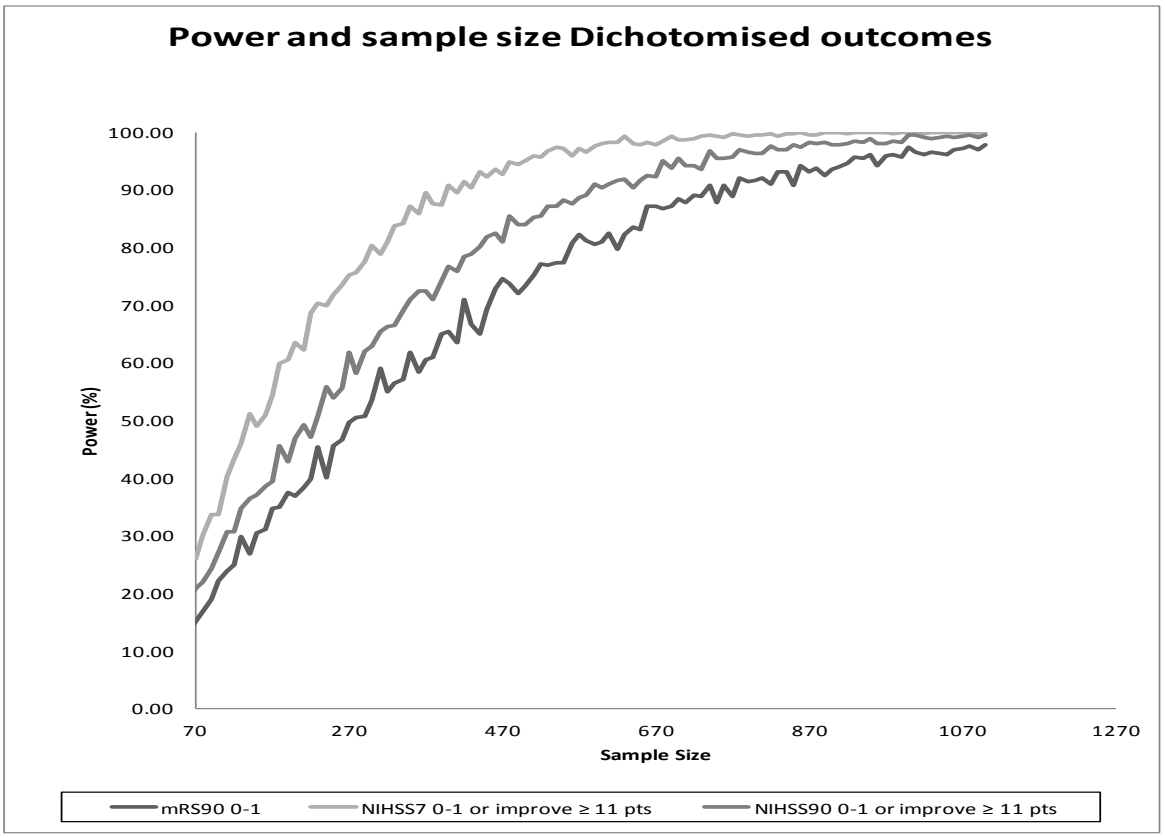

Figure 6-4: Line graph showing the relationship between sample size and statistical power for each of the 3 outcome measures using dichotomised analysis.

Table 6-3 shows the minimum sample sizes required for each outcome measure to achieve statistical power at typical choices of 80% or 90%.

For outcome at day 90 the NIHSS was only moderately better than the mRS, irrespective of whether these were analysed as ordinal or dichotomised measures.

Table 6-3: Table showing the minimum sample size required to achieve statistical power of 80% for each outcome measure using ordinal or dichotomised methods.

Endpoint	Ordinal	Dichotomised
80% Power		
90 day mRS	930	590
90 day NIHSS	910	440
7 day NIHSS	410	300
90% Power		
90 day mRS	>1000	740
90 day NIHSS	>1000	590
7 day NIHSS	590	400

6.4. Discussion

6.4.1. VISTA Analysis

We found NIHSS as measured at 7 days to be the outcome measure most sensitive to the treatment effect of thrombolysis. This is consistent with conclusions from independent datasets. Nevertheless, potential confounding influences should be considered.

Allocation to the thrombolysis and control groups in our study was not randomised, with choice to treat based on a range of clinical factors, some undocumented. Also, outcome assessors, although masked to investigational treatment allocation were not blinded to use of thrombolysis. Some reassurance derives from correspondence of the overall treatment effect, as measured by 90day mRS, with results of the pooled randomised trials.[7] Given these limitations our results deserve external validation using randomised trial data.

The VISTA analysis presented here complements reports by Young *et al*[299] and Broderick *et al*[230]. Both found dichotomised NIHSS endpoints to be more sensitive to a simulated treatment effect than various dichotomised disability endpoints. Broderick et al reported 24h dichotomised NIHSS to be the most powerful endpoint. Our analysis is based on a sample that is eight-fold larger, and could adjust for known baseline prognostic factors such as age and NIHSS. We have considered the recently favoured and statistically more powerful ordinal approaches.

The NIHSS reflects impairment, the clinical domain in which the effects of acute stroke therapies are likely to be most marked[299-300]. In contrast, the mRS covers a broader domain and is influenced by extraneous factors[300] that acute stroke therapies may not influence. Restricting such background noise may improve sensitivity to acute treatment effects.

This interpretation may also contribute to the greater sensitivity of the NIHSS as measured at 7 days compared to 90 days. Extraneous factors may have a greater influence on 90 day than 7 day NIHSS, with the latter better reflecting treatment effects in isolation from the myriad factors which come into play after discharge.

For pivotal trials in acute stroke, EMEA supports neurological scales mainly as a supplement to ordinal analysis of an activity scale (primarily mRS)[297]. However for early phase, exploratory research, 7 day NIHSS score has much to recommend it as an endpoint. It is sensitive to treatment effects, requiring comparatively low sample sizes to achieve desirable levels of statistical power. Moreover by recording outcome at 7 days, when most participants may still be in hospital, losses to follow-up will be minimised and unrelated adverse events should have less influence on detection of adverse reactions.

Trial management decisions, such as dose escalation between patient cohorts, can be expedited and costs contained.

In summary, seven day NIHSS score appears an ideal endpoint for the early exploratory testing of novel agents. Promising agents could then be validated in larger phase III trials using the mRS at 90 days to inform licensing and purchasing decisions, with validation of the 7-day NIHSS endpoint on the same sample to permit use of prior data as the necessary supporting evidence for regulatory submissions.

6.4.2. Validation Analysis

We found NIHSS as measured at 7 days to be the outcome measure most sensitive to the treatment effect of thrombolysis[222]. This is consistent with NINDS trial data[230].

Our present analysis is based on an observed rather than simulated treatment effect, and is adjusted for baseline prognostic factors such as age and NIHSS. In addition, we consider ordinal endpoints, which have recently been favoured over dichotomisation[221 294].

The validation analysis presented here favours dichotomised endpoints for all outcomes presented. This contrasts with most trial circumstances, including explorations within VISTA, in which ordinal analysis is considered more powerful. Whilst this differs also from our prior publication which for all outcomes other than NIHSS day 7 favoured ordinal analysis, the case mix in the pooled RCT data includes a large cohort of patients in whom thrombolysis treatment was initiated after 3h. This cohort makes up over half of the data[171]. Patients treated with thrombolysis between 3 and 4.5 h derive benefit according to mRS 0-1 but there is also a neutral or even marginal adverse effect on severe outcomes of disability and mortality:[4] in this case mix, the ordinal analysis still better represents the net effect of treatment across a population but the dichotomised approach is more

sensitive to the biological signal. It is illustrated by comparison of the mRS distributions shown in figure 2 of the updated pooled analysis paper [171].

90 day mRS remains the preferred outcome measure for pivotal trials in acute stroke, as the functional domain it measures is more relevant to individual patients and healthcare providers than the domain of disability measured by the NIHSS[140 240].

However, 7 day NIHSS has much to commend it for use in exploratory trials of novel agents. Its sensitivity would allow desirable levels of statistical power to be achieved with comparatively small sample sizes; and by recording outcome at 7 days when most patients will still be in hospital losses to follow-up will be minimised. During conduct of a trial, early access to relevant data could inform trial management decisions such as dose escalation. The use of 7 day NIHSS would allow exploratory phase II trials to be undertaken more quickly and with smaller sample sizes, reducing costs and hastening the arrival of promising agents to the market (or more timely discontinuation of development). Promising agents could then be validated in larger phase III trials using the more relevant outcome measure of 90 day mRS.

In summary, validation on external RCT data has shown that seven day NIHSS score offers statistical advantages as an endpoint for the early exploratory testing of novel agents, particularly reperfusion strategies. Detecting an early signal of treatment benefit or futility in acute stroke trials is economically, scientifically and ethically desirable.

Chapter 7
Exploration of time-course combinations of outcome scales in stroke recovery.

7.1. Background

Clinical trials for treatment of acute ischaemic stroke require large numbers of patients and are expensive to conduct. Treatment is typically administered within the first hours or days after stroke onset. Outcome is usually assessed by a single measure, the most common being the modified Rankin scale (mRS)[127 260]. Any strategy that can reduce cost or deliver more reliable answers on safety and efficacy of the investigational treatment would be welcome for future exploratory testing of novel treatments.

Some trialists have taken advantage of the fact that there are numerous scales available to measure various domains of neurological and functional recovery[96 200]. The ICTUS trial was designed to replicate results of meta-analysis of citicoline in acute stroke. The chosen primary outcome was recovery at 90 days as measured by a global test combining the favourable responses from Barthel index (BI), mRS, and National Institutes of Health Stroke Scale (NIHSS). If the direction of change generated by the experimental treatment is the same on each of these measures, then the statistical analysis is most efficiently achieved by testing with a global procedure[96]. The clinical relevance of this procedure

may be strengthened if it combines outcome scales that measure different aspects of expected stroke recovery, e.g. mobility and cognitive function. Both the ICTUS and NINDS trial Part 2 investigators considered that a multidimensional outcome measure was desirable[96 200 301]. Pocock has indicated that the use of such a combined outcome can enhance statistical power[302].

However, global procedures need not be limited to data from a day 90 assessment. All clinical trials include longitudinal assessments of outcome which may include day 7 NIHSS score, day 30 mRS and day 90 mRS. The incorporation of all these measures could afford a more subtle assessment of outcome and recovery. For example a participant with an mRS score of 3 at both 30 and 90 days after stroke has perhaps experienced a more favourable outcome than a participant who attained a day 90 mRS score of 3 after a day 30 mRS of 5. A subjective ranking technique allows each patient in a clinical trial to be ordered according to all their trial experiences and the distribution of ranks between treatment groups can then be compared. This method was applied to data from the Systolic Hypertension in the Elderly Program[303] and there provided a sensitive measure of treatment effect.

In Chapter 6 it was shown that an outcome assessment just 7 days after stroke onset gives a statistically stronger indication for putative treatment effect than a 90 day functional recording[222]. Feng et al[304] proposed repeated-measures analysis as an alternative method for assessing treatment effect, which could be used in future stroke trials if outcomes of interest are collected across several time points. In their analysis, the mRS, NIHSS, BI and GOS were each sensitive to treatment effect with this approach. Li et al[305] examined the various mRS time points used in NINDS rt-PA stroke trial and showed that a treatment effect could be observed at 7–10 days, 3 months, 6 months, and 12

months post stroke onset. We aimed to examine whether the combination by a global testing procedure of early and late assessments into an integrated view of recovery across the first 90 days after stroke offers yet more statistical information or power than standard outcome assessment. We also investigated a more simple approach by, combining ranks of different scores.

Feng et al showed that both NIHSS and mRS were sensitive to treatment effect at 90 and 7 days post stroke[304]. We hypothesised that after adjustment of outcomes for known prognostic variables (age, baseline NIHSS), the apparent treatment effect of the proven therapy intravenous thrombolysis would be detected more strongly if the outcome measure uses a combination of early and late measures (e.g. 7 day NIHSS combined with 90-day mRS) than either early or late measure alone.

7.2. Methods

7.2.1. Data source and patients

Data were sought and extracted from the Virtual International Stroke Trials Archive (VISTA) on participants enrolled in neuroprotectant trials from between 1998-2007. To be eligible for analysis, participants needed to have recordings available for initial stroke severity (baseline NIHSS score), and prognostic factors such as age, prior atrial fibrillation, hypertension and diabetes available; documentation of use or avoidance of thrombolysis; and post-treatment measurements of outcome. We restricted our analysis to patients who had measurements of modified Rankin scale at 90 days (mRS90), 30 days (mRS30), 7 days (mRS7) and National Institutes of Health Stroke Scale at 90 days (NIHSS90) and 7 days (NIHSS7).

7.2.2. Preliminary analysis

Interdependence between early and late outcome measures was examined using Spearman-Rank correlation calculated as partial correlations, adjusting for prognostic covariates. Amongst other values, this provided correlation coefficients (r) i.e. the strength and direction of the relationship between two outcome scales. For reporting purposes we calculated coefficients of determination (R^2), i.e. the percentage variation of one scale explained by another scale after adjusting for the effect of any covariates. Due to the ordinal nature of mRS, R^2 values adjusted for covariates were calculated as pseudo R^2. This was done using the Logistic procedure in SAS, discussed further in Chapter 2.

Treatment responsiveness was explored, analysing adjusted outcomes among the following two populations: thrombolysed (as standard of care) patients versus non-thrombolysed (as standard of care) patients. Odds ratios (OR) were indicative of this sensitivity to treatment effect. For the thrombolysed group, patients were selected regardless of the time from treatment onset at which t-PA was started, though this was generally below 3 hours. For the control group, the time window from onset time to treatment was between 1 and 7 hours. Patients were not excluded on account of stroke severity. Outcomes were handled as ordinal measures as appropriate to maximise power, and adjustment for major covariates was undertaken.

MRS at 90 days, 30 days and 7 days (mRS90, mRS30 and mRS7 respectively) and NIHSS at 90 days and 7 days (NIHSS90 and NIHSS7 respectively) were considered as ordinal scales. For global outcome, three combinations were considered: [mRS7, mRS30, mRS90], [NIHSS7, NIHSS90] and [NIHSS7, mRS90]. MRS was analysed in its standard form, whereas NIHSS was stratified into groups: ≤4, 5-8, 9-12, 13-16, 17-20, 21-24, and ≥25[222 272]. Death was considered as a separate group.

We analysed sensitivity to treatment effect, using proportional odds logistic regression[272] for ordinal scales individually. Generalised Estimating Equations (GEEs) were used, when analysing this for scales combined in a global outcome. The OR that we derived are allowed comparison between ordinal scales and global outcomes, discussed further below. We chose ordinal logistic regression in favour of linear regression since the scales are neither truly linear nor normally distributed. We adjusted for age and baseline NIHSS, treated as continuous variables. Our choice of baseline factors for adjustment was based mainly on the fact that age and baseline severity are the two most powerful prognostic factors for stroke[63 241] and are the most common adjusting factors used in analysis of stroke data. Testing for significance of any association, with ordinal scales, was performed using the Cochran Mantel Haenszel (CMH) statistic. The CMH test provides a conservative estimate of statistical significance[272] and was performed for illustrative purposes. Stratification by covariates in the CMH test is limited by the sample size and precludes simultaneous adjustment for all possible variables.

7.2.3. Simple combination of ranks

A simple method of combining scales within each time-point was initially proposed. NIHSS was converted to an ordinal measure in order to make it comparable to the ordinal NIHSS that was used within the global test. This method treats scale1 as the scale taken at the latest time point. The next scale was ordered within each grade of the first scale and so on. This is done using the formula $(1 \times Scale1) + (100 \times Scale2) + (10{,}000 \times Scale3)$. For example: if we were interested in a subject with an mRS90 score of 3, an mRS30 score of 2 and an mRS7 score of 2 their combined score would be 20203. Each of these calculated scores were then ranked in ascending order to generate a discrete outcome measure.

7.2.4. Global test

The calculation of a global test statistic is somewhat complex. For binary outcomes, calculation is based on the proportion of good outcome (success) in each group, as discussed by Dymova et al[306] and Bolland et al[307]. Calculating this statistic manually becomes more difficult when the scales have more than two groups so an alternative approach was sought. Here we used the approach outlined by Bolland et al[307] allowing for prognostic covariates, applying generalised estimating equations (GEE) using PROC GENMOD in SAS[308] and the non-linear mixed effects library in R[309]. The GEE approach was used as an initial strategy in the ICTUS trial[307] and was also applied in the NINDS trial of rt-PA part 2[96]. The method proposed here has been adapted for use with ordinal scales by using the multinomial distribution.

For each global test combination a GEE multinomial regression model was fitted to the patient data. A factor distinguishing each assessment scale was included in the model and prognostic covariates were fitted allowing for a separate effect for each outcome. This was done by fitting each covariate as an interaction with outcome scale. The parameter estimate obtained for treatment effect from this regression corresponds to a common log-odds ratio for treatment on all of the combined scales. This along with the standard error for this estimate can be used to calculate the OR and the 95% CI for OR.

A sequential design approach to the global test has also been proposed by Whitehead et al[301] but was not used here. The benefit of the GEE approach is both its simplicity and compatibility across statistical packages so simulations could be performed in R.

7.2.5. Simulations

We performed multiple re-sampling to assess relation between sample size and power for each ordinal scale and global outcome. From our VISTA dataset, 10,000 random samples of patients were drawn at each of a series of sample sizes ranging from 1000 downwards in decrements of 10. Each sample consisted of equal numbers of rt-PA treated and control patients. For every sample size, the percentage of trials yielding a statistically significant treatment effect was recorded. This percentage approximates the statistical power of each outcome measure to detect thrombolysis treatment effects at that sample size. We present this graphically, plotting sample size against power, setting $p<0.05$. For treatment effect, if 95 out of 100 simulated trials achieved $p<0.05$, this is 95% power.

Simulations were performed using R version 2.12.1, all other analyses used SPSS 19 and SAS 9.2.

7.3. Results

7.3.1. Preliminary Analysis

Data on 4077 patients were available, containing a value for each of the outcome measures. The age range of patients included in the cohort was 20 to 97 years, with a mean of 68.0 years. Baseline characteristics are displayed in Table 7-1.

Table 7-1: Baseline characteristics of patients included in VISTA dataset. Percentages refer to baseline characteristic % within each treatment group

		rt-PA		
		Yes	**No**	Total
Previous Stroke n(%)	**Yes**	210 (13.6)	589 (23.2)	799 (19.6)
	No	1330 (86.4)	1948 (76.8)	3278 (80.4)
Hypertension n(%)	**Yes**	473 (30.7)	597 (23.5)	1070 (26.2)
	No	1067 (69.3)	1940 (76.5)	3007 (73.8)
Diabetes n(%)	**Yes**	287 (18.6)	617 (24.3)	904 (22.2)
	No	1253 (81.4)	1920 (75.7)	3173 (77.8)
Myocardial Infarction n(%)	**Yes**	210 (13.6)	296 (11.7)	506 (12.4)
	No	1330 (86.4)	2241 (88.3)	3571 (87.6)
Atrial Fibrillation n(%)	**Yes**	344(22.3)	639 (25.2)	983 (24.1)
	No	1196(77.7)	1898 (74.8)	3094 (75.9)
Gender n(%)	**Male**	900 (58.4)	1355 (53.4)	2255 (55.3)
	Female	640 (41.6)	1182 (46.6)	1822 (44.7)
Time to treatment	**Median (IQR)**	3 (1)	4 (2)	-
Age	**Mean (SD)**	67 (13)	69 (12)	-
Baseline NIHSS	**Mean (SD)**	13 (5)	11 (5)	-

We assessed the interdependence between scales amongst patients who had a measurement for each of mRS90, mRS30, mRS7, NIHSS90 and NIHSS7. Estimates of adjusted interdependence among the five outcome scales are displayed in Table 7-2 as both partial correlation coefficients (r) and adjusted coefficients of determination (R^2). Within this cohort of 4,077 patients, median scores (and IQR) for the outcomes were as follow: 3 (3) for mRS90; 3 (3) for mRS30; 4 (2) for mRS7; 3 (7) for NIHSS90; and 5 (9) for

NIHSS7. Median scores of NIHSS relate to the full scale and not the ordinal form used in further analysis.

Table 7-2: Interdependence of ordinal stroke scales. Values are adjusted for baseline NIHSS and age.

		mRS90	mRS30	mRS7	NIHSS90	NIHSS7
mRS90	r R^2 (%) p-value n	1 100 4077	0.811 73.8 <.0001 4077	0.70 61.1 <.0001 4077	0.719 69.2 <.0001 4077	0.658 51.1 <.0001 4077
mRS30	r R^2 (%) p-value n		1 100 4077	0.820 74.6 <.0001 4077	0.613 47.8 <.0001 4077	0.716 59.6 <.0001 4077
mRS7	r R^2 (%) p-value n			1 100 4077	0.531 40.3 <.0001 4077	0.723 60.9 <.0001 4077
NIHSS90	r R^2 (%) p-value n				1 100 4077	0.672 52.4 <.0001 4077

All scales appeared sensitive to rt-PA therapy, after adjustment for age and baseline NIHSS. The mRS90 had an unadjusted OR for better outcome with rt-PA of 1.16 and adjusted OR of 1.56. The NIHSS90 had an unadjusted OR of 1.10 and adjusted OR of 1.62. The adjusted ORs for all scales individually, incrementally ranked and combined with a global test are given in Table 7-3 below. Two combinations of the global test give the strongest sensitivity to treatment effect with adjusted OR for the global outcomes of (mRS90, NIHSS7) and (NIHSS90, NIHSS7) were 1.69 and 1.73 respectively. The combination of mRS90, mRS30 and mRS7 is the least sensitive to treatment effect when using both the global test and the incremental ranking method. It would appear from

these results that by incrementally ranking the outcome measures no benefit is gained in comparison to the use of individual scales.

