EXPLORING HETEROGENEITY IN META-ANALYSES USING SUMMARY AND INDIVIDUAL PATIENT DATA METHODOLOGIES FROM STROKE TRIALS

Jo Leonardi-Bee, BSc (Hons), MSc

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

April 2005

TABLE OF CONTENTS

ABSTRACT

Until recently, meta-analyses have usually been performed based on summary data methods. Individual patient data methods are becoming more popular, but the advantages of using these methods have not been fully investigated with regards to assessing and exploring heterogeneity.

This thesis has assessed whether there are any clinically important differences in the results from analysing data from three meta-analyses of randomised controlled trials in the area of stroke medicine, using summary and individual patient data methodologies.

Blood pressure in Acute Stroke Collaboration (BASC)

The management of blood pressure during the acute phase of stroke remains an enigma, therefore a systematic review and meta-analysis of existing randomised controlled trials was conducted to assess the effects of vasoactive drugs on outcome. Trends towards an increase in the risk of death, and death or dependency at the end of trial were found in patients randomised to a vasoactive drug as compared to those randomised to control. When baseline systolic blood pressure was taken into account in the analyses, patients randomised to a vasoactive drug had a significantly higher risk of death at the end of trial. Analyses also indicated that patients recruited early and within 48 hours has significant increases in the risk of death at the end of trial. However, no significant effects were seen for early change in systolic blood pressure.

Community occupational therapy in stroke patients

An evaluation of the efficacy of occupational therapy given in the community to stroke patients was performed using a systematic review and meta-analysis of randomised controlled trials. At the end of intervention, patient randomised to occupational therapy had significantly higher scores for extended