Table 7-3: Scales and combinations ranked in order of responsiveness to treatment effect. Adjusted for Baseline NIHSS & Age

Scale(s)	OR	95% CI	p-value
Global test2 (NIHSS7, NIHSS90)	1.73	(1.52, 1.95)	<0.0001
Global test3 (mRS90, NIHSS7)	1.69	(1.51, 1.90)	<0.0001
NIHSS7	1.69	(1.50, 1.92)	<0.0001
Rank2 (NIHSS90, NIHSS7)	1.67	(1.47, 1.89)	<0.0001
NIHSS90	1.62	(1.41, 1.87)	<0.0001
Rank3 (mRS90, NIHSS7)	1.61	(1.44, 1.80)	<0.0001
mRS90	1.56	(1.39, 1.75)	<0.0001
Global test1 (mRS90, mRS30, mRS7)	1.55	(1.39, 1. 27)	<0.0001
Rank1 (mRS90, mRS30, mRS7)	1.53	(1.37, 1.71)	<0.0001
mRS30	1.47	(1.31, 1.65)	<0.0001
mRS7	1.40	(1.25, 1.58)	<0.0001

7.3.2. Simulations

Table 7-4 shows the simulation results giving minimum sample sizes required for each end point to achieve statistical power at typical choices of 80% or 90%. The smallest sample size required obtaining statistical power of ≥ 80% for mRS90, NIHSS7, and global outcomes of mRS0 and NIHSS7 combined and NIHSS90 and NIHSS7 combined, were 500, 490, 400, and 380, respectively. A global outcome combining NIHSS7 with NIHSS90 was found to be the most sensitive measure and mRS7 the least.

Table 7-4: Estimates of minimum sample size required to achieve power of 80% and 90% for ordinal scales, global outcomes and ranking combinations of these scales. mRS indicates modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale.

	80% Power	90% Power
End-point	**Sample size**	**Sample size**
mRS7	-	-
mRS30	740	-
mRS90	500	650
NIHSS7	490	630
NIHSS90	610	800
Rank1 (mRS90, mRS30, mRS7)	540	710
Rank2 (NIHSS90, NIHSS7)	460	580
Rank3 (mRS90, NIHSS7)	440	590
Global test1 (mRS90, mRS30, mRS7)	700	-
Global test2 (NIHSS90, NIHSS7)	400	530
Global test3 (mRS90, NIHSS7)	380	480

As was anticipated, given the GEE regressions in Table 7-3 the incremental ranking method provides no benefit when compared to individual measures. Figure 7-1 below shows the relationship between desired power and required sample size for each of the three ranking methods rank1, rank2 and rank3, the 'gold standard' outcome measure of mRS day 90 most frequently used in trials and ordinal NIHSS day 7 shown in Chapter 6 to be the most sensitive ordinal measure when testing the treatment effect of thrombolysis. This confirms the assumption that none of the three ranking methods provide any benefit to the individual scales as there are no clear distinctions between the lines and there is a great deal of overlap.

Figure 7-2 below shows the relationship between desired power and required sample size for each of the three global test using the GEE methods, the 'gold standard' outcome measure of mRS day 90 and ordinal NIHSS day 7. This confirms the results from Table 7-3, showing that the global test combination 2(NIHSS90, NIHSS7) and global test combination 3 (mRS90, NIHSS7) are the most sensitive to treatment effect. Looking at the plot both perform consistently better than all other scales with the (NIHSS90, NIHSS7) combination proving the most sensitive to the treatment effect of thrombolysis. Figure 7-3 shows the global testing and ranking methods alone. It can be seen that the global test 2 (NIHSS7, NIHSS90) and 3 (mRS90, NIHSS7) perform consistently better than all of the incremental ranking methods. The combination of (mRS90, mRS30 and mRS7) using both the ranking (ranking1) and global test (global1) appears to perform worse than any other.

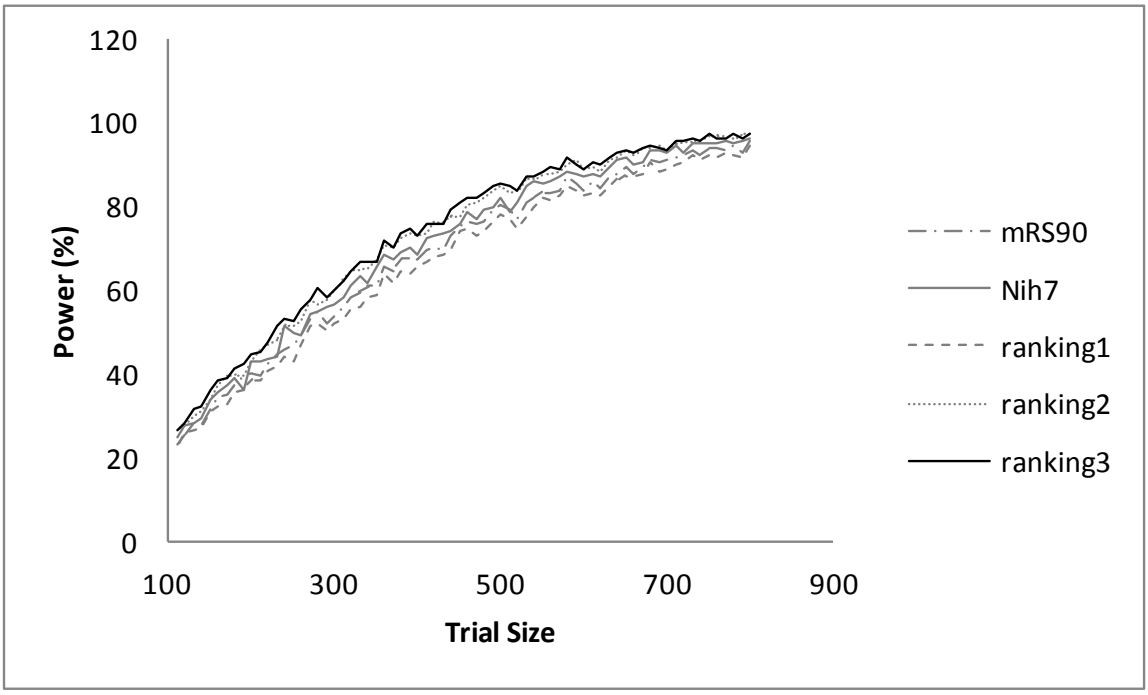

Figure 7-1: Relationship between power and sample size for each of the combinations of ranks as well as the standard ordinal outcome of mRS day 90 and the most sensitive outcome measure found in Chapter 6, NIHSS day 7.

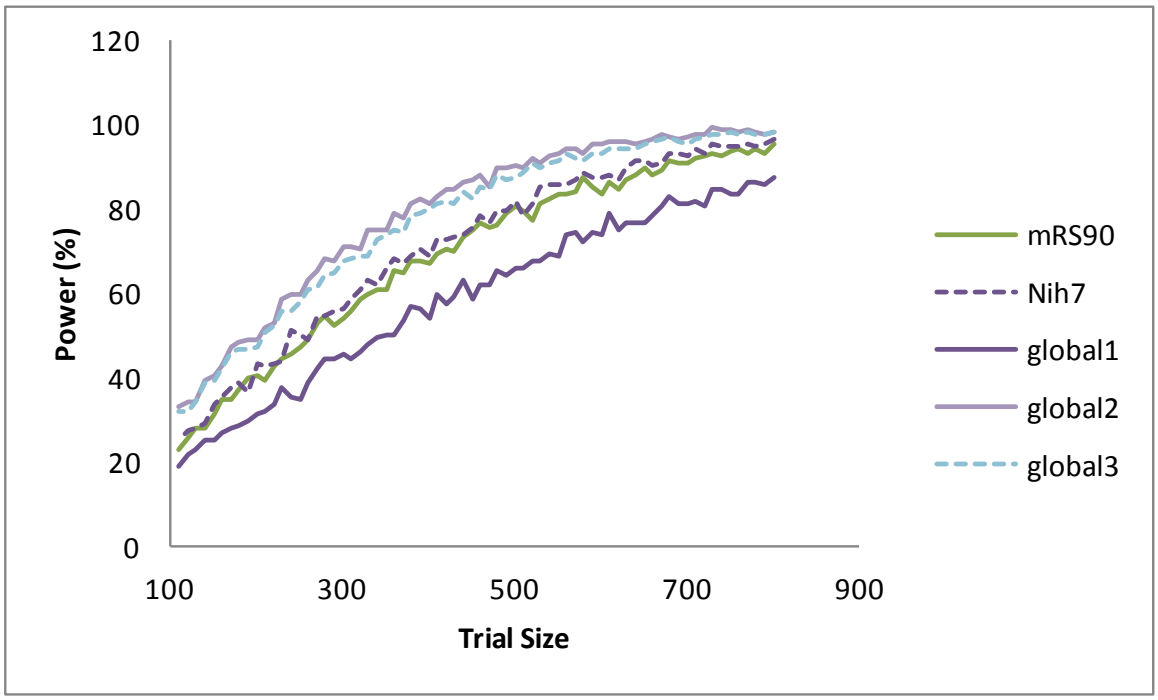

Figure 7-2: Relationship between power and sample size for each of the global test combinations as well as the standard ordinal outcome of mRS day 90 and the most sensitive outcome measure found in Chapter 6, NIHSS day 7.

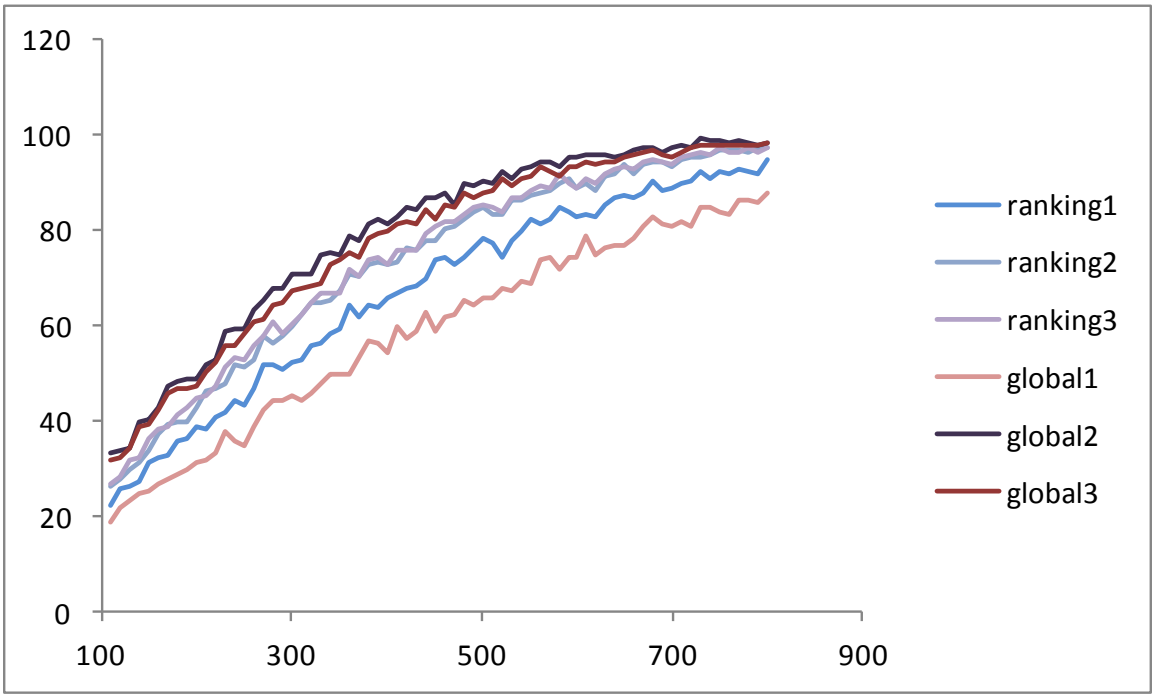

Figure 7-3: Relationship between power and sample size for each of the combinations of ranks and each of the global tests.

7.4. Discussion

The global test combination of NIHSS90 with NIHSS7 appears to offer incremental sensitivity to treatment effect compared to the ordinal forms of these scales alone. The combination of mRS90 with NIHSS7 did not increase the sensitivity to treatment effect when compared to NIHSS alone, but offers a broader clinical measure without loss of statistical power. This may have practical importance, since regulatory bodies may consider NIHSS at day 7 to be too early a measure and too neurological, i.e. divorced from function, for treatment approval purposes; but may be more open to consider the enhancement of traditional 90-day mRS with NIHSS7.

As expected, we found greater advantage from the combination of measures that correlated with each other only modestly. However the use of a more simple incremental approach combining the scales using a ranking method offered no benefit to the ordinal scales.

The increments in treatment effect estimates that we observed when using the global test approach were modest but the impact that these limited benefits deliver to sample sizes is surprisingly great. For an enhancement in the odds ratio for treatment benefit of only 0.13 (from 1.56 with mRS 90 to 1.69 with mRS90 and NIHSS7), the necessary sample size for 80% power fell by 20% from 500 to 400 patients. A second advantage may accrue from the combination approach: treatments that reduce early disability should also limit in-hospital and rehabilitation costs: it is desirable that we should detect such benefits as part of the outcome measure.

Our analyses gain strength from the large size of our dataset, the rigorous conditions under which the data were collected and the breadth of the trials and centres from which they derive. However, our estimates are weakened by the non-random allocation in this sample between "treatment" and "control" and by differences in baseline prognostic factors between our two treatment groups. In particular, there may be a difference in the timing of baseline NIHSS measurements between our treatment groups, with NIHSS assessments possibly undertaken earlier in the thrombolysed patients than controls. This may affect the overall estimate of "treatment" effect size, but should not interfere with the relative power of various outcome measures.

Analysis of the NIHSS or mRS across time has been described in the statistical literature[305]. Feng et al recently demonstrated that "repeated-measures analysis allows for a more comprehensive understanding of the clinical benefit of study intervention at any defined study point or throughout the entire study period"[304]. Our findings take this a step further since unlike Li et al or Feng et al, we combined different scales across time (i.e. day 90 mRS and day 7 NIHSS).

Consideration of the time course of recovery and statistical approaches to combine measures that best reflect the nature of the deficit at set intervals after stroke both have biological plausibility. Our report suggests that this deserves further exploration since statistical gains could be meaningful and inexpensive.

Chapter 8
Exploration of case-control matching using historical controls

8.1. Background

Clinical trials for treatment of acute ischaemic stroke require large numbers of patients enrolled under tight timelines across many centres and are expensive to conduct. Patients or their relatives may be asked to consent urgently to research participation that carries life-threatening clinical implications during a time of extreme anxiety and diagnostic uncertainty.

With increasing use of thrombolysis[96 171 310] and in some centres endovascular approaches, and with disappointing results from recent trials of neuroprotectant approaches[188 195 311], translational development of treatments for acute stroke is becoming more challenging. Investigators are seeking research strategies that will ease ethical burdens and will limit sample sizes and cost without compromising reliability: any valid method to detect or exclude a biological signal with a new treatment and reach an early "go/no go" decision would be attractive.

For early phase research, especially when endovascular devices are involved, employing invasive procedures for which a placebo control may be considered impractical or

unethical, researchers are turning towards single arm designs or comparisons against historical controls[312-313] Also in pilot studies and when patients are considered at very high risk if untreated, i.e. if clinical equipoise is not considered to be present, comparisons against historical controls have been considered[314-317].

Abandoning the rigour of the blinded RCT carries substantial penalty in loss of reliability and should not be undertaken lightly. However, there may be statistical strategies that can strengthen these compromises when they have been considered unavoidable. Numerous strategies have been discussed in statistical literature[263 268 318-323] with diverse opinions. Such strategies are now attracting at least tacit support from regulatory bodies[320 324].

With the increasing availability of data-banks and registries such as VISTA and SITS, opportunities to access historical controls are likely to be grasped, possibly without close attention to the reliability of the ensuing comparisons. Rather than opting for a 'quick fix' prompted by ready access to generous historical patient data, we wished to explore aspects of reliability in a systematic manner.

Within the Virtual International Stroke Trials Archive (VISTA), we have access to datasets from several trials. Most of these trials made careful recordings of a variety of baseline measures along with recordings of outcome at three months after stroke onset, giving us the opportunity to investigate the matching process. Various methods of matching as well as a variety of variable combinations for matching will be assessed under different conditions. Variables for matching are generally chosen based on clinical knowledge of the disease itself[325],however we will use a multiple re-sampling approach, discussed

below, to assess which combination of the variables widely accepted as being predictive of outcome is the most responsive for matching.

8.2. Methods

8.2.1. Data Source and Patients

Data were extracted from VISTA that met overarching selection criteria on data availability, namely: NIHSS recorded within 6h of stroke onset and at 90 days; modified Rankin Scale score recorded at 90 days; age; exposure to iv thrombolysis (or not) and date of trial falling within period 1998-2008. All available medical history data were sought where available.

The data contained two distinct populations. One population originating from pooled neuroprotectant trials and the other from pooled thrombolysis trials. Due to the nature of thrombolytic therapy, inclusion into a thrombolysis trial would be more restrictive than that of a neuroprotectant. Neuroprotectant trials tend to be more inclusive and to have had a wider time window for inclusion (Figure 8-1). Inclusion criteria of OTT <6h, baseline NIHSS 4-20 and an age range of 18-80 were applied to all data to ensure consistency with the entry criteria for European thrombolysis trials. Initially no further restriction was placed to ensure the largest sample for matching.

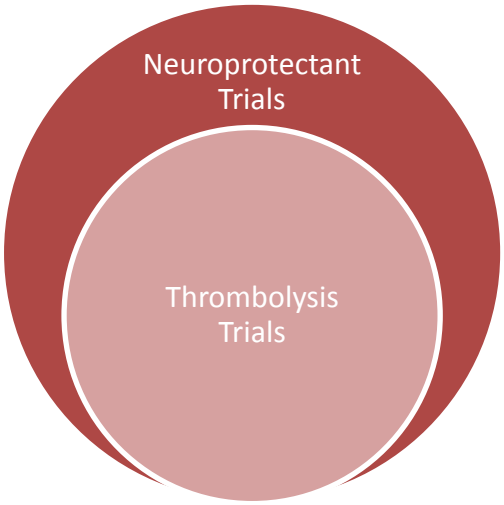

Figure 8-1: Venn diagram illustrating overlapping inclusion in thrombolysis and neuroprotectant trials

8.2.2. Simulation study comparing methods and models for matching

The goal was to simulate a trial population using VISTA data that consisted of a sampled 'case' group and a 'control' group. The control group are matched to the case group by propensity scores utilising various matching methods. Propensity score matching can be performed with any combination of covariates. Within the simulation study six variations of covariate models were considered (Table 8-1). Each model generated a different matched group by attempting to balance all covariates between the case and control groups.

Propensity score matching was performed using the MatchIt package in R. The matching methods utilised were nearest, full, optimal and genetic as discussed in Chapter 2. The exact matching method was not utilised as this has been previously investigated within VISTA[178 236]. In any case finding exact matches on all continuous variables such as age, glucose and diastolic blood pressure would be problematic. Due to the limitations of the MatchIt package only subjects with recorded values for all covariates can be employed for the matching process.

Table 8-1: Covariate models for matching

Model Number	Variables for matching
1	Age, baseline NIHSS and Hemisphere of stroke
2	Age, Baseline NIHSS, Hemisphere and onset time to study drug (OTT)
3	Age, Baseline NIHSS, Hemisphere, OTT, Glucose at baseline (mmol/L), Diastolic blood pressure (DBP) at baseline (mg/Hg)
4	Age, Baseline NIHSS, Hemisphere, OTT, History of Diabetes, History of Hypertension
5	Age, Baseline NIHSS, Hemisphere, OTT, Diabetes, Hypertension, Glucose at baseline and DBP at baseline
6	Age, Baseline NIHSS, Hemisphere, OTT, Diabetes, Hypertension, Glucose at baseline and DBP at baseline, Atrial Fibrillation (AF)

We undertook simulations to model the error that occurs due to the matching process. Using R statistical programming language, 1,000 mock trials were simulated. For each trial 500 subjects were sampled from all eligible untreated subjects, without replacement. This generated a case group with inclusion criteria already applied. A control group was generated from the remaining untreated subjects by propensity score matching as illustrated in Figure 8-2. In order to directly compare methods and covariate models for matching each were applied simultaneously to the data throughout the 1,000 simulations.

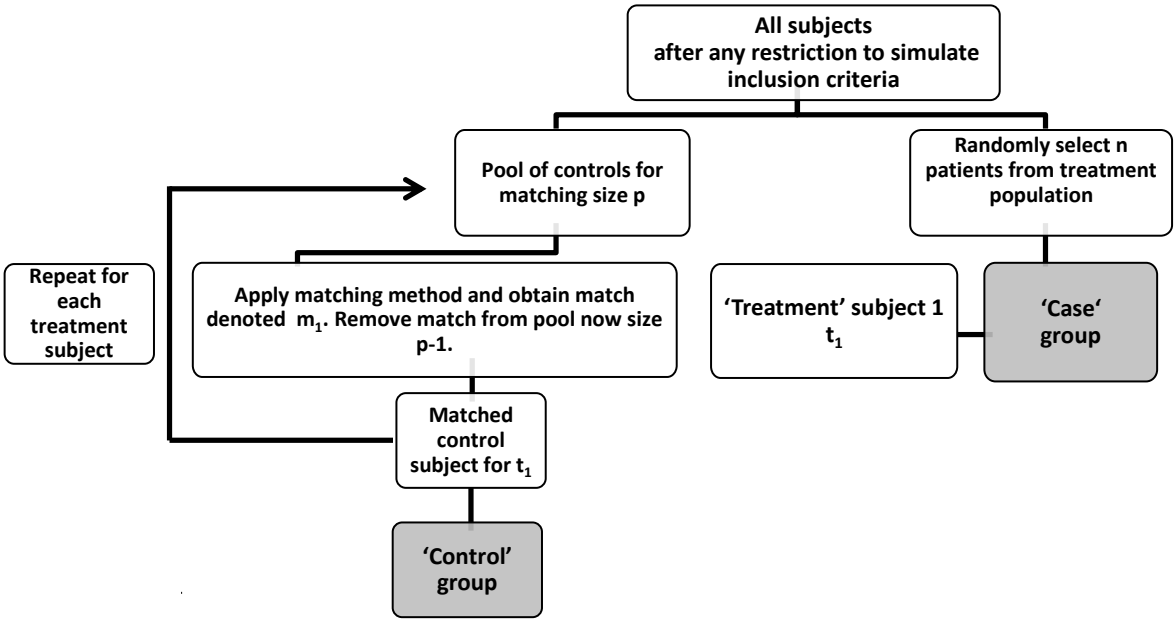

Figure 8-2: How matching for each subject works within each simulation

The aim of matching is to create a data set similar to what you would expect from a randomised design. Therefore we need to assess how close the distributions of the covariates are between the two groups, also referred to as the balance[269]. As outlined in Chapter 2, a propensity score is calculated for each subject based on all matching variables; this gives a one dimensional continuous variable for matching, sometimes referred to as 'distance'. The mean difference in propensity score between the case and control groups was recorded before and after matching. The smaller this difference is the more balanced the control and treatment groups are. This measure of balance is based solely on the matching variables. Along with these differences the overall percentage balance improvement in propensity score was recorded.

While these measures can be informative they can also be misleading: if the two groups were reasonably balanced before matching there could be very little improvement in balance. Alternatively, a large reduction in difference leading to a high percentage balance improvement may still yield unbalanced groups.

To investigate the matching process further: each simulated trial was analysed for treatment effect based on ordinal mRS day 90 and ordinal NIHSS day 90 illustrated in Figure 8-3. This treatment effect was analysed using proportional odds logistic regression across the full scale of the mRS and NIHSS, stratified into groups: ≤4, 5-8, 9-12, 13-16, 17-20, 21-24, and ≥25[222 272], recording both the p-value and odds ratio for treatment effect. In order to account for the variation related to the matching variables the analysis was adjusted for all covariates used for the propensity score calculation[263].

By sampling the two groups from one untreated population we exclude the influence of a treatment effect as well as any bias occurring when sampling from two separate populations. Given this we would expect to yield no treatment effect between groups. Consequently retaining the p-values from the analysis allows us to obtain the type I error rates. This is calculated as the proportion of times the null hypothesis was wrongly rejected. If matching was successful the type I error rate would be expected to be <0.05. This is assuming 5% error may occur by chance alone.

After separating the data into the two distinct populations, these simulations were repeated. Here the case group was sampled from the untreated thrombolysis trial population and the control group matched from the untreated neuroprotectant trial population. Again if matching was successful we would expect no observable treatment effect. By matching a control group to a case group taken from a different population we aimed to investigate if matching accounts for the disparity between the two populations.

Further restrictions were applied to the inclusion criteria, reducing the OTT to <4.5 hours and the simulations repeated again. Due to the lack of data availability covariate models 4, 5 and 6 (Table 8-1) could not be used for these simulations.

Finally, for investigative purposes only, untreated subjects from neuroprotectant trials were matched to those treated with thrombolysis as standard of care. Due to VISTA regulations subjects from the thrombolysis trial population could not be used here. Retaining the p-values when there is a known treatment effect allows us to calculate the type II error rate. This is calculated as the proportion of times a null hypothesis was wrongly accepted. The power of a study can be calculated from this as 1-type II error rate. Ideally this error rate would be as low as possible ensuring an adequately powered and precise study. The odds ratios allow us to see the direction of any observed treatment effect.

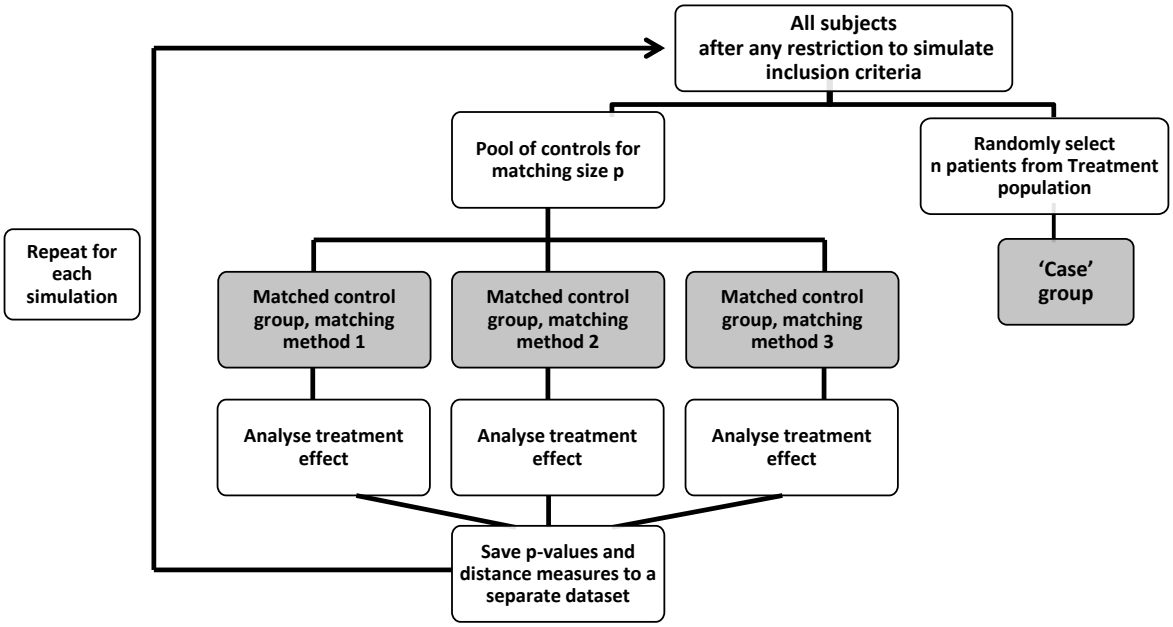

Figure 8-3: Outline of simulation process

8.3. Results

8.3.1. Available Data

After applying the initial restrictions to the entire dataset, data on 5076 patients were available. 2118 (41.7%) of subjects were thrombolysed. Baseline characteristics are displayed in Table 8-2.

Table 8-2: Baseline demographics of VISTA dataset after initial restriction applied. Demographics displayed for entire data broken up by treatment group as well as population.

Variable	Statistic	All Controls	All Treated	Control Neuroprotectant	Control Thrombolysis	Treated Neuroprotectant
Age (years)	N	2958	2118	2339	619	1526
	Mean(SD)	65.9 (10.9)	64.9 (11.5)	66.1 (10.9)	65.3 (11.1)	12 (7)
Onset time to treatment (hours)	N	2958	2118	2339	619	1526
	Mean(SD)	3.9 (1.1)	2.69 (1.1)	4.04(0.96)	3.45 (1.52)	2.42 (0.57)
Baseline NIHSS	N	2958	2118	2339	619	1526
	Median (IQR)	10 (7)	12 (7)	10 (7)	12 (7)	12 (7)
Glucose baseline (mmol/l)	N	2958	2118	2339	619	1526
	Mean (SD)	7.8 (3.4)	7.5 (3.5)	7.8 (3.3)	7.9 (3.9)	7.3 (2.9)
Diastolic blood pressure at baseline (mmHg)	N	2952	2107	2338	614	1526
	Mean (SD)	85 (15)	83 (15)	85.3 (16)	85.8 (13)	82.8 (15.9)
Sex						
Female	N (%)	1311 (44.3)	880 (41.6)	1010 (43.2)	301 (48.6)	607 (39.8)
Male	N (%)	1647 (55.7)	1238 (58.4)	1329 (56.8)	318 (51.4)	919 (60.2)
Hemisphere of Stroke						
Left	N (%)	1302 (44.0)	911 (43.0)	1018 (43.5)	284 (45.9)	646 (42.3)
Right	N (%)	1656 (56.0)	1207 (57.0)	1321 (56.5)	335 (54.1)	880 (57.7)
Diabetes						
No	N (%)	2265 (76.6)	1733 (81.9)	1731 (74.0)	534 (86.3)	1218 (79.8)
Yes	N (%)	693 (23.4)	384 (18.1)	608 (26.0)	85 (13.7)	308 (20.2)
Hypertension						
No	N (%)	825 (29.0)	731 (36.4)	565 (24.2)	260 (51.1)	486 (31.8)
Yes	N (%)	2023 (71.0)	1276 (63.6)	1774 (75.8)	249 (48.9)	1040 (68.2)
AF						
No	N (%)	2095 (78.8)	1486 (81.9)	1815 (77.6)	280 (88.1)	1224 (80.2)
Yes	N (%)	562 (21.2)	328 (18.1)	524 (22.4)	38 (11.9)	302 (19.8)

8.3.2. Simulation Study

8.3.2.1. Matching from the same population

It was postulated that sampling and matching from the same population would generate very similar groups regardless of the covariate model used. This was investigated in Table 8-3 below.

Table 8-3: Difference in propensity score between groups before and after matching along with percentage balance improvement between groups after matching. Case and control groups sampled from the same population.

Distance information for each model					
Covariate model	**Difference in propensity score between groups**	**Full matching**	**Nearest matching**	**Optimal matching**	**Genetic matching**
1	Mean before Mean after % Balance improvement	0.000824 0.00000397 96.13	0.00202 0.0000143 99.20	0.00124 0.0000224 96.88	0.00127 0.0000172 99.30
2	Mean before Mean after % Balance improvement	0.00135 -0.0000033 98.23	0.00225 0.000021 99.01	0.00147 0.000020 98.12	0.00211 0.000000297 99.55
3	Mean before Mean after % Balance improvement	0.00177 -0.00000204 98.77	0.00409 0.0000359 99.16	0.00259 0.0000134 99.10	0.00275 -0.00000328 97.89
4	Mean before Mean after % Balance improvement	0.00275 0.000000122 99.21	0.00272 0.0000257 99.04	0.00243 0.0000182 98.84	0.00320 0.00000494 99.20%
5	Mean before Mean after % Balance improvement	0.00345 0.000000308 99.21	0.00458 0.0000392 99.08	0.00429 0.0000279 99.24	0.00413 -0.00000343 96.92
6	Mean before Mean after % Balance improvement	0.00375 0.00000534 99.35	0.0538 0.0000538 99.07	0.00451 0.000268 99.32%	0.00559 -0.0000362 96.91

In Table 8-3 it is clear that for all covariate models matching has successfully reduced the mean difference in propensity score. Regardless of matching method used the percentage balance improvement is very high. The consequence of propensity score

matching on outcome analysis at day 90 was investigated using type I error rates given in Table 8-4.

Table 8-4: Type I error rate for each method and model being evaluated. Outcomes analysed are mRS day 90 and ordinal NIHSS day 90. Case and control groups sampled from the same population.

Type I error rates					
Covariate model	**Scale used for outcome assessment**	**Full matching**	**Nearest matching**	**Optimal matching**	**Genetic matching**
1	mRS 90 NIH 90	12.1% 0.5%	2.1% 0.6%	0.4% 0.8%	0.9% 0.9%
2	mRS 90 NIH 90	11.5% 0.5%	0.9% 0.9%	0.3% 1.1%	0.1% 4.3%
3	mRS 90 NIH 90	12.4% 0.8%	0.7% 1.6%	0.4% 0.9%	1.3% 1.9%
4	mRS 90 NIH 90	8.1% 0.5%	0.9% 0.9%	0.1% 0.9%	0.9% 2.1%
5	mRS 90 NIH 90	9.2% 0.7%	1.4% 1.7%	0.3% 0.4%	0.8% 2.2%
6	mRS 90 NIH 90	9.3% 0.8%	1.7% 1.7%	0.2% 0.8%	1.6% 2.9%

Comparing the type I error rates given in Table 8-4 the full matching method performs poorly in comparison to other methods for all covariate models. This is highlighted in Figure 8-4. If matching was successful, we would anticipate the OR for treatment effect to be close to 1. For all matching methods excluding full matching the odds ratios cluster around 1.

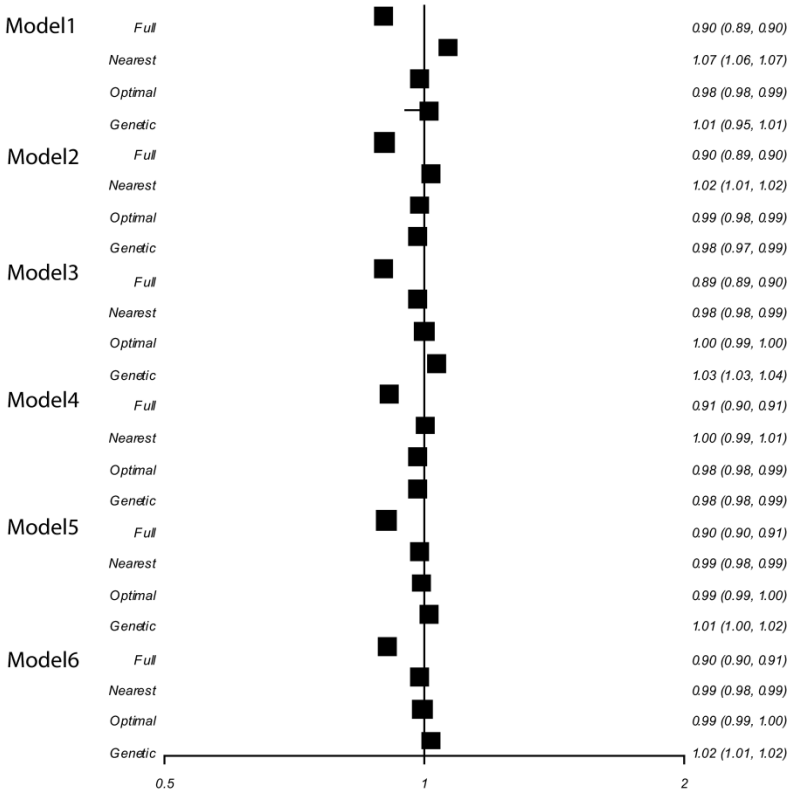

Figure 8-4: Forest plot illustrating the odds ratio for treatment effect measured by full scale mRS day 90. Mean OR and 95% CI given for each model with each matching method. Case and control groups sampled from the same population.

This analysis was repeated using the entire pool of treated subjects. Results are given in Appendix B. Due to the extensive computational time taken to run, the genetic matching method was not used for any further simulations.

8.3.2.2. Matching from a different population

By matching an untreated control group to an untreated case group taken from a different population it was hypothesised that the difference between the two groups would be larger compared to the simulations above. Table 8-5 illustrates the difference in propensity score before and after matching. This is given for both the original data restriction and the further restriction applied. Due to data availability only covariate models 1, 2 and 3 could be used here.

Table 8-5: Difference in propensity score between groups before and after matching along with percentage balance improvement between groups after matching. Case and control groups sampled from different populations. Results are given for initial inclusion criteria and after further restriction were applied.

Distance information for each model			
Covariate model and matching method	**Difference in propensity score between groups**	**Initial inclusion criteria applied**	**Further inclusion criteria applied**
1 Full matching	Mean before Mean after % Balance improvement	0.0210 -0.0000718 99.68	0.029059 -0.0000683 99.78%
1 Nearest matching	Mean before Mean after % Balance improvement	0.0210 0.0000831 99.57	0.029059 0.00009970 99.67%
1 Optimal matching	Mean before Mean after % Balance improvement	0.0210 -0.0000346 99.81	0.029059 0.00000453 99.89%
2 Full matching	Mean before Mean after % Balance improvement	0.0797 0.000897 99.15%	0.2299 0.0026094 98.92%
2 Nearest matching	Mean before Mean after % Balance improvement	0.0797 0.0308 61.45	0.229933 0.101947 55.70%
2 Optimal matching	Mean before Mean after % Balance improvement	0.0797 0.0307 61.47	0.229933 0.1019414 55.70%
3 Full matching	Mean before Mean after % Balance improvement	0.190 0.000703 99.68	0.3066 0.002358 99.27%
3 Nearest matching	Mean before Mean after % Balance improvement	0.190 0.0139 92.84	0.30666 0.12546 59.11%
3 Optimal matching	Mean before Mean after % Balance improvement	0.190 0.0115 94.16	0.30659 0.125443 59.11%

This was investigated further by analysing treatment effect. The type I error rates are displayed in Table 8-6 .

Table 8-6: Type I error rate for each method and model being evaluated. Outcomes analysed are mRS day 90 and ordinal NIHSS day 90. Case and control groups sampled from different populations. Results are given for initial inclusion criteria and after further restriction were applied.

Type I error rates			
Covariate model and matching method	**Scale used for outcome assessment**	**Initial inclusion criteria applied**	**Further inclusion criteria applied**
1 Full matching	mRS 90 NIH 90	89.6% 0.1%%	8.4% 0%
1 Nearest matching	mRS 90 NIH 90	65.3% 0.3%	28% 0.4%
1 Optimal matching	mRS 90 NIH 90	50.7% 0.1%	2.6% 0.4%
2 Full matching	mRS 90 NIH 90	85.3% 0%	0% 0%
2 Nearest matching	mRS 90 NIH 90	23.8% 1.9%	0% 13.2%
2 Optimal matching	mRS 90 NIH 90	20.6% 0.8%	0% 10.8%
3 Full matching	mRS 90 NIH 90	93.0% 0%	13% 0%
3 Nearest matching	mRS 90 NIH 90	72.5% 0.2%	0.2% 1.5%
3 Optimal matching	mRS 90 NIH 90	67.1% 0.2%	0.2% 2.2%

With the initial inclusion criteria applied the type I error rates are very high when analysing outcome as mRS day 90. When further restriction is applied to OTT these error rates reduce dramatically.

The plots in Figure 8-5 illustrate the odds ratios for treatment effect with initial restriction (top) and further restriction (bottom). Further restriction results in odds ratios closer to 1 than initial restriction but in both cases there is substantial deviation from 1.

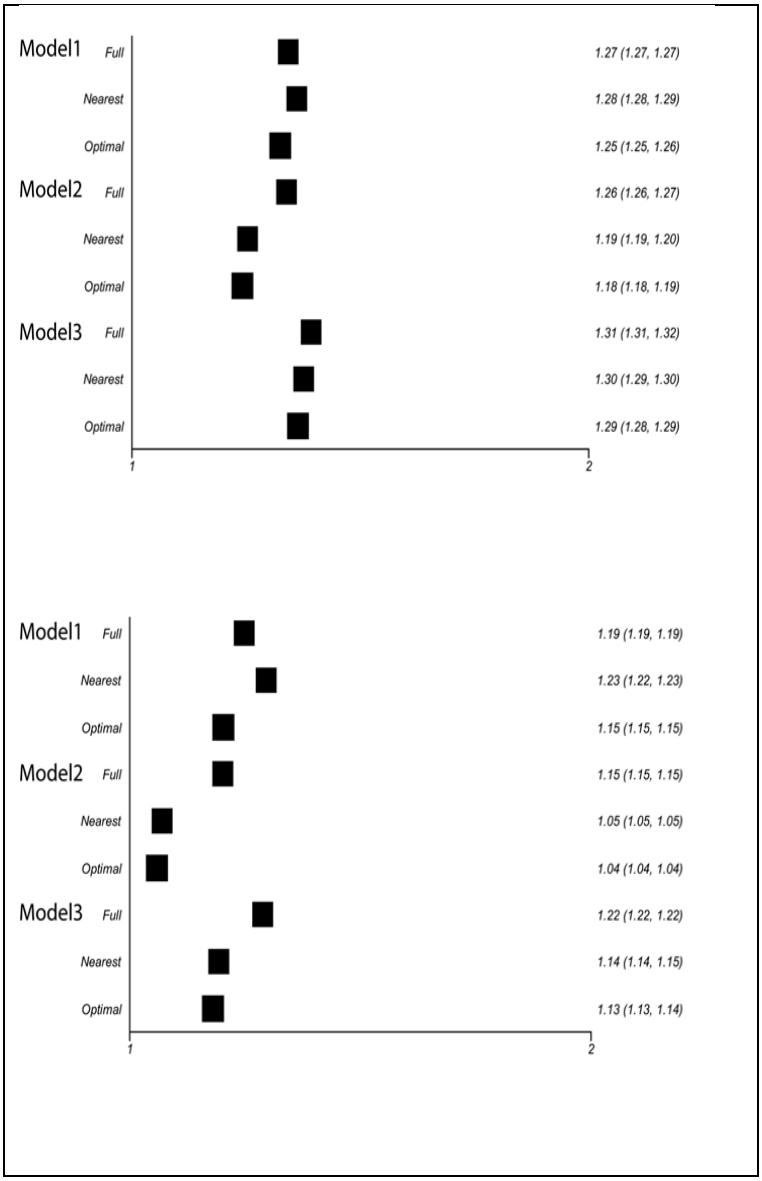

Figure 8-5: Forest plot illustrating the odds ratio for treatment effect measured by full scale mRS day 90. Mean OR and 95% CI given for each model with each matching method. Case and control groups sampled from different populations. Plots are given for initial inclusion criteria (top) and after further restriction were applied (bottom).

8.3.2.3. Matching controls to a treatment population

The aim of these simulations was to investigate the variability that occurs when sampling non-randomised subjects treated with thrombolysis as the case group and generating a

matched control group. The difference in propensity score between the two groups before and after matching is given in Table 8-7.

Table 8-7: Difference in propensity score between groups before and after matching along with percentage balance improvement between groups after matching. Control subjects matched to a group treated as standard of care.

Distance information for each model				
Covariate model	**Difference in propensity score between groups**	**Full matching**	**Nearest matching**	**Optimal matching**
1	Mean before Mean after % Balance improvement	0.035921 0.00000765 99.90%	0.035921 0.0003426 99.07%	0.035921 0.0001357 99.61%
2	Mean before Mean after % Balance improvement	0.519092 0.0002227 99.96%	0.519092 0.313179 39.71%	0.519092 0.313791 39.71%
3	Mean before Mean after % Balance improvement	0.587363 0.000440 99.93%	0.587363 0.401051 31.76%	0.587363 0.4010522 31.76%

When more variables are added to the model the difference in propensity score between the two groups becomes substantially higher. For these simulations the full matching method has reduced the mean difference in propensity score to a larger extent than any other method for all covariate models. Treatment effect was analysed and the type II error rates are given in Table 8-8.

Table 8-8: Type II error rates for each method and model being evaluated. Outcomes analysed are mRS day 90 and ordinal NIHSS day 90. Untreated subjects matched to nonrandomised subjects treated as standard of care.

Type II error rates				
Covariate model	**Scale used for outcome assessment**	**Full matching**	**Nearest matching**	**Optimal matching**
1	mRS 90 NIH 90	0% 0%	0% 0.7%	0% 0.4%
2	mRS 90 NIH 90	23.4 31.8	73.0% 80.4%	73.4% 80.4%
3	mRS 90 NIH 90	56.6% 58.1%	88.4% 92.9%	88.0% 92.9%

Adding more covariates into the model substantially increases the type II error rates. The full matching method has the lowest error rates in comparison to other matching methods. Figure 8-6 illustrates the odds ratios for treatment effect for all methods of matching with each covariate model used.

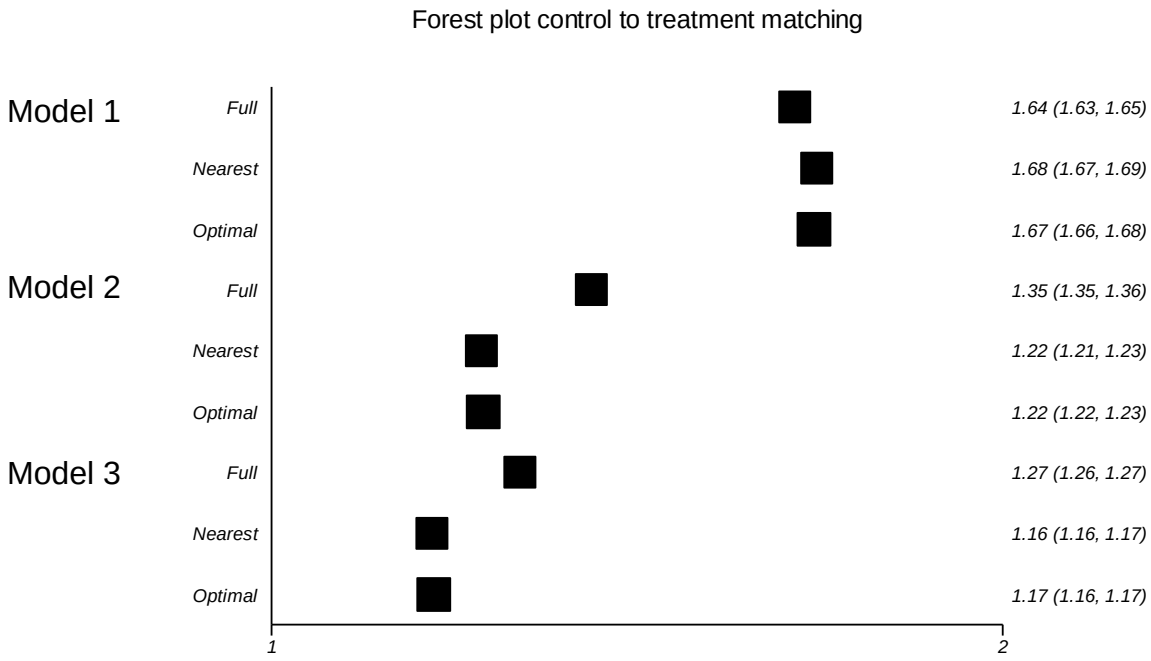

Figure 8-6: Forest plot illustrating the odds ratio for treatment effect measured by full scale mRS day 90. Mean OR and 95% CI given for each model with each matching method. Untreated subjects matched to nonrandomised subjects treated as standard of care.

8.4. Discussion

The analysis presented here provides a within-VISTA investigation of propensity score matching using the MatchIt package in R. Different covariate models have been considered for the matching process. These models include the variables most often associated with outcome, age, baseline NIHSS and onset time to treatment[171 178].

It is important to note that matching based on propensity score attempts to balance on all observed covariates added into the matching model. This is in contrast to an RCT which should balance on all variables both observed and unobserved. Thus propensity score matching may not necessarily eliminate all systematic differences between the groups as it does not balance the unobserved variables[326]. This imbalance can ultimately result in a biased estimate of the treatment effect[266].

From the initial simulations it appears that matching based on all covariate models across all methods substantially reduces the mean difference in propensity score (Table 8-3). With the exception of the full matching method the type I error rates are reassuringly low, falling below 5% (Table 8-4). Similarly the mean odds ratios for treatment effect all centre around 1 with the exception of full matching. These results suggest that very little error occurs by chance by matching alone, regardless of covariate model used. The deviation of the full matching method from other methods could possibly be due to the nature of the full matching method. This method allows for a larger matched group than the other methods which have a fixed case to control ratio.

When matching subjects originating from two different trial populations, the error rates decrease substantially when further restrictions were applied to the data before matching (Table 8-6). This illustrates the importance of placing restrictions upon the data before matching to ensure similarity of the two populations.

The large type II error rates generated when matching untreated subjects to treated subjects (Table 8-8) from neuroprotectant trials may be due to the conditional independence assumption (also known as ignorability) being violated. This assumption states, given a set of observable covariates which are not affected by treatment, potential outcomes are independent of treatment assignment[327]. In this case because treatment was given as standard of care the exact timing of the other covariates is uncertain. It may be the case that baseline NIHSS and blood pressure measurements etc. were taken after thrombolytic therapy was administered. This could result in the baseline covariates being misrepresented.

The methods used for analysis of treatment effect within a matched study are a subject of much debate throughout the literature[266 328-330]. It is often the case that matched pairs analysis is used in order to account for the matching itself[266 331]. However, if ignorability holds and matching creates adequate balance between groups there is evidence to support using techniques similar to those in an RCT setting[328-329 332-333]. The use of matched pair methods such as conditional logistic regression may result in loss of power, particularly when matching on the estimated liner propensity score has not created close pairwise matches[328]. This could be investigated further by repeating the simulations and analysing treatment effect allowing for the matched pairs design using methods such as generalised estimating equations.

There are several limitations to this study making it difficult to draw any definitive conclusions about matching in RCTs. Many of the results generated are spurious, in particular the incredibly low type I error rates given when matching control subjects within the same population (Table 8-4) in comparison to the large type I error rates obtained from matching different populations (Table 8-6). As mentioned above, this could partly be due to the matching model used. Employing unnecessary variables in the matching process may lead to inefficiency in the analysis rather than benefit from matching. However the large degree of inconsistency is more likely to be a result of the analytical method employed.

As previously discussed there is a great deal of debate in the literature regarding the correct method of analysis of matched data. The aim of this study was to assess the success of matching by evaluating if matching increases the power or precision of the analysis. The analytical method employed assumes that matching has created an adequate balance between groups, if unbalanced this method may be sub-optimal. For

this analysis, methods that account for matched pairs design such as logistic regression with GEE methods may have been more appropriate, leading to more consistent results. These methods would account for the lack of independence in the data and allow us to observe the true conditional treatment effect in the matched sample. This would be expected to generate more accurate estimates of the type I and type II error rates for assessing the success of matching.

We obtained our data from a single non-randomised data source. Further investigation is needed matching a historical control population to an external RCT control population and matching non-randomised subjects treated with thrombolysis to a true RCT thrombolysis group. The matched trial population can then be analysed and the results compared to those from the true population.

In summary, the high error rates observed coupled with high percentage balance improvement suggest substantial further work is needed to assess the applicability of historical controls for use as a matched control group in future studies. Considerations need to be made into the model used for matching, the efficiency of the matching method used, the comparability of the two populations being matched and the analytical approach to investigate treatment effect.

Chapter 9
Cluster Trials and Reliability of Multicentre Stroke Trials within VISTA

9.1. Background

While randomised control trials are the 'gold standard' in stroke research, cluster randomised trials, which randomise patients by groups, are becoming a more widely used approach. When evaluating strategies to promote the transfer of research findings into clinical practice, i.e. in "Implementation Research", a cluster randomised trial design is of advantage.

However, compared to standard randomised control trials, cluster randomised trials decrease statistical efficiency requiring more patients to obtain equivalent statistical power[228]. Observations on individuals in the same cluster tend to be correlated and this reduces the effective sample size[334].

Cluster trials have to be inflated by a factor called the design effect which depends on the average cluster size and the degree of correlation within clusters (intracluster correlation coefficient or ICC). ICCs are calculated as the between-cluster variance divided by the sum of the within-cluster and between-cluster variance. Thus, they refer to the proportion of variance that can be attributed to the cluster-level.[335]

Knowledge of the ICCs of specific outcome measures within specific trial settings is essential to compute the desired sample size for cluster trials. Up to now there is limited information on the ICCs in stroke trials and quality of reporting of those few cluster trials in the literature has been poor[336].

A literature search was performed by a colleague Dr Benedikt Frank and only four articles presenting the results of a cluster randomised trial in secondary (i.e. hospital) stroke care including outcome variables could be identified[337-340]. Three of the trials missed their primary endpoint. Two of the latter underestimated their ICC in the planning phase and thereby underpowered their study[337 339].

Within the Virtual International Stroke Trials Archive (VISTA)[204], we have access to datasets from numerous randomised multicentre stroke trials that made careful recordings of relevant baseline variables and of outcome measures at 90 days after stroke onset.

This gives us the opportunity to compute in our data the ICCs for specific baseline and outcome measures and thereby to support the planning of future cluster trials in the field of stroke research. With reliable estimations of ICCs, the risk of under- or over-powering of studies should be reduced. In addition, we report the ICCs of different levels of clustering within VISTA (trial, continent, and year of enrolment) to provide a measure for the reliability of this trial archive.

9.2. Methods

9.2.1. Data source and patients

We gathered demographics, clinical data and functional outcome measures from trials in ischaemic stroke conducted in the period from 1992 to 2006. We obtained our data, anonymised in relation to patients, study-centres and trials, from VISTA. We included placebo patients from thrombolysis trials as well as all patients from neuroprotectant trials involving any drug now known not to influence outcome after stroke. We excluded patients for whom we lacked information about age, baseline NIHSS, and thrombolysis administration as standard of care.

9.2.2. Statistical Analysis

For a cluster trial to achieve the equivalent power of a randomised control trial, the standard sample size estimate has to be inflated by the design effect, $1 + (\bar{n} - 1)\rho$ where ρ is the estimated ICC and $\bar{n}$ is the average number of observations in each cluster. This estimated ICC is the total variation in the outcome attributed to the difference between clusters[334]:

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}.$$

Where σ_b^2 is the between cluster variance and σ_w^2 is the within cluster variance.

To estimate ICCs, we analysed the data using a linear mixed model (SAS PROC MIXED) for metric variables, respectively a generalised linear mixed model (SAS PROC GLIMMIX) for binary and ordinal variables, and hereby obtained the maximum likelihood estimates for the components of variance. Detailed justification for selection of this method has previously been published[341-342].

For the purposes of this analysis ICCs are reported as unadjusted and adjusted estimates. We adjusted for age, baseline NIHSS and treatment with thrombolysis as confounders, as previously justified in detail[241 284]. For analytical purposes the NIHSS was converted to an ordinal measure, as discussed in previous chapters. Analyses were undertaken using SAS 9.2.

9.2.3. Cluster levels

The reported levels of clustering were study-centres within each trial, as well as trial, continent, and year of enrolment within VISTA. Continents were defined as North America, South America, Europe, Africa, Asia, and Australia/Oceania. In 657/11841 patients the information on the exact year of enrolment was missing and the respective start year of the trial was used.

9.3. Results

We computed the ICCs of every trial separately, using centre as the level of clustering and present the median, minimum and maximum ICC for specified baseline and follow-up data in Table 9-1 with n representing the total number and $\bar{n}$ the average number. The ICCs for mRS and NIHSS at different time points using centre as the level of clustering are reported in Table 9-2. We restricted this sub-analysis to patients with outcomes recorded at every time point. The ICCs across VISTA using trial, continent and year of enrolment as the level of clustering are presented in Table 9-3 to Table 9-5.

Table 9-1: ICC's with centre as level of clustering. Presented as the median ICC for centre across all trials within the VISTA dataset (n) containing the measurement.

	Trials (n)	Cluster ($\bar{n}$)	Patients ($\bar{n}$)	Cluster Size ($\bar{n}$)	Unadjusted ICC Median (Min-Max)	Adjusted ICC Median (Min-Max)
Baseline Data						
Age	9	102.2	731.4	8.6	0.040 (<.001-0.106)	0.038 (<.001-0.102)
thrombolysis	3	112.7	1003.0	8.9	0.554 (0.251-0.615)	0.592 (0.244-0.604)
Baseline NIHSS	9	102.2	731.4	8.6	0.106 (<.001-0.136)	0.104 (<.001-0.146)
Glucose	6	88.0	529.2	8.3	0.008 (<.001-0.071)	0.006 (<.001-0.060)
Diastolic BP	7	109.0	831.7	9.6	0.061 (<.001-0.089)	0.062 (<.001-0.083)
Systolic BP	7	109.0	832.0	9.7	0.031 (<.001-0.059)	0.026 (<.001-0.053)
Outcome Data						
mRS day 90 (ordinal)	8	112.4	753.0	6.5	0.025 (<.001-0.057)	0.007 (<.001-0.039)
mRS day 90 0-1 (Dichotomised)	8	112.4	753.0	6.5	0.058 (0.013-0.113)	0.031 (0.015-0.123)
mRS day 90 0-2 (Dichotomised)	8	112.4	753.0	6.5	0.054 (<.001-0.085)	0.019 (<.001-0.179)
NIHSS day 90 (Ordinal)	8	112.4	753.0	6.5	0.023 (<.001-0.074)	0.029 (0.008-0.085)
NIHSS day 90 0-1 (Dichotomised)	8	112.4	753.0	6.5	0.015 (<.001-0.061)	0.012 (<.001-0.078)
BI day 90 >95 (Dichotomised)	8	112.4	753.0	6.5	<.001 (<.001-0.051)	0.008 (<.001-0.062)
Mortality day 90	8	102.8	687.0	8.3	0.014 (<.001-0.082)	0.010 (<.001-0.024)
Glucose 24-48h	8	102.8	687.0	8.3	0.013 (<.001-0.097)	0.007 (<.001-0.058)
Diastolic BP 4-7 days	9	101.4	711.8	8.5	0.027 (<.001-0.112)	0.030 (<.001-0.110)
Systolic BP 4-7 days	9	101.4	711.8	8.5	0.013 (<.001-0.056)	0.015 (<.001-0.069)

Table 9-2: ICC's for outcomes at other time-points with centre as level of clustering. Presented as the median ICC for centre across all trials within the VISTA dataset (n) containing the measurement. Data for each outcome is restricted to patients with recordings for all time-points of each outcome measure.

	Trials (n)	Cluster ($\overline{n}$)	Patients ($\overline{n}$)	Cluster Size ($\overline{n}$)	Unadjusted ICC Median (Min-Max)	Adjusted ICC Median (Min-Max)
mRS day 90 (ordinal)	5	117.4	826.6	6.5	0.027 (<.001-0.084)	0.015 (0.004-0.030)
mRS day 30 (ordinal)	5	117.4	826.6	6.5	0.017 (<.001-0.058)	0.016 (<.001-0.037)
NIHSS baseline (Ordinal)	6	103.0	711.0	9.0	0.080 (<.001-0.163)	0.079 (<.001-0.161)
NIHSS day 30 (Ordinal)	6	103.0	711.0	9.0	0.007 (<.001-0.064)	0.013 (<.001-0.032)
NIHSS day 90 (Ordinal)	6	103.0	711.0	9.0	0.006 (<.001-0.044)	0.007 (<.001-0.041)

Table 9-3: ICC's treating each anonymised trial as a cluster.

	Cluster (n)	Patients (n)	Cluster Size ($\bar{n}$)	unadjusted ICC	Adjusted ICC
Baseline data					
Age	11	11841	1076.5	0.036	0.043
thrombolysis	4	7955	1988.8	0.123	0.112
Baseline NIHSS	11	11841	1076.5	0.044	0.051
Glucose	7	7888	1126.9	<.001	0.001
Diastolic BP	9	11060	1228.9	0.023	0.019
Systolic BP	9	11062	1229.1	0.006	0.004
Outcome Data					
mRS day 90 (ordinal)	10	11154	1115.4	0.018	0.006
mRS day 90 0-1 (Dichotomised)	10	11154	1115.4	0.014	0.008
mRS day 90 0-2 (Dichotomised)	10	11154	1115.4	0.018	0.002
NIHSS day 90 (Ordinal)	10	10554	1055.4	0.022	0.003
NIHSS day 90 0-1 (Dichotomised)	10	10554	1055.4	0.027	0.005
BI day 90 >95 (Dichotomised)	11	11533	1048.5	0.021	0.003
Mortality day 90	11	11841	1076.5	0.020	0.006
Glucose 24-48h	3	5698	1899.3	0.002	<.001
Diastolic BP 4-7 days	4	5470	1367.5	0.018	0.012
Systolic BP 4-7 days	4	5470	1367.5	<.001	<.001

Table 9-4: ICC's treating each continent as a cluster.

	Cluster (n)	Patients (n)	Cluster Size ($\bar{n}$)	unadjusted ICC	Adjusted ICC
Baseline data					
Age	6	11633	1938.8	0.041	0.041
thrombolysis	6	7955	1325.8	0.173	0.188
Baseline NIHSS	6	11633	1938.8	0.013	0.009
Glucose	6	7857	1309.5	0.007	0.009
Diastolic BP	6	10875	1812.5	0.021	0.017
Systolic BP	6	10877	1812.8	0.014	0.006
Outcome Data					
mRS day 90 (ordinal)	10947	1824.5	0.004	<.001	10947
mRS day 90 0-1 (Dichotomised)	10947	1824.5	0.007	0.001	10947
mRS day 90 0-2 (Dichotomised)	10947	1824.5	0.004	<.001	10947
NIHSS day 90 (Ordinal)	6	10349	1724.8	0.005	<.001
NIHSS day 90 0-1 (Dichotomised)	6	10349	1724.8	0.004	<.001
BI day 90 >95 (Dichotomised)	6	11326	1887.7	0.002	0.007
Mortality day 90	6	11633	1938.8	0.007	0.004
Glucose 24-48h	6	5698	949.7	0.009	0.009
Diastolic BP 4-7 days	6	5470	911.7	0.024	0.020
Systolic BP 4-7 days	6	5470	911.7	0.015	0.011

Table 9-5: ICC's treating year of enrolment as a cluster.

	Cluster (n)	Patients (n)	Cluster Size ($\bar{n}$)	unadjusted ICC	Adjusted ICC
Baseline data					
Age	11841	740.1	0.046	0.059	11841
thrombolysis	7955	994.4	0.196	0.196	7955
Baseline NIHSS	11841	740.1	0.026	0.026	11841
Glucose	7888	563.4	<.001	0.001	7888
Diastolic BP	11060	850.8	0.008	0.004	11060
Systolic BP	11062	850.9	0.003	0.001	11062
Outcome Data					
mRS day 90 (ordinal)	15	11154	743.6	0.012	0.004
mRS day 90 0-1 (Dichotomised)	15	11154	743.6	0.007	0.005
mRS day 90 0-2 (Dichotomised)	15	11154	743.6	0.007	0.001
NIHSS day 90 (Ordinal)	16	10554	659.6	0.018	0.003
NIHSS day 90 0-1 (Dichotomised)	16	10554	659.6	0.012	0.004
BI day 90 >95 (Dichotomised)	16	11533	720.8	0.013	0.001
Mortality day 90	16	11841	740.1	0.021	0.007
Glucose 24-48h	9	5698	633.1	0.001	<.001
Diastolic BP 4-7 days	12	5470	455.8	0.014	0.010
Systolic BP 4-7 days	12	5470	455.8	<.001	<.001

9.4. Discussion

The analysis presented here provides a collection of ICCs for various baseline and outcome measures used in contemporary stroke trials, including the three most prevalent scales, being mRS, NIHSS, and Barthel Index[343].

ICCs are known to be strongly influenced by event rate changes, underlining the importance of the knowledge of ICCs for specific outcome measures[344]. The use of incorrect ICC estimates in the planning phase of a trial can lead to an underpowered study and unreliable results.

Alongside unadjusted ICCs, we present estimates adjusted for relevant covariates (age, thrombolysis and baseline NIHSS). Typically this will render smaller ICCs, as we confirmed in our sample, as some of the between cluster variation may be explained by the cluster level factor[345]. This should reduce sample size requirements but it is crucial that any subsequent analysis is also adjusted for these covariates. For example, a study with an average cluster size of 150 patients should be inflated by a design effect of 4.7 when using our median unadjusted ICC of 0.025 for ordinal mRS and only 2.0 when using the adjusted ICC estimation of 0.007. Additionally, the selection of outcome measures has a significant impact on required sample size: the choice of mRS as a binary measure, dichotomised between 1 and 2, would result in a design effect of 9.6 (unadjusted ICC of 0.058), respectively 5.6 (adjusted ICC of 0.031) for an assumed average cluster size of 150 patients.

We obtained our data for the analysis from trials randomising at patient level. This contrasts with prior cluster randomised trials in secondary stroke care, as our average cluster size was smaller and number of clusters larger. However a previous simulation

study, evaluating the impact of different factors on ICC estimates, showed that cluster size had only a minimal effect on the ICC and its upper 95% confidence limit[344]. The same study observed a negligible change on the ICC estimate when increasing the number of clusters, but a significant impact on confidence intervals: as the number of clusters increases, the 95% confidence intervals become tighter. Both findings support the reliability of our ICC estimates. Additionally, our estimates largely track with the ICCs from a recently published, cluster randomised trial, reporting the same binary outcomes. Their adjusted ICC for mRS at 90 days dichotomised between 2 and 3 was reported to be 0.018, versus a median of 0.019 in our analysis. The ICCs for Barthel Index at 90 days dichotomised between 90 and 95 were lower in the recently published cluster randomised trial with 0.015 versus our finding of 0.030.

By calculating the ICCs for various levels of clustering within VISTA, we analysed the reliability of this trials archive. The adjusted ICCs for the most commonly used outcome measures (mRS, NIHSS, and Barthel Index) were all below 0.01. Thus, trial, continent, and year of enrolment explained less than 1% of the variability of follow-up data within VISTA. In contrast, the majority of median adjusted ICCs calculated within each trial using centre as level of clustering were above 0.01.

In summary, the low contribution of trials, year or continent of enrolment to overall variation in outcome offers reassurance that analyses using pooled data from multiple trials in VISTA are unlikely to suffer from bias from these sources. We present potentially valuable and reliable estimates of ICCs for specific baseline and follow-up data to support future sample size calculations for cluster randomised trials.

Chapter 10
Discussion and conclusions

This thesis describes a compendium of projects assessing the impact of applying different methods of design, inclusion and outcome measurement to limit sample size and strengthen analysis in clinical trials in acute stroke.

First some simple inclusion criteria were investigated (Chapter 3). When considering subjects for inclusion into a trial several factors need to be taken into consideration. It is important to exclude those in whom, given treatment, potential harm could outweigh any benefit gained. Inclusion criteria considered here were the combined effect of age and onset time to treatment (OTT) and history of atrial Fibrillation (AF) after treatment with thrombolysis.

The relationship between OTT and age could be investigated by assessing how the effect of thrombolysis changes over onset time to treatment (OTT). By looking across the entire range of OTT and assessing the interaction between the two covariates this provided complementary data to a previous VISTA analysis conducted by Mishra et al[63]. In a non-randomised VISTA comparison, it was found that across the full range of OTT, up to 3.5h, the treatment effect of thrombolysis in very elderly stroke patients (>80 years old) was comparable to that of their younger counterparts.

AF has been considered a risk factor for poor outcome from acute stroke and may influence response to thrombolytic therapy. Due to the potential confounding with age and stroke severity, supporting evidence for this is limited. The association of AF and

modified Rankin Scale (mRS) at day 90 was assessed in a non-randomised VISTA analysis. Multiple logistic regression analysis adjusted for age and baseline National Institutes of Health Stroke Scale (NIHSS) showed that history of AF had no independent impact on stroke outcome. Compared to untreated comparators, the magnitude of the outcome following thrombolytic therapy was comparatively equal whether in presence or absence of AF.

These data lend support to the use of thrombolysis across all age groups of stroke patients with and without history of AF within approximately 3.5h of stroke onset. AF appears to be a marker for high age and baseline NIHSS rather than an independent risk factor for poor outcome post stroke.

For therapies such as thrombolysis, treatment must be initiated in the hyperacute period post stroke. The large sample size required for such trials may result from natural population heterogeneity and variation in baseline NIHSS collected when patients are unstable. For neurorestorative treatments that may be initiated later after stroke onset, e.g. at 24h, a more homogeneous population with more predictable outcome may be available. In Chapter 4 it was postulated that deferred selection would permit more powerful and thus smaller trials.

The relationship between hyperacute vs. 24h NIHSS recordings and 90 day mRS was examined in subjects from VISTA. A simulation study was then performed to model the sample size required to detect a 'shift' in mRS outcome equivalent to a 5% absolute difference in proportion achieving mRS 0-2 versus 3-6, setting power at 80% and assuming adjustment for entry age and NIHSS. Two restrictions of NIHSS were investigated 4-20 and 7-20.

It was found that extending the time window for patient selection provides a measurement which has a stronger more predictive relationship with outcome. This subsequently allows a larger population to benefit from treatment. Such extension of the time window must be balanced against any anticipated decay in biological effect of the treatment with increasing delay from stroke onset. However, with recent interest in neurorestorative treatments, as discussed by Steven Cramer[288], any strategy that can limit the cost of proof of concept trials is desirable.

Due to the simulated nature of the treatment effect no definitive conclusions can be made, only recommendations. Validation of these results on a randomised control trial would be desirable however no applicable data are available from current stroke trials.

Trial inclusion was explored further in Chapter 5 by investigating selection for delayed treatment with thrombolysis based on a prognostic score. Pooled analysis of RCT suggests that additional patients could benefit but others may be harmed with initiation of thrombolysis beyond 4.5 hours after stroke onset. Prognostic scoring methods were proposed to identify a strategy for patient selection to be applied first to an existing trial dataset and then validated in the pooled RCT 4.5-6h data.

500 patients treated with thrombolysis and 500 controls from VISTA were selected, matching Rankin (mRS) outcomes to those from pooled RCT 4.5-6h data. We ranked patients by prognostic score (from age and NIHSS). Prognostic score limits were chosen to optimise the sample for a net treatment benefit significant at p=0.01 by Cochran Mantel Haenszel test and by ordinal logistic regression. More inclusive limits were also defined based on p=0.05 criteria. We iteratively chose lower and upper score limits to exclude patients with extreme predicted outcomes. After finalising prognostic score limits, for

validation they were applied by an independent statistician to the pooled RCT data for 4.5-6h. All analyses were adjusted for age and NIHSS.

While the validation analysis based on ordinal outcomes failed to deliver a population in whom treatment >4.5h was safe and effective, analysis based on net benefit (mRS 0-1) showed significance. The analysis of trial data according to net benefit has proponents and opponents, the latter arguing that it may conceal useful treatment effects among subpopulations. Unless we can prospectively select patients for these sub-populations, the ordinal approach to interpretation may remain optimal as it better reflects the true outcome of clinical practice.

Clinical trials for acute ischaemic stroke treatment require large numbers of participants and are expensive to conduct. Methods that enhance statistical power are therefore desirable. In the past some trialists have investigated the use of earlier endpoints on single trial datasets[230] and taken advantage of the fact that numerous outcome scales are available to measure various domains of neurological and functional recovery[96 200].

Clinical trials in stroke typically measure outcome after 90 days. Chapter 6 provides an exploration and validation analysis investigating the effect on power when using earlier outcome assessment.

First a within-VISTA exploration study was conducted to compare the sensitivity of four outcome measures (mRS at 30 and 90 days, and NIHSS at 7 and 90 days, analysed as ordinal measures) to the established treatment effect of thrombolysis. This was a non-randomised comparison using a multiple re-sampling approach, day 7 NIHSS was found to be the most sensitive endpoint. Dichotomised analyses supported these results. However this needed validation in a randomised trial dataset for use in exploratory stroke trials.

The validation stage included data from patients who had been enrolled and treated within 270 minutes of stroke onset, in any of 8 published randomised trials[171]. A simulation approach was performed following a predefined analysis plan and was conducted by an independent statistician.

The validation study reinforced the results from the non-randomised VISTA study. Both found day 7 NIHSS score offers statistical advantages as an endpoint for the early, exploratory testing of novel agents. Detecting an early signal of treatment benefit or futility in acute stroke trials is economically, scientifically and ethically desirable.

In Chapter 7 it was hypothesised that the apparent treatment effect of the proven therapy thrombolysis would be detected more strongly if the outcome measure uses a combination of early and late measures (e.g. 7 day NIHSS combined with 90-day mRS) than either early or late measure alone.

Global outcome measures were generated using combinations of typical outcome scales at different time-points. A simulation approach was undertaken to assess relations between sample size and power for ordinal scales and corresponding global outcomes. A simple incremental ranking method was also investigated for the combination of endpoints.

The global test combination of NIHSS90 with NIHSS7 appears to offer incremental sensitivity to treatment effect compared to the ordinal forms of these scales alone. The combination of mRS90 with NIHSS7 did not increase the sensitivity to treatment effect when compared to NIHSS alone, but offers a broader clinical measure without loss of statistical power. The use of a more simple incremental approach combining the scales using a ranking method offered no benefit in comparison to the ordinal scales.

When data concerning both early and late outcomes are combined into a global measure there is an increased sensitivity to treatment effect compared to solitary ordinal scales. This delivers a 20% reduction in required sample size at 80% power. Combining early with late outcomes merits further consideration and requires validation on external RCT data.

If a placebo control is deemed impractical, researchers consider comparisons against historical controls. Abandoning the rigour of the blinded RCT carries substantial penalty in loss of reliability and should not be undertaken lightly. However, there may be strategies that can strengthen these compromises when considered unavoidable. In Chapter 8 a within-VISTA exploration of case-control matching is presented.

The results indicate that caution must be taken when using historical controls to generate a matched control group. Looking at the equivalence of two trial populations highlights the importance of model selection and the application of inclusion criteria before matching. Substantial further work matching to external data and validation to RCT data is needed.

Chapter 9 considers some elements in the design of cluster randomised trials. Reliable estimates of intracluster correlation coefficients (ICCs) for specific outcome measures are crucial for sample size calculations of future cluster randomised trials. ICCs indicate the proportion of data variability that is explained by defined levels of clustering.

ICCs were estimated from linear and generalised linear mixed models using maximum likelihood estimation for common measures used in stroke research, including modified Rankin Skale (mRs), National Institutes of Health Stroke Scale (NIHSS) and Barthel Index (BI).

These estimates of relevant ICCs should assist trial planning. For example the sample size for a cluster trial with 150 patients per centre using ordinal analysis of mRS should be inflated by 2.0 due to the ICC of 0.007; whereas the ICC of 0.031 using mRS dichotomised above mRS 0-1, requires inflation by 5.6. The low contribution of trials, year or continent of enrolment to overall variation in outcome offers reassurance that analyses using pooled data from multiple trials in VISTA are unlikely to suffer from bias from these sources.

While this thesis has concentrated on clinical trials in the area of acute stroke, some methods are applicable across different areas of medical research. The use of a global test statistic discussed in Chapter 7 is not unique to stroke research. Similar to the work presented in this thesis, a global test was used in the re-analysis of clinical trials in Parkinson's disease[346] and found to be a more powerful and clinically advantageous alternative to analysis of multiple individual outcomes, each assessing a single dimension of recovery. A similar study was performed evaluating the use of the global test in rheumatoid arthritis trials[347]. The global test statistic has also been utilised in other areas of medical research such as rehabilitation[348], environmental health[306], multiple sclerosis[349], asthma[350] and genetics[351-353].

Similarly, cluster randomised trials are becoming increasingly popular in medical research, particularly for non-pharmaceutical interventions[354-355] and primary care[356-358]. Cluster randomised trials are an attractive alternative when an RCT is not practical. For example when assessing organisational changes[359], implementing an educational programme[360-361] or assessing screening methods[362] it would be impractical to randomise at the individual patient level. The ICC's presented in this thesis are only applicable to a similar population of stroke patients. However, the development and expansion of databases

such as the General Practitioners Research Database[363] (GPRD) and the Safe Haven database[364] present the opportunity to calculate ICC's for future UK based cluster trials investigating certain aspects of primary care. This could be expanded further into medical interventions by using medical databases for ICC calculation such as: the Society of Thoracic Surgeons (STS) national database for cardiothoracic surgery, the Stanford Cancer Centre Research Database (SCCRD)[365], the UK based National Cancer Research Institute (NRCI) cancer research database[366] and the Virtual International Cardiovascular and Cognitive Trials archive (VICCTA) currently under development.

In conclusion, this research has shown that there are several areas in the design of clinical trials of acute stroke that merit further investigation. Several strategies have been highlighted that could potentially reduce sample size whilst retaining optimal levels of statistical power. Most analyses presented were performed retrospectively using non-randomised historical data. Without validation within a prospective cohort the applicability of some of the results to current clinical practice may be limited. For example, when validated on external RCT data, the prognostic score limits found in Chapter 5 failed to find a safe treatment population. Without access to quality imaging data any further investigation would likely be futile. Similarly, substantial work is needed into the investigation of the use of a matched control group using historical controls. The use of historical controls also brings into question the applicability of historical data in current practice. Prospective controls would likely be a better, more reliable matched population.

The baseline severity measurement most commonly used in acute stroke trials was found to be sub-optimal in comparison to a later measure. Substantial decreases in required sample size were observed by changing the time-point for baseline severity measurement

from hyperacute (<6h) to 24h. Lack of validation would make this difficult to implement in a new RCT. However there is a clinical premise for the use of later baseline measurements if the nature of the drug under study allows it. It has been shown that systolic BP, diastolic BP and serum glucose level fall within the first 24h post stroke. This decrease is greater when thrombolysis treatment has been given as standard of care[367]. Similarly it was shown in Chapter 4 that 24h NIHSS is more predictive of outcome making it a better measurement for inclusion.

Translational development of treatments for acute stroke is both costly and challenging. The results shown here have important implications for the design of future trials in acute stroke. Some methods presented here could be implemented into a new RCT without additional cost. The most common outcome scale used, mRS day 90[127], lacked statistical power when compared to NIHSS day 7. This suggests implementing an earlier endpoint may be more optimal to detect futility in a trial with the conventional 90 day mRS retained as the primary efficacy endpoint, at least for reperfusion strategies. Alternatively, a global endpoint could be used incorporating both mRS day 90 and NIHSS day 7 as the primary efficacy endpoint. This could potentially reduce the cost of a trial by decreasing required sample size for an equivalent level of statistical power. When considering a cluster randomised trial as an alternative to an RCT, reliable estimates of ICC's for sample size calculation could increase statistical efficiency. However other aspects such as patient selection and the nature of the intervention under study can affect trial cost and statistical power and need to be taken under consideration.

Appendix A

Appendix for Chapter 3

Table 10-1: Details of the fitted model for all data looking at full scale mRS day 90. Coefficient and standard errors given for each reported interaction.

All data Outcome mRS90	Interaction OTT and thrombolysis	Interaction OTT and age	Interaction age and thrombolysis	Interaction OTT and thrombolysis after removing all other non-significant interactions	Interaction OTT, age and thrombolysis after removing all non-significant interactions
Coefficient	0.3011	-0.00178	-0.0107	0.3076	0.000877
S.E coefficient	0.1288	0.00479	0.0060	0.1276	0.00139

Table 10-2: Details of the fitted model for those ≤ 80 looking at full scale mRS day 90. Coefficient and standard errors given for each reported interaction.

≤80 Outcome mRS90	Interaction OTT and thrombolysis	Interaction OTT and age	Interaction age and thrombolysis	Interaction OTT and thrombolysis after removing all other non-significant interactions
Coefficient	0.3803	0.00889	-0.00477	0.3505
S.E coefficient	0.1425	0.00599	0.00747	0.1412

Table 10-3: Details of the fitted model for those > 80 looking at full scale mRS day 90. Coefficient and standard errors given for each reported interaction.

> 80 Outcome mRS90	Interaction OTT and thrombolysis	Interaction OTT and age	Interaction age and thrombolysis	Interaction OTT and thrombolysis after removing all other non-significant interactions
Coefficient	0.3011	0.0164	0.0859	0.2348
S.E coefficient	0.3118	0.0408	0.0499	0.3107

Table 10-4: Details of the fitted model for all data looking at mortality day 90. Coefficient and standard errors given for each reported interaction.

All data Outcome mortality 90	Interaction OTT and thrombolysis	Interaction OTT and age	Interaction age and thrombolysis	Interaction OTT and thrombolysis after removing all other non-significant interactions	Interaction OTT, age and thrombolysis after removing all non-significant interactions
Coefficient	0.5597	-0.00235	-0.0118	0.5644	0.00399
S.E coefficient	0.2076	0.00925	0.0115	0.2054	0.00230

Table 10-5: Details of the fitted model for those ≤ 80 looking at mortality day 90. Coefficient and standard errors given for each reported interaction.

≤80 Outcome mortality 90	Interaction OTT and thrombolysis	Interaction OTT and age	Interaction age and thrombolysis	Interaction OTT and thrombolysis after removing all other non-significant interactions
Coefficient	0.7107	0.0179	-0.00028	0.6737
S.E coefficient	0.2526	0.0123	0.0155	0.2509

Table 10-6: Details of the fitted model for those > 80 looking at mortality day 90. Coefficient and standard errors given for each reported interaction.

> 80 Outcome mortality 90	Interaction OTT and thrombolysis	Interaction OTT and age	Interaction age and thrombolysis	Interaction OTT and thrombolysis after removing all other non-significant interactions
Coefficient	0.4773	0.0159	0.0888	0.4230
S.E coefficient	0.3822	0.0492	0.0618	0.3817

Table 10-7: Details of the fitted model for all data, adjusting for dichotomised age, looking at mRS day 90. Coefficient and standard errors given for each reported interaction.

All data Outcome mRS 90	Interaction OTT and thrombolysis after removing all other non-significant interactions	Interaction OTT, dichotomised age and thrombolysis after removing all non-significant interactions
Coefficient	0.3871	-0.0709
S.E coefficient	0.1283	0.0666

Table 10-8: Details of the fitted model for all data adjusting for dichotomised age, looking at mortality day 90. Coefficient and standard errors given for each reported interaction.

All data Outcome mortality 90	Interaction OTT and thrombolysis after removing all other non-significant interactions	Interaction OTT, age and thrombolysis after removing all non-significant interactions
Coefficient	0.6785	-0.1065
S.E coefficient	0.2083	0.0913

Appendix B

Appendix for Chapter 8

Table 10-9: Difference in propensity score between groups before and after matching along with percentage balance improvement between groups after matching. Case and control groups sampled from the same population of treated subjects.

Distance information for each model					
Covariate model	**Difference in propensity score between groups**	**Full matching**	**Nearest matching**	**Optimal matching**	**Genetic matching**
1	Mean before Mean after % Balance improvement	0.00194 0.00000573 98.71	0.00312 0.0000950 96.95	0.00181 0.00000235 98.59	0.00138 0.0000144 99.56
2	Mean before Mean after % Balance improvement	0.00221 0.00000817 98.79	0.00342 0.000124 96.52	0.00210 0.0000206 97.10	0.00169 0.0000144 99.18
3	Mean before Mean after % Balance improvement	0.00392 -0.0000227 98.62	0.00617 0.000576 91.64	0.00304 0.000224 95.28	0.00239 0.0000811 94.62

Table 10-10: Type I error rate for each method and model being evaluated. Outcomes analysed are mRS day 90 and ordinal NIHSS day 90. Case and control groups sampled from the same population of treated subjects.

Type I error rates					
Covariate model	**Scale used for outcome assessment**	**Full matching**	**Nearest matching**	**Optimal matching**	**Genetic matching**
1	mRS 90 NIH 90	0.8% 0.2%	0.4% 5.7%	52.9% 21.2%	0.5% 0.3%
2	mRS 90 NIH 90	0.8% 0.2%	1.2% 4.7%	43.5% 12.5%	0.8% 0.5%
3	mRS 90 NIH 90	1.2% 0.2%	1.3% 8.0%	34.5% 6.2%	1.0% 0.7%

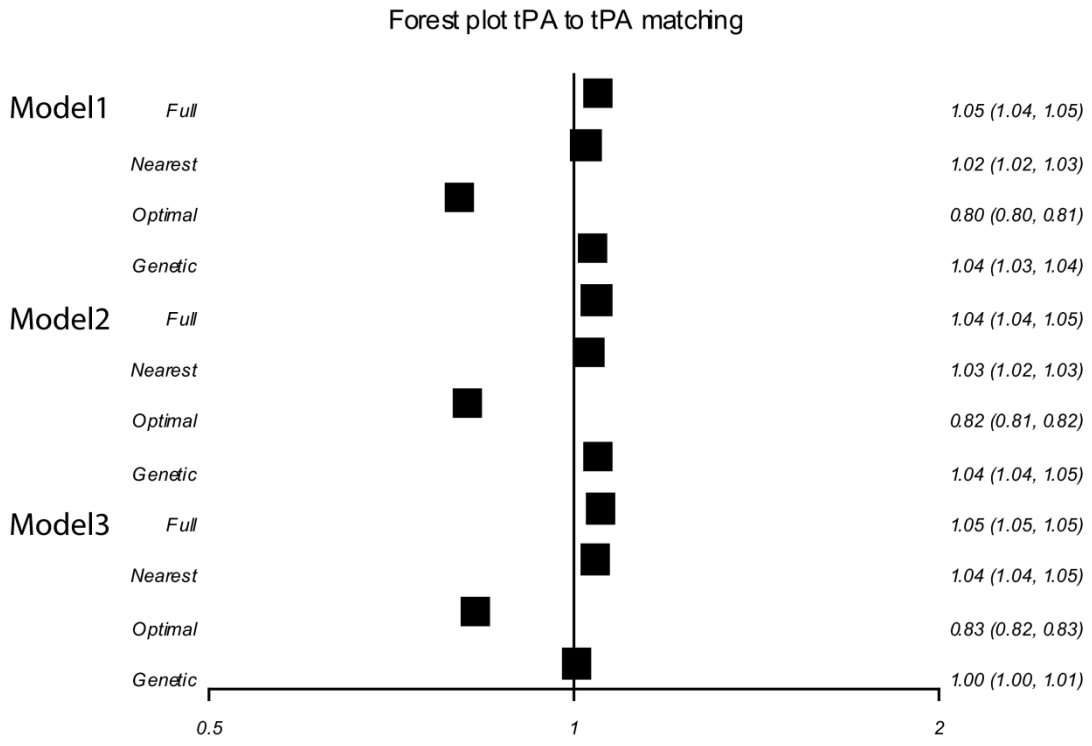

Figure 10-1: Forest plot illustrating the odds ratio for treatment effect measured by full scale mRS day 90. Mean OR and 95% CI given for each model with each matching method. Case and control groups sampled from the same population of treated subjects.

Bibliography

1. Investigators WMP. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. *J Clin Epidemiol* 1988;41(2):105-14.
2. Hankey GJ. *Stroke: Your Questions Answered*: Elsevier Health Sciences, 2007.
3. Bonita R. Epidemiology of stroke. *The Lancet* 1992;339(8789):342-44.
4. Network SIG. Management of patients with stroke or TIA: assessment, investigation, immediate management and secondary prevention. A national clinical guideline.: NHS, 2010.
5. Lindley RI. *Stroke*: Oxford University Press, USA, 2008.
6. Group GBoDSE. Methodology of the global and regional burden of stroke study. *Neuroepidemiology* 2012;38(1):30-40.
7. Rathore SS, Hinn AR, Cooper LS, Tyroler HA, Rosamond WD. Characterization of Incident Stroke Signs and Symptoms. *Stroke* 2002;33(11):2718-21.
8. Mohr JP, Caplan LR, Melski JW, Goldstein RJ, Duncan GW, Kistler JP, et al. The Harvard Cooperative Stroke Registry: a prospective registry. *Neurology* 1978;28(8):754-62.
9. Caplan LR. *Caplan's Stroke: A Clinical Approach*. 4th ed. Philadelphia: Saunders-Elsevier, 2009.
10. Brainin M, Heiss WD. *Textbook of Stroke Medicine*: Cambridge University Press, 2009.
11. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993;24(1):35-41.
12. Pittock SJ, Meldrum D, Hardiman O, Thornton J, Brennan P, Moroney JT. The Oxfordshire Community Stroke Project classification: Correlation with imaging, associated complications, and prediction of outcome in acute ischemic stroke. *Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association* 2003;12(1):1-7.
13. Wardlaw JM, Dennis MS, Lindley RI, Sellar RJ, Warlow CP. The validity of a simple clinical classification of acute ischaemic stroke. *J Neurol* 1996;243(3):274-9.
14. Kobayashi A, Wardlaw JM, Lindley RI, Lewis SC, Sandercock PAG, Czlonkowska A, et al. Oxfordshire Community Stroke Project Clinical Stroke Syndrome and Appearances of Tissue and Vascular Lesions on Pretreatment CT in Hyperacute Ischemic Stroke Among the First 510 Patients in the Third International Stroke Trial (IST-3). *Stroke* 2009;40(3):743-48.
15. Hanger HC, Wilkinson TJ, Fayez-Iskander N, Sainsbury R. The risk of recurrent stroke after intracerebral haemorrhage. *Journal of Neurology, Neurosurgery & Psychiatry* 2007;78(8):836-40.
16. Sacco S, Marini C, Toni D, Olivieri L, Carolei A. Incidence and 10-year survival of intracerebral hemorrhage in a population-based registry. *Stroke* 2009;40(2):394-9.
17. Morgenstern LB, Hemphill JC, 3rd, Anderson C, Becker K, Broderick JP, Connolly ES, Jr., et al. Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2010;41(9):2108-29.
18. van Gijn J, Rinkel GJE. Subarachnoid haemorrhage: diagnosis, causes and management. *Brain* 2001;124(2):249-78.

19. Connolly ES, Jr., Rabinstein AA, Carhuapoma JR, Derdeyn CP, Dion J, Higashida RT, et al. Guidelines for the management of aneurysmal subarachnoid hemorrhage: a guideline for healthcare professionals from the American Heart Association/american Stroke Association. *Stroke* 2012;43(6):1711-37.
20. Kim YW, Lawson MF, Hoh BL. Nonaneurysmal subarachnoid hemorrhage: an update. *Curr Atheroscler Rep* 2012;14(4):328-34.
21. Shah S. Stroke Pathophysiology. *FERNE*.
22. Herrmann M, Ehrenreich H. Brain derived proteins as markers of acute stroke: their relation to pathophysiology, outcome prediction and neuroprotective drug monitoring. *Restor Neurol Neurosci* 2003;21(3-4):177-90.
23. Strandgaard S, Paulson OB. Pathophysiology of Stroke. *Journal of Cardiovascular Pharmacology* 1990;15:S38-S42.
24. Cramer SC. Repairing the human brain after stroke. II. Restorative therapies. *Annals of Neurology* 2008;63(5):549-60.
25. Pancioli AM, Broderick J, Kothari R, Brott T, Tuchfarber A, Miller R, et al. Public perception of stroke warning signs and knowledge of potential risk factors. *JAMA* 1998;279(16):1288-92.
26. Wall HK, Beagan BM, O'Neill J, Foell KM, Boddie-Willis CL. Addressing stroke signs and symptoms through public education: the Stroke Heroes Act FAST campaign. *Prev Chronic Dis* 2008;5(2):A49.
27. Wolfe CD, Rudd T. The Burden of Stroke White Paper: Raising awareness of the global toll of stroke-related disability and death, 2007.
28. Robinson TG, Reid A, Haunton VJ, Wilson A, Naylor AR. The face arm speech test: does it encourage rapid recognition of important stroke warning symptoms? *Emergency Medicine Journal* 2012.
29. Yew KS, Cheng E. Acute stroke diagnosis. *Am Fam Physician* 2009;80(1):33-40.
30. Bonaffini N, Altieri M, Rocco A, Di Piero V. Functional neuroimaging in acute stroke. *Clin Exp Hypertens* 2002;24(7-8):647-57.
31. Muir KW, Santosh C. Imaging of acute stroke and transient ischaemic attack. *Journal of Neurology, Neurosurgery & Psychiatry* 2005;76(suppl 3):iii19-iii28.
32. Pedraza S, Puig J, Remollo S, Quiles A, Gomez1 E, Laguillo G, et al. Magnetic resonance imaging in the diagnosis of stroke. *Expert Opinion on Medical Diagnostics* 2008;2(7):843-52.
33. Feigin VL, Lawes CM, Bennett DA, Anderson CS. Stroke epidemiology: a review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. *Lancet Neurol* 2003;2(1):43-53.
34. Truelsen T, Bonita R, Jamrozik K. Surveillance of stroke: a global perspective. *International Journal of Epidemiology* 2001;30(suppl 1):S11.
35. Mukherjee D, Patil CG. Epidemiology and the global burden of stroke. *World Neurosurg* 2011;76(6 Suppl):S85-90.
36. Bonita R, Mendis S, Truelsen T, Bogousslavsky J, Toole J, Yatsu F. The Global Stroke Initiative. *The Lancet Neurology* 2004;3(7):391-93.
37. Truelsen T, Begg S, Colin M. The global burden of cerebrovascular disease, 2006.
38. Organisation TWH. The world health report 2003-shaping the future. Geneva, 2003.
39. Johnston SC, Mendis S, Mathers CD. Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. *Lancet Neurol* 2009;8(4):345-54.
40. Addo J, Ayerbe L, Mohan KM, Crichton S, Sheldenkar A, Chen R, et al. Socioeconomic status and stroke: an updated review. *Stroke* 2012;43(4):1186-91.

41. Thorvaldsen P, Asplund K, Kuulasmaa K, Rajakangas AM, Schroll M. Stroke incidence, case fatality, and mortality in the WHO MONICA project. World Health Organization Monitoring Trends and Determinants in Cardiovascular Disease. *Stroke* 1995;26(3):361-7.
42. Feigin VL, Lawes CMM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *The Lancet Neurology* 2009;8(4):355-69.
43. Bank W, editor. *World development indicators in 2004*. Washington D.C, 2004.
44. Lee S, Shafe ACE, Cowie MR. UK stroke incidence, mortality and cardiovascular risk management 1999–2008: time-trend analysis from the General Practice Research Database. *BMJ Open* 2011;1(2).
45. Carandang R, Seshadri S, Beiser A, Kelly-Hayes M, Kase CS, Kannel WB, et al. Trends in incidence, lifetime risk, severity, and 30-day mortality of stroke over the past 50 years. *JAMA* 2006;296(24):2939-46.
46. Prevalence of stroke--United States, 2006-2010. *MMWR Morb Mortal Wkly Rep* 2012;61(20):379-82.
47. Strong K, Mathers C, Bonita R. Preventing stroke: saving lives around the world. *The Lancet Neurology* 2007;6(2):182-87.
48. Organisation WH. The top 10 causes of death by broad income group, 2008.
49. Organisation WH. The global burden of disease 2004 update, 2008.
50. Feigin VL, Krishnamurthi R. Stroke prevention in the developing world. *Stroke* 2011;42(12):3655-8.
51. Di Carlo A. Human and economic burden of stroke. *Age Ageing* 2009;38(1):4-5.
52. Saka Ö, McGuire A, Wolfe C. Cost of stroke in the United Kingdom. *Age and Ageing* 2009;38(1):27-32.
53. Roger VL, Go AS, Lloyd-Jones DM, Adams RJ, Berry JD, Brown TM, et al. Heart Disease and Stroke Statistics—2011 Update / 1. About 1. About These Statistics / 2. American Heart Association's 2020 Impact Goals / 3. Cardiovascular Diseases / 4. Subclinical Atherosclerosis / 5. Coronary Heart Disease, Acute Coronary Syndrome, and Angina Pectoris / 6. Stroke (Cerebrovascular Disease) / 7. High Blood Pressure / 8. Congenital Cardiovascular Defects / 9. Cardiomyopathy and Heart Failure / 10. Other Cardiovascular Diseases / 11. Family History and Genetics / 12. Risk Factor: Smoking/Tobacco Use / 13. Risk Factor: High Blood Cholesterol and Other Lipids / 14. Risk Factor: Physical Inactivity / 15. Risk Factor: Overweight and Obesity / 16. Risk Factor: Diabetes Mellitus / 17. End-Stage Renal Disease and Chronic Kidney Disease / 18. Metabolic Syndrome / 19. Nutrition / 20. Quality of Care / 21. Medical Procedures / 22. Economic Cost of Cardiovascular Disease / 23. At-a-Glance Summary Tables / 24. Glossary. *Circulation* 2011;123(4):e18-e209.
54. Adamson J, Beswick A, Ebrahim S. Is stroke the most common cause of disability? *Journal of Stroke and Cerebrovascular Diseases* 2004;13(4):171-77.
55. Young J, Forster A. Review of stroke rehabilitation. *BMJ* 2007;334(7584):86-90.
56. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* 2006;367(9524):1747-57.
57. Rothwell PM. The high cost of not funding stroke research: a comparison with heart disease and cancer. *The Lancet* 2001;357(9268):1612-16.
58. Scarborough P, Peto V, Bhatnagar P, Kaur A, Leal J, Fernandez RL, et al. Stroke Statistics 2009 Edition: Brittish Heart Foundation, 2009.

59. Moon L, Moise O, Jacobzone S, Group A-SE. Stroke Care in OECD countries: A comparison of Treatment, Costs and Outcome in 17 Countries, 2003.
60. Pendlebury ST. Worldwide under-funding of stroke research. *Int J Stroke* 2007;2(2):80-4.
61. Pendlebury ST, Rothwell PM, Algra A, Ariesen M-J, Bakac G, Czlonkowska A, et al. Underfunding of Stroke Research. *Stroke* 2004;35(10):2368-71.
62. Broderick J. Gross underfunding of stroke research. *The Lancet Neurology* 2004;3(12):705-06.
63. Mishra NK, Diener HC, Lyden PD, Bluhmki E, Lees KR. Influence of age on outcome from thrombolysis in acute stroke: a controlled comparison in patients from the Virtual International Stroke Trials Archive (VISTA). *Stroke* 2010;41(12):2840-8.
64. Feigin VL. Stroke in developing countries: can the epidemic be stopped and outcomes improved? *The Lancet Neurology* 2007;6(2):94-97.
65. Wolf PA, Clagett GP, Easton JD, Goldstein LB, Gorelick PB, Kelly-Hayes M, et al. Preventing Ischemic Stroke in Patients With Prior Stroke and Transient Ischemic Attack. *Stroke* 1999;30(9):1991-94.
66. Hier DB, Foulkes MA, Swiontoniowski M, Sacco RL, Gorelick PB, Mohr JP, et al. Stroke recurrence within 2 years after ischemic infarction. *Stroke* 1991;22(2):155-61.
67. Zhang H, Thijs L, Staessen JA. Blood Pressure Lowering for Primary and Secondary Prevention of Stroke. *Hypertension* 2006;48(2):187-95.
68. Ohkubo T, Asayama K, Kikuya M, Metoki H, Obara T, Saito S, et al. Prediction of ischaemic and haemorrhagic stroke by self-measured blood pressure at home: the Ohasama study. *Blood Press Monit* 2004;9(6):315-20.
69. Staessen JA, Thijs L, Fagard R, O'Brien ET, Clement D, de Leeuw PW, et al. Predicting cardiovascular risk using conventional vs ambulatory blood pressure in older patients with systolic hypertension. Systolic Hypertension in Europe Trial Investigators. *JAMA* 1999;282(6):539-46.
70. Thompson PD. Cerebrovascular disease in western Australia: an analysis of stroke type and associated risk factors. *Aust N Z J Med* 1983;13(5):497-503.
71. MacMahon S, Peto R, Collins R, Godwin J, Cutler J, Sorlie P, et al. Blood pressure, stroke, and coronary heart disease: Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *The Lancet* 1990;335(8692):765-74.
72. Wolf PA. Risk factors for stroke. *Stroke* 1985;16(3):359-60.
73. Luepker RV, Steffen LM, Jacobs DR, Zhou X, Blackburn H. Trends in Blood Pressure and Hypertension Detection, Treatment and Control 1980-2009: The Minnesota Heart Survey. *Circulation* 2012.
74. Han JL. Actions to control hypertension among adults in Oklahoma. *Prev Chronic Dis* 2011;8(1):A10.
75. Alderman MH. Blood pressure control to reduce cardiovascular morbidity and mortality: today and tomorrow. *Eff Clin Pract* 1998;1(1):23-5.
76. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* 1991;22(8):983-8.
77. Preliminary Report of the Stroke Prevention in Atrial Fibrillation Study. *New England Journal of Medicine* 1990;322(12):863-68.
78. Oldgren J, Wallentin L, Alexander JH, James S, Jönelid B, Steg G, et al. New oral anticoagulants in addition to single or dual antiplatelet therapy after an acute coronary syndrome: a systematic review and meta-analysis. *European Heart Journal* 2013.

79. Iqbal MB, Taneja AK, Lip GYH, Flather M. Recent developments in atrial fibrillation. *BMJ* 2005;330(7485):238-43.
80. Sowers JR. Stroke in Patients With Diabetes. *The Journal of Clinical Hypertension* 2004;6(2):62-63.
81. BARRETT-CONNOR E, KHAW K-T. DIABETES MELLITUS: AN INDEPENDENT RISK FACTOR FOR STROKE? *American Journal of Epidemiology* 1988;128(1):116-23.
82. Vermeer SE, Sandee W, Algra A, Koudstaal PJ, Kappelle LJ, Dippel DWJ, et al. Impaired Glucose Tolerance Increases Stroke Risk in Nondiabetic Patients With Transient Ischemic Attack or Minor Ischemic Stroke. *Stroke* 2006;37(6):1413-17.
83. Chiuve SE, Rexrode KM, Spiegelman D, Logroscino G, Manson JE, Rimm EB. Primary Prevention of Stroke by Healthy Lifestyle. *Circulation* 2008;118(9):947-54.
84. Kurth T MSCGJ, et al. HEalthy lifestyle and the risk of stroke in women. *Archives of Internal Medicine* 2006;166(13):1403-09.
85. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *The Lancet* 2010;376(9735):112-23.
86. Goldstein LB, Adams R, Alberts MJ, Appel LJ, Brass LM, Bushnell CD, et al. Primary Prevention of Ischemic Stroke. *Stroke* 2006;37(6):1583-633.
87. Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, et al. The reasons for geographic and racial differences in stroke study: objectives and design. *Neuroepidemiology* 2005;25(3):135-43.
88. Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, et al. Traditional Risk Factors as the Underlying Cause of Racial Disparities in Stroke. *Stroke* 2011;42(12):3369-75.
89. Kleindorfer D. Sociodemographic Groups at Risk: Race/Ethnicity. *Stroke* 2009;40(3 suppl 1):S75-S78.
90. Ostchega Y, Dillon CF, Hughes JP, Carroll M, Yoon S. Trends in Hypertension Prevalence, Awareness, Treatment, and Control in Older U.S. Adults: Data from the National Health and Nutrition Examination Survey 1988 to 2004. *Journal of the American Geriatrics Society* 2007;55(7):1056-65.
91. Sacco RL, Boden-Albala B, Abel G, Lin I-F, Elkind M, Hauser WA, et al. Race-Ethnic Disparities in the Impact of Stroke Risk Factors. *Stroke* 2001;32(8):1725-31.
92. Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and Thrombolytic Therapy for Ischemic Stroke. *CHEST Journal* 2001;119(1_suppl):300S-20S.
93. Leys D, Cordonnier C. rt-PA for ischaemic stroke: what will the next question be? *The Lancet*;379(9834):2320-21.
94. Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, et al. Guidelines for the Early Management of Adults With Ischemic Stroke. *Stroke* 2007;38(5):1655-711.
95. Wardlaw JM, Murray V, Berge E, Del Zoppo GJ. Thrombolysis for acute ischaemic stroke. *Cochrane Database Syst Rev* 2009;7(4):CD000213.
96. The National Institute Of Neurological D, Stroke R-p, Stroke Study G. Tissue plasminogen activator for acute ischemic stroke. *New England Journal of Medicine* 1995;333(24):1581-87.
97. Califf RM, Fortin DF, Tenaglia AN, Sane DC. Clinical risks of thrombolytic therapy. *Am J Cardiol* 1992;69(2):12A-20A.
98. Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Guidetti D, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med* 2008;359(13):1317-29.

99. Larrue V, von Kummer R, Müller A, Bluhmki E. Risk Factors for Severe Hemorrhagic Transformation in Ischemic Stroke Patients Treated With Recombinant Tissue Plasminogen Activator : A Secondary Analysis of the European-Australasian Acute Stroke Study (ECASS II). *Stroke* 2001;32(2):438-41.
100. The European Stroke Organisation Executive C, the ESOWC. Guidelines for Management of Ischaemic Stroke and Transient Ischaemic Attack 2008. *Cerebrovasc. Dis.* 2008;25(5):457-507.
101. group Ew. Should the time window for thrombolysis be extended? *Karolinska stroke update*, 2008.
102. Lyden PD. *Thrombolytic therapy for stroke*: Humana Press, 2001.
103. Collaborative overview of randomised trials of antiplatelet therapy--I: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. Antiplatelet Trialists' Collaboration. *BMJ* 1994;308(6921):81-106.
104. The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *The Lancet* 1997;349(9065):1569-81.
105. Khaja AM, Grotta JC. Established treatments for acute ischaemic stroke. *The Lancet*;369(9558):319-30.
106. Smith WS, Sung G, Saver J, Budzik R, Duckwiler G, Liebeskind DS, et al. Mechanical Thrombectomy for Acute Ischemic Stroke. *Stroke* 2008;39(4):1205-12.
107. Kulcsár Z, Bonvin C, Pereira VM, Altrichter S, Yilmaz H, Lövblad KO, et al. Penumbra System: A Novel Mechanical Thrombectomy Device for Large-Vessel Occlusions in Acute Stroke. *American Journal of Neuroradiology* 2010;31(4):628-33.
108. Broderick JP, Palesch YY, Demchuk AM, Yeatts SD, Khatri P, Hill MD, et al. Endovascular Therapy after Intravenous t-PA versus t-PA Alone for Stroke. *New England Journal of Medicine* 2013;368(10):893-903.
109. Kidwell CS, Jahan R, Alger JR, Schaewe TJ, Guzy J, Starkman S, et al. Design and rationale of the Mechanical Retrieval and Recanalization of Stroke Clots Using Embolectomy (MR RESCUE) Trial. *Int J Stroke* 2012;13(10):1747-4949.
110. Kidwell CS, Jahan R, Gornbein J, Alger JR, Nenov V, Ajani Z, et al. A Trial of Imaging Selection and Endovascular Treatment for Ischemic Stroke. *New England Journal of Medicine* 2013;368(10):914-23.
111. del Zoppo GJ, Higashida RT, Furlan AJ, Pessin MS, Rowley HA, Gent M. PROACT: A Phase II Randomized Trial of Recombinant Pro-Urokinase by Direct Arterial Delivery in Acute Middle Cerebral Artery Stroke. *Stroke* 1998;29(1):4-11.
112. Furlan A HRWL, et al. Intra-arterial prourokinase for acute ischemic stroke: The proact ii study: a randomized controlled trial. *JAMA* 1999;282(21):2003-11.
113. Ogawa A, Mori E, Minematsu K, Taki W, Takahashi A, Nemoto S, et al. Randomized Trial of Intraarterial Infusion of Urokinase Within 6 Hours of Middle Cerebral Artery Stroke: The Middle Cerebral Artery Embolism Local Fibrinolytic Intervention Trial (MELT) Japan. *Stroke* 2007;38(10):2633-39.
114. Macleod MR, Davis SM, Mitchell PJ, Gerraty RP, Fitt G, Hankey GJ, et al. Results of a multicentre, randomised controlled trial of intra-arterial urokinase in the treatment of acute posterior circulation ischaemic stroke. *Cerebrovasc Dis* 2005;20(1):12-7.
115. Ducrocq X, Bracard S, Taillandier L, Anxionnat R, Lacour JC, Guillemin F, et al. Comparison of intravenous and intra-arterial urokinase thrombolysis for acute

ischaemic stroke: Randomised study of 27 patients. *Journal of Neuroradiology* 2005;32(1):26-32.
116. Nam J, Jing H, O'Reilly D. Intra-arterial thrombolysis vs. standard treatment or intravenous thrombolysis in adults with acute ischemic stroke: a systematic review and meta-analysis. *Int. J. Stroke* 2013:n/a-n/a.
117. Lindsberg PJ, Mattle HP. Therapy of Basilar Artery Occlusion: A Systematic Analysis Comparing Intra-Arterial and Intravenous Thrombolysis. *Stroke* 2006;37(3):922-28.
118. Lisboa RC, Jovanovic BD, Alberts MJ. Analysis of the Safety and Efficacy of Intra-Arterial Thrombolytic Therapy in Ischemic Stroke. *Stroke* 2002;33(12):2866-71.
119. Rha J-H, Saver JL. The Impact of Recanalization on Ischemic Stroke Outcome: A Meta-Analysis. *Stroke* 2007;38(3):967-73.
120. Fields JD, Khatri P, Nesbit GM, Liu KC, Barnwell SL, Lutsep HL, et al. Meta-analysis of randomized intra-arterial thrombolytic trials for the treatment of acute stroke due to middle cerebral artery occlusion. *Journal of NeuroInterventional Surgery* 2011;3(2):151-55.
121. Lee M, Hong K-S, Saver JL. Efficacy of Intra-Arterial Fibrinolysis for Acute Ischemic Stroke: Meta-Analysis of Randomized Controlled Trials. *Stroke* 2010;41(5):932-37.
122. Tenser MS, Amar AP, Mack WJ. Mechanical Thrombectomy for Acute Ischemic Stroke Using the MERCI Retriever and Penumbra Aspiration Systems. *World Neurosurgery* 2011;76(6):S16-S23.
123. Ciccone A, Valvassori L, Nichelatti M, Sgoifo A, Ponzio M, Sterzi R, et al. Endovascular Treatment for Acute Ischemic Stroke. *New England Journal of Medicine* 2013;368(10):904-13.
124. Baker WL, Colby JA, Tongbram V, Talati R, Silverman IE, White CM, et al. Neurothrombectomy Devices for the Treatment of Acute Ischemic Stroke: State of the Evidence. *Annals of Internal Medicine* 2011;154(4):243-52.
125. Kasner SE. Clinical interpretation and use of stroke scales. *The Lancet Neurology* 2006;5(7):603-12.
126. Baker K, Cano SJ, Playford ED. Outcome Measurement in Stroke. *Stroke* 2011;42(6):1787-94.
127. Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials. *Int. J. Stroke* 2009;4(3):200-05.
128. Quinn TJ, Dawson J. Acute 'strokenomics': efficacy and economic analyses of alteplase for acute ischemic stroke. *Expert Rev Pharmacoecon Outcomes Res* 2009;9(6):513-22.
129. Quinn T, Dawson J, Walters M, Lees KR. Variability in modified rankin scoring across a large cohort of observers. *Stroke* 2008;39(2):P509.
130. Quinn TJ, Dawson J, Walters MR, Lees KR. Reliability of the Modified Rankin Scale A Systematic Review. *Stroke* 2009;40(10):3393-95.
131. Lees KR. Modified Rankin Scale. Glasgow: University of Glasgow, 1999.
132. Lees KR, Bath PM, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al. Contemporary outcome measures in acute stroke research: choice of primary outcome measure. *Stroke* 2012;43(4):1163-70.
133. Goldstein LB, Samsa GP. Reliability of the National Institutes of Health Stroke Scale. Extension to non-neurologists in the context of a clinical trial. *Stroke* 1997;28(2):307-10.
134. Patrick Lyden RR, Lin Liu, Marian Emr, Margo Warren, John Marler. National Institute of Health Stroke Scale certification is Reliable Across Multiple Venues. *Stroke* 2009;40:2507-11.

135. Woo D, Broderick JP, Kothari RU, Lu M, Brott T, Lyden PD, et al. Does the National Institutes of Health Stroke Scale favor left hemisphere strokes? NINDS t-PA Stroke Study Group. *Stroke* 1999;30(11):2355-9.
136. Lyden PD, Hantson L. Assessment scales for the evaluation of stroke patients. *J Stroke Cerebrovasc Dis* 1998;7(2):113-27.
137. Middleton S, Levi C, Ward J, Grimshaw J, Griffiths R, D'Este C, et al. Fever, hyperglycaemia and swallowing dysfunction management in acute stroke: A cluster randomised controlled trial of knowledge transfer. *Implementation Science* 2009;4(1):16.
138. Tilley BC, Marler J, Geller NL, Lu M, Legler J, Brott T, et al. Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke t-PA Stroke Trial. *Stroke* 1996;27(11):2136-42.
139. EMA. CPMP Points to Consider on Clinical Investigation of Medicinal Products for the Treatmant of Acute Stroke. 2001.
140. Lees KR, Bath PM, Schellinger PD, Kerr DM, Fulton R, Hacke W, et al. Contemporary outcome measures in acute stroke research: choice of primary outcome measure.
141. Koenig IR, Ziegler A, Bluhmki E, Hacke W, Bath PMW, Sacco RL, et al. Predicting long-term outcome after acute ischemic stroke - A simple index works in patients from controlled clinical trials. *Stroke* 2008;39(6):1821-26.
142. Saver JL, Altman H. Relationship Between Neurologic Deficit Severity and Final Functional Outcome Shifts and Strengthens During First Hours After Onset. *Stroke* 2012.
143. Weimar C, Konig IR, Kraywinkel K, Ziegler A, Diener HC. Age and National Institutes of Health Stroke Scale Score Within 6 Hours After Onset Are Accurate Predictors of Outcome After Cerebral Ischemia: Development and External Validation of Prognostic Models. *Stroke* 2004;35(1):158-62.
144. The NINDS t-PA Stroke Study Group. Generalized Efficacy of t-PA for Acute Stroke. *Stroke* 1997;28(11):2119-25.
145. Luitse MJA, Biessels GJ, Rutten GEHM, Kappelle LJ. Diabetes, hyperglycaemia, and acute ischaemic stroke. *The Lancet Neurology* 2012;11(3):261-71.
146. Bruno A, Levine SR, Frankel MR, Brott TG, Lin Y, Tilley BC, et al. Admission glucose level and clinical outcomes in the NINDS rt-PA Stroke Trial. *Neurology* 2002;59(5):669-74.
147. Hu GC, Hsieh SF, Chen YM, Hu YN, Kang CL, Chien KL. The prognostic roles of initial glucose level and functional outcomes in patients with ischemic stroke: difference between diabetic and nondiabetic patients. *Disabil Rehabil* 2012;34(1):34-9.
148. Stead LG, Gilmore RM, Bellolio MF, Mishra S, Bhagra A, Vaidyanathan L, et al. Hyperglycemia as an independent predictor of worse outcome in non-diabetic patients presenting with acute ischemic stroke. *Neurocrit Care* 2009;10(2):181-86.
149. Kruyt ND, Biessels GJ, DeVries JH, Roos YB. Hyperglycemia in acute ischemic stroke: pathophysiology and clinical management. *Nat Rev Neurol* 2010;6(3):145-55.
150. Schurr A. Bench-to-bedside review: A possible resolution of the glucose paradox of cerebral ischemia. *Critical Care* 2002;6(4):330 - 34.
151. Kaplan RCP, Tirschwell DLMD, Longstreth WTJMDMPH, Manolio TAMDP, Heckbert SRMDP, LeValley AJMS, et al. Blood Pressure Level and Outcomes in Adults Aged 65 and Older with Prior Ischemic Stroke. *Journal of the American Geriatrics Society September* 2006;54(9):1309-16.

152. Leira R, Millán M, Díez-Tejedor E, Blanco M, Serena J, Fuentes B, et al. Age Determines the Effects of Blood Pressure Lowering During the Acute Phase of Ischemic Stroke. The TICA Study. *Hypertension* 2009.
153. Sare GM, Ali M, Shuaib A, Bath PM. Relationship between hyperacute blood pressure and outcome after ischemic stroke: data from the VISTA collaboration. *Stroke* 2009;40(6):2098-103.
154. Boreas AM, Lodder J, Kessels F, de Leeuw PW, Troost J. Prognostic value of blood pressure in acute stroke. *J Hum Hypertens* 2002;16(2):111-6.
155. Willmot M, Leonardi-Bee J, Bath PM. High blood pressure in acute stroke and subsequent outcome: a systematic review. *Hypertension* 2004;43(1):18-24.
156. Sprigg Na, Gray LJa, Bath PMWa, Boysen Gb, De Deyn PPc, Friis Pd, et al. Relationship between outcome and baseline blood pressure and other haemodynamic measures in acute ischaemic stroke: data from the TAIST trial. *Journal of Hypertension July* 2006;24(7):1413-17.
157. Leonardi Bee JM, Bath PMWF, Phillips SJF, Sandercock PAGF, for the ISTCG. Blood Pressure and Clinical Outcomes in the International Stroke Trial. *Stroke May* 2002;33(5):1315-20.
158. Kimura K, Minematsu K, Yamaguchi T. Atrial fibrillation as a predictive factor for severe stroke and early death in 15,831 patients with acute ischaemic stroke. *J Neurol Neurosurg Psychiatry* 2005;76(5):679-83.
159. Jorgensen HSMD, Nakayama HMDP, Reith JMD, Raaschou HOMD, Olsen TSMDP. Acute Stroke With Atrial Fibrillation: The Copenhagen Stroke Study. *Stroke October* 1996;27(10):1765-69.
160. Jorgensen HSMDP, Nakayama HMDP, Reith JM, Raaschou HOM, Olsen TSMDP. Stroke recurrence: Predictors, severity, and prognosis. The Copenhagen Stroke Study. *Neurology April* 1997;48(4):891-95.
161. Frankel MR, Morgenstern LB, Kwiatkowski T, Lu M, Tilley BC, Broderick JP, et al. Predicting prognosis after stroke: a placebo group analysis from the National Institute of Neurological Disorders and Stroke rt-PA Stroke Trial. *Neurology* 2000;55(7):952-9.
162. II STAIR. Recommendations for Clinical Trial Evaluation of Acute Stroke Therapies. *Stroke* 2001;32(7):1598-606.
163. Concato J, Shah N, Horwitz RI. Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs. *New England Journal of Medicine* 2000;342(25):1887-92.
164. Pickering RM, Kunkel D, Fitton C, Ashburn A, Jenkinson D. In-hospital recruitment to observational studies of stroke. *Int J Rehabil Res* 2010;33(1):56-63.
165. Kidwell CS, Liebeskind DS, Starkman S, Saver JL. Trends in Acute Ischemic Stroke Trials Through the 20th Century. *Stroke* 2001;32(6):1349-59.
166. Hacke W, Kaste M, Fieschi C, Toni D, Lesaffre E, von Kummer R, et al. Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric stroke. The European Cooperative Acute Stroke Study (ECASS). *JAMA* 1995;274(13):1017-25.
167. Hacke W, Kaste M, Fieschi C, von Kummer R, Davalos A, Meier D, et al. Randomised double-blind placebo-controlled trial of thrombolytic therapy with intravenous alteplase in acute ischaemic stroke (ECASS II). Second European-Australasian Acute Stroke Study Investigators. *Lancet* 1998;352(9136):1245-51.
168. Clark WM, Albers GW, Madden KP, Hamilton S. The rtPA (alteplase) 0- to 6-hour acute stroke trial, part A (A0276g) : results of a double-blind, placebo-controlled,

multicenter study. Thromblytic therapy in acute ischemic stroke study investigators. *Stroke* 2000;31(4):811-6.
169. Clark WM, Wissman S, Albers GW, Jhamandas JH, Madden KP, Hamilton S. Recombinant tissue-type plasminogen activator (Alteplase) for ischemic stroke 3 to 5 hours after symptom onset. The ATLANTIS Study: a randomized controlled trial. Alteplase Thrombolysis for Acute Noninterventional Therapy in Ischemic Stroke. *JAMA* 1999;282(21):2019-26.
170. Davis SM, Donnan GA, Parsons MW, Levi C, Butcher KS, Peeters A, et al. Effects of alteplase beyond 3 h after stroke in the Echoplanar Imaging Thrombolytic Evaluation Trial (EPITHET): a placebo-controlled randomised trial. *Lancet Neurol* 2008;7(4):299-309.
171. Lees KR, Bluhmki E, von Kummer R, Brott TG, Toni D, Grotta JC, et al. Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. *Lancet* 2010;375(9727):1695-703.
172. Grp ISTC. The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial IST-3): a randomised controlled trial (vol 379, pg 2352, 2012). *Lancet* 2012;380(9843):730-30.
173. Sanossian N, Ovbiagele B. Prevention and management of stroke in very elderly patients. *Lancet Neurol* 2009;8(11):1031-41.
174. Sanossian N, Ovbiagele B. Prevention and management of stroke in very elderly patients. *The Lancet Neurology* 2009;8(11):1031-41.
175. Hallevi H, Grotta JC. SITS-MOST: old myths laid to rest. *The Lancet Neurology* 2007;6(4):292-93.
176. Lees KR, Ford GA, Muir KW, Ahmed N, Dyker AG, Atula S, et al. Thrombolytic therapy for acute stroke in the United Kingdom: experience from the safe implementation of thrombolysis in stroke (SITS) register. *QJM* 2008;101(11):863-69.
177. Lyden PD. In Anticipation of International Stroke Trial-3 (IST-3). *Stroke* 2012;43(6):1691-94.
178. Mishra NK, Ahmed N, Andersen G, Egido JA, Lindsberg PJ, Ringleb PA, et al. Thrombolysis in very elderly people: controlled comparison of SITS International Stroke Thrombolysis Registry and Virtual International Stroke Trials Archive. *BMJ* 2010;341:c6046.
179. Wahlgren N, Ahmed N, Dávalos A, Ford GA, Grond M, Hacke W, et al. Thrombolysis with alteplase for acute ischaemic stroke in the Safe Implementation of Thrombolysis in Stroke-Monitoring Study (SITS-MOST): an observational study. *The Lancet*;369(9558):275-82.
180. Hacke W, Donnan G, Fieschi C, Kaste M, von Kummer R, Broderick JP, et al. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet* 2004;363(9411):768-74.
181. Wahlgren N, Ahmed N, Eriksson N, Aichner F, Bluhmki E, Davalos A, et al. Multivariable Analysis of Outcome Predictors and Adjustment of Main Outcome Results to Baseline Data Profile in Randomized Controlled Trials Safe Implementation of Thrombolysis in Stroke-MOnitoring STudy (SITS-MOST). *Stroke* 2008;39(12):3316-22.
182. Sacco Rl DJTHJE, et al. Glycine antagonist in neuroprotection for patients with acute stroke: Gain americas: a randomized controlled trial. *JAMA: The Journal of the American Medical Association* 2001;285(13):1719-28.

183. Lees KR, Asplund K, Carolei A, Davis SM, Diener H-C, Kaste M, et al. Glycine antagonist (gavestinel) in neuroprotection (GAIN International) in patients with acute stroke: a randomised controlled trial. *The Lancet* 2000;355(9219):1949-54.
184. Diener HC, Cortens M, Ford G, Grotta J, Hacke W, Kaste M, et al. Lubeluzole in Acute Ischemic Stroke Treatment. *Stroke* 2000;31(11):2543-51.
185. Davis SM, Lees KR, Albers GW, Diener HC, Markabi S, Karlsson G, et al. Selfotel in Acute Ischemic Stroke. *Stroke* 2000;31(2):347-54.
186. Albers Gw GLBHDLLMftAASI. Aptiganel hydrochloride in acute ischemic stroke: A randomized controlled trial. *JAMA: The Journal of the American Medical Association* 2001;286(21):2673-82.
187. Investigators TR. A Randomized Trial of Tirilazad Mesylate in Patients With Acute Stroke (RANTTAS). *Stroke* 1996;27(9):1453-58.
188. Muir KW, Lees KR, Ford I, Davis S. Magnesium for acute stroke (Intravenous Magnesium Efficacy in Stroke trial): randomised controlled trial. *Lancet* 2004;363(9407):439-45.
189. Enlimomab Acute Stroke Trial I. Use of anti-ICAM-1 therapy in ischemic stroke: results of the Enlimomab Acute Stroke Trial. *Neurology* 2001;57(8):1428-34.
190. Lees KR. Does neuroprotection improve stroke outcome? *The Lancet* 1998;351(9114):1447-48.
191. Muir KW, Grosset DG. Neuroprotection for Acute Stroke. *Stroke* 1999;30(1):180-82.
192. Cheng YD, Al-Khoury L, Zivin JA. Neuroprotection for ischemic stroke: two decades of success and failure. *NeuroRx* 2004;1(1):36-45.
193. Teal P, Davis S, Hacke W, Kaste M, Lyden PD, Investigators ftmS, et al. A Randomized, Double-Blind, Placebo-Controlled Trial to Evaluate the Efficacy, Safety, Tolerability, and Pharmacokinetic/Pharmacodynamic Effects of a Targeted Exposure of Intravenous Repinotan in Patients With Acute Ischemic Stroke. *Stroke* 2009;40(11):3518-25.
194. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener H-C, et al. NXY-059 for Acute Ischemic Stroke. *New England Journal of Medicine* 2006;354(6):588-600.
195. Diener HC, Lees KR, Lyden P, Grotta J, Davalos A, Davis SM, et al. NXY-059 for the treatment of acute stroke: pooled analysis of the SAINT I and II Trials. *Stroke* 2008;39(6):1751-8.
196. Shuaib A, Lees KR, Lyden P, Grotta J, Davalos A, Davis SM, et al. NXY-059 for the Treatment of Acute Ischemic Stroke. *New England Journal of Medicine* 2007;357(6):562-71.
197. Davalos A, Castillo J, Alvarez-Sabin J, Secades JJ, Mercadal J, Lopez S, et al. Oral citicoline in acute ischemic stroke: an individual patient data pooling analysis of clinical trials. *Stroke* 2002;33(12):2850-7.
198. Clark WM, Warach SJ, Pettigrew LC, Gammans RE, Sabounjian LA. A randomized dose-response trial of citicoline in acute ischemic stroke patients. Citicoline Stroke Study Group. *Neurology* 1997;49(3):671-8.
199. Clark WM, Williams BJ, Selzer KA, Zweifler RM, Sabounjian LA, Gammans RE, et al. A Randomized Efficacy Trial of Citicoline in Patients With Acute Ischemic Stroke. *Stroke* 1999;30(12):2592-97.
200. Dávalos A, Alvarez-Sabín J, Castillo J, Díez-Tejedor E, Ferro J, Martínez-Vila E, et al. Citicoline in the treatment of acute ischaemic stroke: an international, randomised, multicentre, placebo-controlled study (ICTUS trial). *The Lancet*;380(9839):349-57.

201. Dachs RJ, Burton JH, Joslin J. A user's guide to the NINDS rt-PA stroke trial database. *PLoS Med* 2008;5(5):e113.
202. Sandercock P, Niewada M, Czlonkowska A, Group tISTC. The International Stroke Trial database. *Trials* 2011;12(1):101.
203. Myzoon A, Bath P, Brady M, Davis S, Diener HC, Donnan G, et al. Development, expansion, and use of a stroke clinical trials resource for novel exploratory analyses. *Int J Stroke* 2012;7(2):133-8.
204. Ali M, Bath PMW, Curram J, Davis SM, Diener HC, Donnan GA, et al. The virtual international stroke trials archive. *Stroke* 2007;38(6):1905-10.
205. Meretoja A, Roine RO, Kaste M, Linna M, Juntunen M, Erilä T, et al. Stroke Monitoring on a National Level. *Stroke* 2010;41(10):2239-46.
206. Tu JV, Willison DJ, Silver FL, Fang J, Richards JA, Laupacis A, et al. Impracticability of informed consent in the Registry of the Canadian Stroke Network. *N Engl J Med* 2004;350(14):1414-21.
207. Mohan KM, Crichton SL, Grieve AP, Rudd AG, Wolfe CDA, Heuschmann PU. Frequency and predictors for the risk of stroke recurrence up to 10 years after stroke: the South London Stroke Register. *Journal of Neurology, Neurosurgery & Psychiatry* 2009;80(9):1012-18.
208. Hamann GF, Weimar C, Glahn J, Busse O, Diener HC. Adherence to Secondary Stroke Prevention Strategies – Results from the German Stroke Data Bank. *Cerebrovasc. Dis.* 2003;15(4):282-88.
209. Foulkes MA, Wolf PA, Price TR, Mohr JP, Hier DB. The Stroke Data Bank: design, methods, and baseline characteristics. *Stroke* 1988;19(5):547-54.
210. investigators TS. SITS Publications.
211. Elkins JS, Khatabi T, Fung L, Rootenberg J, Johnston SC. Recruiting Subjects for Acute Stroke Trials. *Stroke* 2006;37(1):123-28.
212. Hemmen TM, Rapp KS, Emond JA, Raman R, Lyden PD. Analysis of the National Institute of Neurological Disorders and Stroke tissue plasminogen activator studies following European Cooperative Acute Stroke Study III patient selection criteria. *J Stroke Cerebrovasc Dis* 2010;19(4):290-3.
213. Saver JL, Albers GW, Dunn B, Johnston KC, Fisher M, Consortium ftSV. Stroke Therapy Academic Industry Roundtable (STAIR) Recommendations for Extended Window Acute Stroke Therapy Trials. *Stroke* 2009;40(7):2594-600.
214. Grotta JC, Welch KMA, Fagan SC, Lu M, Frankel MR, Brott T, et al. Clinical Deterioration Following Improvement in the NINDS rt-PA Stroke Trial. *Stroke* 2001;32(3):661-68.
215. Lees KR. Neuroprotection Is Unlikely to Be Effective in Humans Using Current Trial Designs: An Opposing View. *Stroke* 2002;33(1):308-09.
216. Dorman PJ, Sandercock PAG. Considerations in the Design of Clinical Trials of Neuroprotective Therapy in Acute Stroke. *Stroke* 1996;27(9):1507-15.
217. Raab GM, Day S, Sales J. How to Select Covariates to Include in the Analysis of a Clinical Trial. *Controlled Clinical Trials* 2000;21(4):330-42.
218. Optimising the Analysis of Stroke Trials C. Should stroke trials adjust functional outcome for baseline prognostic factors? *Stroke* 2009;40(3):888-94.
219. Senn SJ. Covariate imbalance and random allocation in clinical trials. *Stat Med* 1989;8(4):467-75.
220. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *The Lancet* 2000;355(9209):1064-69.

221. Saver JL. Novel end point analytic techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials. *Stroke* 2007;38(11):3055-62.
222. Kerr DM, Fulton RL, Lees KR. Seven-Day NIHSS Is a Sensitive Outcome Measure for Exploratory Clinical Trials in Acute Stroke: Evidence From the Virtual International Stroke Trials Archive. *Stroke* 2012;2:2.
223. Optimising Analysis of Stroke Trials C, Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J. Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. *Stroke* 2007;38(6):1911-15.
224. Fisher M, Hanley DF, Howard G, Jauch EC, Warach S, Group ftS. Recommendations From the STAIR V Meeting on Acute Stroke Trials, Technology and Outcomes. *Stroke* 2007;38(2):245-48.
225. Rohrig B, du Prel JB, Wachtlin D, Kwiecien R, Blettner M. Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2010;107(31-32):552-6.
226. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134(8):663-94.
227. Weaver CS, Leonardi-Bee J, Bath-Hextall FJ, Bath PMW. Sample Size Calculations in Acute Stroke Trials: A Systematic Review of Their Reporting, Characteristics, and Relationship With Outcome. *Stroke* 2004;35(5):1216-24.
228. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328(7441):702-08.
229. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. *Lancet* 2004;363(9411):768-74.
230. Broderick JP, Lu M, Kothari R, Levine SR, Lyden PD, Haley EC, et al. Finding the most powerful measures of the effectiveness of tissue plasminogen activator in the NINDS tPA Stroke Trial. *Stroke* 2000;31(10):2335-41.
231. Bruno A. Selecting acute stroke patients for thrombolytic therapy. *ACC Current Journal Review* 2000;9(3):79-88.
232. Bruno A, Biller J, Adams HP, Jr., Clarke WR, Woolson RF, Williams LS, et al. Acute blood glucose level and outcome from ischemic stroke. Trial of ORG 10172 in Acute Stroke Treatment (TOAST) Investigators. *Neurology* 1999;52(2):280-4.
233. Werner Hacke EB, Thorsten Steiner, Turgut Tatlisumak, Marie-Helene Mahagne, Marisa-Luisa Saccetti, Deiter Meier. Dichotomised Efficacy End Points and Global End-Point analysis Applied to the ECASS intention-to-treat Dataset: Post Hoc analysis of ECASS I. *Stroke* 1998;29:2073-75.
234. Young FB, Weir CJ, Lees KR. Comparison of the National Institutes of Health Stroke Scale with disability outcome measures in acute stroke trials. *Stroke* 2005;36(10):2187-92.
235. Cramer SC, Procaccio V. Correlation between genetic polymorphisms and stroke recovery: analysis of the GAIN Americas and GAIN International Studies. *Eur J Neurol* 2012;19(5):718-24.
236. Ali M. The Virtual International Stroke Trials Archive (VISTA) : promulgation of a clinial trial resource. University of Glasgow.
237. Ali M, Bath PM, Curram J, Davis SM, Diener HC, Donnan GA, et al. The Virtual International Stroke Trials Archive. *Stroke* 2007;38(6):1905-10.
238. Ali M, Lyden P, Sacco RL, Shuaib A, Lees KR. Natural history of complications after intracerebral haemorrhage. *Eur J Neurol* 2009;16(5):624-30.

239. Hallevi H, Albright KC, Martin-Schild S, Barreto AD, Grotta JC, Savitz SI. The complications of cardioembolic stroke: lessons from the VISTA database. *Cerebrovasc Dis* 2008;26(1):38-40.
240. Dawson J, Lees JS, Chang TP, Walters MR, Ali M, Davis SM, et al. Association between disability measures and healthcare costs after initial treatment for acute stroke. *Stroke* 2007;38(6):1893-8.
241. Mishra NK, Lyden P, Grotta JC, Lees KR. Thrombolysis is associated with consistent functional improvement across baseline stroke severity: a comparison of outcomes in patients from the Virtual International Stroke Trials Archive (VISTA). *Stroke* 2010;41(11):2612-7.
242. Dawson J, Lees KR, Weir CJ, Quinn T, Ali M, Hennerici MG, et al. Baseline serum urate and 90-day functional outcomes following acute ischemic stroke. *Cerebrovasc Dis* 2009;28(2):202-3.
243. Prosser J, MacGregor L, Lees KR, Diener HC, Hacke W, Davis S. Predictors of early cardiac morbidity and mortality after ischemic stroke. *Stroke* 2007;38(8):2295-302.
244. Quinn TJ, Dawson J, Lees JS, Chang T-P, Walters MR, Lees KR, et al. Time Spent at Home Poststroke. *Stroke* 2008;39(1):231-33.
245. Mishra NK, Shuaib A, Lyden P, Diener H-C, Grotta J, Davis S, et al. Home Time Is Extended in Patients With Ischemic Stroke Who Receive Thrombolytic Therapy. *Stroke* 2011;42(4):1046-50.
246. Mishra NK, Davis SM, Kaste M, Lees KR, Collaboration ftV. Comparison of Outcomes Following Thrombolytic Therapy Among Patients With Prior Stroke and Diabetes in the Virtual International Stroke Trials Archive (VISTA). *Diabetes Care* 2010;33(12):2531-37.
247. Vogt G, Laage R, Shuaib A, Schneider A. Initial lesion volume is an independent predictor of clinical stroke outcome at day 90: an analysis of the Virtual International Stroke Trials Archive (VISTA) database. *Stroke* 2012;43(5):1266-72.
248. Dowlatshahi D, Demchuk AM, Flaherty ML, Ali M, Lyden PL, Smith EE. Defining hematoma expansion in intracerebral hemorrhage: relationship with patient outcomes. *Neurology* 2011;76(14):1238-44.
249. Whitehead J, Bolland K, Valdes-Marquez E, Lihic A, Ali M, Lees K. Using historical lesion volume data in the design of a new phase II clinical trial in acute stroke. *Stroke* 2009;40(4):1347-52.
250. Sedgwick P. Pearson's correlation coefficient. *BMJ* 2012;345.
251. Pirie W. Spearman Rank Correlation Coefficient. *Encyclopedia of Statistical Sciences*: John Wiley & Sons, Inc., 2004.
252. Stevens JW. Partial and semi-partial correlations, 2003.
253. Veall MR, Zimmermann KF. Pseudo-R 2's in the ordinal probit model*. *The Journal of Mathematical Sociology* 1992;16(4):333-42.
254. Windmeijer FAG. Goodness-of-fit measures in binary choice models 1. *Econometric Reviews* 1995;14(1):101-16.
255. group USc. FAQ: What are pseudo R-squareds. *SAS Data analysis examples*.
256. Cressie NAC, Whitford HJ. How to Use the Two Sample t-Test. *Biom. J.* 1986;28(2):131-48.
257. DePuy V, Berger VW, Zhou Y. Wilcoxon–Mann–Whitney Test. *Encyclopedia of Statistics in Behavioral Science*: John Wiley & Sons, Ltd, 2005.
258. Kirkwood B, Sterne J. *Essential Medical Statistics*: John Wiley & Sons, 2003.

259. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*: Springer, 2001.
260. Saver JL. Optimal End Points for Acute Stroke Therapy Trials. *Stroke* 2011;42(8):2356-62.
261. McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, et al. A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the IMPACT Project. *Clin Trials* 2010;7(1):44-57.
262. Kuritz SJ, Landis JR, Koch GG. A general overview of Mantel-Haenszel methods: applications and recent developments. *Annu Rev Public Health* 1988;9:123-60.
263. Bland JM, Altman DG. Statistics notes: Matching. *BMJ* 1994;309(6962):1128.
264. Wang S, Linkletter C, Maclure M, Dore D, Mor V, Buka S, et al. Future cases as present controls to adjust for exposure trend bias in case-only studies. *Epidemiology* 2011;22(4):568-74.
265. Rubin DB. Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *Journal of the American Statistical Association* 1979;74(366):318-28.
266. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat. Med.* 2008;27(12):2037-49.
267. Rosenbaum PR, Rubin DB. THE CENTRAL ROLE OF THE PROPENSITY SCORE IN OBSERVATIONAL STUDIES FOR CAUSAL EFFECTS. *Biometrika* 1983;70(1):41-55.
268. Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician* 1985;39(1):33-38.
269. Ho DE, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software* 2011;42(8).
270. Collaboration TOtAoST. Should Stroke Trials Adjust Functional Outcome for Baseline Prognostic Factors? *Stroke* 2009;40(3):888-94.
271. Bluhmki E, Chamorro A, Davalos A, Machnig T, Sauce C, Wahlgren N, et al. Stroke treatment with alteplase given 3.0-4.5 h after onset of acute ischaemic stroke (ECASS III): additional outcomes and subgroup analysis of a randomised controlled trial. *Lancet Neurol* 2009;8(12):1095-102.
272. Mishra NK, Davis SM, Kaste M, Lees KR. Comparison of outcomes following thrombolytic therapy among patients with prior stroke and diabetes in the Virtual International Stroke Trials Archive (VISTA). *Diabetes Care* 2010;33(12):2531-7.
273. Ford GA, Ahmed N, Azevedo E, Grond M, Larrue V, Lindsberg PJ, et al. Intravenous Alteplase for Stroke in Those Older Than 80 Years Old. *Stroke* 2010;41(11):2568-74.
274. Sandercock P, Lindley R, Wardlaw J, Dennis M, Lewis S, Venables G, et al. Third international stroke trial (IST-3) of thrombolysis for acute ischaemic stroke. *Trials* 2008;9(37):37.
275. Wardlaw J, Murray V, Berge E, Zoppo G, Sandercock P, Lindley R, et al. Recombinant tissue plasminogen activator for acute ischaemic stroke: an updated systematic review and meta-analysis. *Lancet* 2012.
276. The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial [IST-3]): a randomised controlled trial. *Lancet* 2012.
277. Saver JL, Gornbein J, Grotta J, Liebeskind D, Lutsep H, Schwamm L, et al. Number needed to treat to benefit and to harm for intravenous tissue plasminogen

activator therapy in the 3- to 4.5-hour window: joint outcome table analysis of the ECASS 3 trial. *Stroke* 2009;40(7):2433-7.
278. Kimura K, Iguchi Y, Yamashita S, Shibazaki K, Kobayashi K, Inoue T. Atrial fibrillation as an independent predictor for no early recanalization after IV-t-PA in acute ischemic stroke. *J Neurol Sci* 2008;267(1-2):57-61.
279. Kimura K, Iguchi Y, Shibazaki K, Iwanaga T, Yamashita S, Aoki J. IV t-PA therapy in acute stroke patients with atrial fibrillation. *J Neurol Sci* 2009;276(1-2):6-8.
280. Sanak D, Herzig R, Kral M, Bartkova A, Zapletalova J, Hutyra M, et al. Is atrial fibrillation associated with poor outcome after thrombolysis? *J Neurol* 2010;257(6):999-1003.
281. Seet RC, Zhang Y, Wijdicks EF, Rabinstein AA. Relationship between chronic atrial fibrillation and worse outcomes in stroke patients after intravenous thrombolysis. *Arch Neurol* 2011;68(11):1454-8.
282. Generalized efficacy of t-PA for acute stroke. Subgroup analysis of the NINDS t-PA Stroke Trial. *Stroke* 1997;28(11):2119-25.
283. Zhang JB, Ding ZY, Yang Y, Sun W, Hai F, Sui XN, et al. Thrombolysis with alteplase for acute ischemic stroke patients with atrial fibrillation. *Neurol Res* 2010;32(4):353-8.
284. Konig IR, Ziegler A, Bluhmki E, Hacke W, Bath PM, Sacco RL, et al. Predicting long-term outcome after acute ischemic stroke: a simple index works in patients from controlled clinical trials. *Stroke* 2008;39(6):1821-6.
285. Conway DS, Heeringa J, Van Der Kuip DA, Chin BS, Hofman A, Witteman JC, et al. Atrial fibrillation and the prothrombotic state in the elderly: the Rotterdam Study. *Stroke* 2003;34(2):413-7.
286. Roldan V, Marin F, Garcia A, Tello-Montoliu A, Lip GY. Is an advanced age an additive risk factor to the prothrombotic or hypercoagulable state in atrial fibrillation? *Int J Cardiol* 2006;110(2):265-6.
287. Young FB, Lees KR, Weir CJ. Improving trial power through use of prognosis-adjusted end points. *Stroke* 2005;36(3):597-601.
288. Cramer SCMD. Functional Imaging in Stroke Recovery. *Stroke* 2004;35(11):2695-98.
289. Commissariat à l'Energie Atomique Institut National de la Santé Et de la Recherche Médicale FUH, Grenoble. . Cell Therapy by Intravenous Injection of Mesenchymal Stem Cells After Stroke. *ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US)*, 2000.
290. System HFH. Phase II, Randomized, Double-Blinded, Placebo-Controlled Trial of Extended-Release Niacin (Niaspan®) to Augment Subacute Ischemic Stroke Recovery. *ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US)*, 2000.
291. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. *N Engl J Med* 1995;333(24):1581-7.
292. Mishra NK, Albers GW, Davis SM, Donnan GA, Furlan AJ, Hacke W, et al. Mismatch-based delayed thrombolysis: a meta-analysis. *Stroke* 2010;41(1):e25-33.
293. Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials. *Int J Stroke* 2009;4(3):200-5.
294. Savitz SI, Lew R, Bluhmki E, Hacke W, Fisher M. Shift analysis versus dichotomization of the modified Rankin scale outcome scores in the NINDS and ECASS-II trials. *Stroke* 2007;38(12):3205-12.
295. Saver JL. Novel end point analytic techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials. *Stroke* 2007;38(11):3055-62.

296. Mishra NK, Davis SM, Kaste M, Lees KR, Collaboration V. Comparison of outcomes following thrombolytic therapy among patients with prior stroke and diabetes in the Virtual International Stroke Trials Archive (VISTA). *Diabetes Care* 2010;33(12):2531-7.
297. CPMP. Points to Consider on Clinical Investigation of Medicinal Products for the Treatment of Acute Stroke, 2001.
298. Weimar C, König IR, Kraywinkel K, Ziegler A, Diener HC, Collaboration GSS. Age and National Institutes of Health Stroke Scale Score within 6 hours after onset are accurate predictors of outcome after cerebral ischemia: development and external validation of prognostic models. *Stroke* 2004;35(1):158-62.
299. Young FB, Weir CJ, Lees KR, Investigators GITSCa. Comparison of the National Institutes of Health Stroke Scale with disability outcome measures in acute stroke trials. *Stroke* 2005;36(10):2187-92.
300. Kasner SE. Clinical interpretation and use of stroke scales. *Lancet Neurol* 2006;5(7):603-12.
301. Whitehead J, Branson M, Todd S. A combined score test for binary and ordinal endpoints from clinical trials. *Stat Med* 2010;29(5):521-32.
302. Pocock SJ. Clinical trials with multiple outcomes: a statistical perspective on their design, analysis, and interpretation. *Control Clin Trials* 1997;18(6):530-45; discussion 46-9.
303. Brittain E, Palensky J, Blood J, Wittes J. Blinded subjective rankings as a method of assessing treatment effect: A large sample example from the Systolic Hypertension in the Elderly Program (SHEP). *Stat. Med.* 1997;16(6):681-93.
304. Feng W, Vasquez G, Suri MF, Lakshminarayan K, Qureshi AI. Repeated-measures analysis of the National Institute of Neurological Disorders and Stroke rt-PA stroke trial. *J Stroke Cerebrovasc Dis* 2011;20(3):241-6.
305. Li N, Elashoff RM, Li G, Saver J. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Stat Med* 2010;29(5):546-57.
306. Dymova N, Hanumara RC, Enander RT, Gagnon RN. Use of the global test statistic as a performance measurement in a reanalysis of environmental health data. *Am J Public Health* 2009;99(10):1739-41.
307. Bolland K, Whitehead J, Cobo E, Secades JJ. Evaluation of a sequential global test of improved recovery following stroke as applied to the ICTUS trial of citicoline. *Pharm. Stat.* 2009;8(2):136-49.
308. SAS 9.1.3 Help and Documentation [program]. Cary, NC: SAS Institute Inc, 2013.
309. R Development Core Team R. Foundation for Statistical Computing, Vienna, Austria. *R: A language and environment for statistical computing* 2009.
310. Ford G, Freemantle N. ECASS-II: intravenous alteplase in acute ischaemic stroke. *The Lancet* 1999;353(9146):65-65.
311. Teal P, Davis S, Hacke W, Kaste M, Lyden PD, Fierus M. A randomized, double-blind, placebo-controlled trial to evaluate the efficacy, safety, tolerability, and pharmacokinetic/pharmacodynamic effects of a targeted exposure of intravenous repinotan in patients with acute ischemic stroke: modified Randomized Exposure Controlled Trial (mRECT). *Stroke* 2009;40(11):3518-25.
312. Qureshi AI, Janjua N, Kirmani JF, Harris-Lane P, Suri MF, Zhou J, et al. Mechanical disruption of thrombus following intravenous tissue plasminogen activator for ischemic stroke. *J Neuroimaging* 2007;17(2):124-30.

313. Smith WS, Sung G, Starkman S, Saver JL, Kidwell CS, Gobin YP, et al. Safety and efficacy of mechanical embolectomy in acute ischemic stroke: results of the MERCI trial. *Stroke* 2005;36(7):1432-8.
314. Haley EC, Jr., Lyden PD, Johnston KC, Hemmen TM. A pilot dose-escalation safety study of tenecteplase in acute ischemic stroke. *Stroke* 2005;36(3):607-12.
315. Aulicky P, Mikulik R, Goldemund D, Reif M, Dufek M, Kubelka T. Safety of performing CT angiography in stroke patients treated with intravenous thrombolysis. *J Neurol Neurosurg Psychiatry* 2010;81(7):783-7.
316. Jiang WJ, Yu W, Du B, Gao F, Cui LY. Outcome of patients with >/=70% symptomatic intracranial stenosis after wingspan stenting. *Stroke* 2011;42(7):1971-5.
317. Toyoda K, Fujii K, Kamouchi M, Nakane H, Arihiro S, Okada Y, et al. Free radical scavenger, edaravone, in stroke with internal carotid artery occlusion. *J Neurol Sci* 2004;221(1-2):11-7.
318. Campbell* G. Statistics in the World of Medical Devices: The Contrast with Pharmaceuticals. *Journal of Biopharmaceutical Statistics* 2007;18(1):4-19.
319. Yue LQ. Statistical and regulatory issues with the application of propensity score analysis to nonrandomized medical device clinical studies. *J Biopharm Stat* 2007;17(1):1-13; discussion 15-7, 19-21, 23-7 passim.
320. Li H, Yue LQ. Statistical and Regulatory Issues in Nonrandomized Medical Device Clinical Studies. *Journal of Biopharmaceutical Statistics* 2007;18(1):20-30.
321. Rubin DB, Thomas N. Characterizing the Effect of Matching Using Linear Propensity Score Methods with Normal Distributions. *Biometrika* 1992;79(4):797-809.
322. Baksh MF, Todd S, Whitehead J, Lucini MM. Design considerations in the sequential analysis of matched case–control data. *Stat. Med.* 2005;24(6):853-67.
323. Mandava PMDPM, Kent TAMD. A Method to Determine Stroke Trial Success Using Multidimensional Pooled Control Functions. *Stroke May* 2009;40(5):1803-10.
324. US Food and Drug Administration Guidance for Industry and FDA - Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials.
Staff 2010.
325. Sorensen HT, Gillman MW. MATCHING IN CASE-CONTROL STUDIES. *British Medical Journal* 1995;310(6975):329-30.
326. Cepeda MS. The use of propensity scores in pharmacoepidemiologic research. *Pharmacoepidemiology and Drug Safety* 2000;9(2):103-04.
327. Caliendo M, Kopeinig S. SOME PRACTICAL GUIDANCE FOR THE IMPLEMENTATION OF PROPENSITY SCORE MATCHING. *Journal of Economic Surveys* 2008;22(1):31-72.
328. Rubin DB, Thomas N. Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates. *Journal of the American Statistical Association* 2000;95(450):573-85.
329. Hill J. Discussion of research using propensity-score matching: Comments on 'A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003' by Peter Austin, Statistics in Medicine. *Stat. Med.* 2008;27(12):2055-61.
330. Osborne JW. *Best Practices in Quantitative Methods*: SAGE Publications, 2008.
331. Austin PC, Grootendorst P, Normand S-LT, Anderson GM. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. *Stat. Med.* 2007;26(4):754-68.
332. Abadie A, Imbens GW. Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica* 2006;74(1):235-67.

333. Robins JM, Rotnitzky A. Semiparametric Efficiency in Multivariate Regression Models with Missing Data. *Journal of the American Statistical Association* 1995;90(429):122-29.
334. Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med* 2001;20(3):391-9.
335. Snijders TAB, Bosker RJ. *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. 2nd ed: Sage, 2011.
336. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ* 2004;328(7441):702-8.
337. Dirks M, Niessen LW, van Wijngaarden JD, Koudstaal PJ, Franke CL, van Oostenbrugge RJ, et al. Promoting thrombolysis in acute ischemic stroke. *Stroke* 2011;42(5):1325-30.
338. Johnston SC, Sidney S, Hills NK, Grosvenor D, Klingman JG, Bernstein A, et al. Standardized discharge orders after stroke: results of the quality improvement in stroke prevention (QUISP) cluster randomized trial. *Ann Neurol* 2010;67(5):579-89.
339. Lakshminarayan K, Borbas C, McLaughlin B, Morris NE, Vazquez G, Luepker RV, et al. A cluster-randomized trial to improve stroke care in hospitals. *Neurology* 2010;74(20):1634-42.
340. Middleton S, McElduff P, Ward J, Grimshaw JM, Dale S, D'Este C, et al. Implementation of evidence-based treatment protocols to manage fever, hyperglycaemia, and swallowing dysfunction in acute stroke (QASC): a cluster randomised controlled trial. *Lancet* 2011;378(9804):1699-706.
341. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004;57(8):785-94.
342. Moerbeek M, van Breukelen GJ, Berger MP. A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. *J Clin Epidemiol* 2003;56(4):341-50.
343. Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials. *Int J Stroke* 2009;4(3):200-5.
344. Chakraborty H, Moore J, Hartwell TD. Intracluster correlation adjustments to maintain power in cluster trials for binary outcomes. *Contemp Clin Trials* 2009;30(5):473-80.
345. Feng Z, Diehr P, Yasui Y, Evans B, Beresford S, Koepsell TD. Explaining community-level variance in group randomized trials. *Stat Med* 1999;18(5):539-56.
346. Huang P, Goetz CG, Woolson RF, Tilley B, Kerr D, Palesch Y, et al. Using global statistical tests in long-term Parkinson's disease clinical trials. *Movement Disorders* 2009;24(12):1732-39.
347. Tilley BC, Pillemer SR, Heyse SP, Li S, Clegg DO, Alarćon GS. Global statistical tests for comparing multiple outcomes in rheumatoid arthritis trials. *Arthritis & Rheumatism* 1999;42(9):1879-88.
348. Bagiella E. Clinical Trials in Rehabilitation: Single or Multiple Outcomes? *Archives of Physical Medicine and Rehabilitation* 2009;90(11, Supplement):S17-S21.
349. Li DK, Zhao GJ, Paty DW. Randomized controlled trial of interferon-beta-1a in secondary progressive MS: MRI results. *Neurology* 2001;56(11):1505-13.
350. Shames RS, Heilbron DC, Janson SL, Kishiyama JL, Au DS, Adelman DC. Clinical Differences Among Women with and Without Self-Reported Perimenstrual Asthma. *Annals of Allergy, Asthma & Immunology* 1998;81(1):65-72.

351. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004;20(1):93-99.
352. Benyo M, Flasko T, Molnar Z, Kerenyi A, Batta Z, Jozsa T, et al. Follow-Up of Thrombin Generation after Prostate Cancer Surgery: Global Test for Increased Hypercoagulability. *PLoS ONE* 2012;7(12):e51299.
353. Jung K, Becker B, Brunner E, Beißbarth T. Comparison of global tests for functional gene sets in two-group designs and selection of potentially effect-causing genes. *Bioinformatics* 2011;27(10):1377-83.
354. Clyne B, Bradley M, Smith S, Hughes C, Motterlini N, Clear D, et al. Effectiveness of medicines review with web-based pharmaceutical treatment algorithms in reducing potentially inappropriate prescribing in older people in primary care: a cluster randomized trial (OPTI-SCRIPT study protocol). *Trials* 2013;14(1):72.
355. Papadakis S, McDonald P, Pipe A, Letherdale S, Reid RD, Brown KS. Effectiveness of telephone-based follow-up support delivered in combination with a multi-component smoking cessation intervention in family practice: A cluster-randomized trial. *Preventive Medicine* (0).
356. Butler CC, Simpson SA, Hood K, Cohen D, Pickles T, Spanou C, et al. Training practitioners to deliver opportunistic multiple behaviour change counselling in primary care: a cluster randomised trial. *BMJ* 2013;346.
357. Kaner E, Bland M, Cassidy P, Coulton S, Dale V, Deluca P, et al. Effectiveness of screening and brief alcohol intervention in primary care (SIPS trial): pragmatic cluster randomised controlled trial. *BMJ* 2013;346(346):e8501.
358. Gregory J, Robling M, Bennert K, Channon S, Cohen D, Crowne E, et al. Development and evaluation by a cluster randomised trial of a psychosocial intervention in children and teenagers experiencing diabetes: the DEPICTED study. *Health Technol Assess* 2011;15(29):1-202.
359. Foy R, Hawthorne G, Gibb I, Eccles M, Steen N, Hrisos S, et al. A cluster randomised controlled trial of educational prompts in diabetes care: study protocol. *Implementation Science* 2007;2(1):22.
360. Butler CC, Simpson SA, Dunstan F, Rollnick S, Cohen D, Gillespie D, et al. Effectiveness of multifaceted educational programme to reduce antibiotic dispensing in primary care: practice based randomised controlled trial. *BMJ* 2012;344.
361. Kulier R, Khan K, Gulmezoglu AM, Carroli G, Cecatti J, Germar M, et al. A cluster randomized controlled trial to evaluate the effectiveness of the clinically integrated RHL evidence -based medicine course. *Reproductive Health* 2010;7(1):8.
362. Crespi CM, Maxwell AE, Wu S. Cluster randomized trials of cancer screening interventions: Are appropriate statistical methods being used? *Contemporary Clinical Trials* 2011;32(4):477-84.
363. Jick SS, Kaye JA, Vasilakis-Scaramozza C, Rodríguez LAG, Ruigómez A, Meier CR, et al. Validity of the General Practice Research Database. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 2003;23(5):686-89.
364. Clyde NGGa. Safe Haven, 2012.
365. University S. The Stanford Cancer Research Database. Stanford, 2008.
366. O'Toole L, Nurse P, Radda G. An analysis of cancer research funding in the UK. *Nat Rev Cancer* 2003;3(2):139-43.
367. Kerr DM, Fulton RL, Higgins P, Bath PM, Shuaib A, Lyden P, et al. Response of Blood Pressure and Blood Glucose to Treatment With Recombinant Tissue-Type

Plasminogen Activator in Acute Ischemic Stroke: Evidence From the Virtual International Stroke Trials Archive. *Stroke* 2011;10:10.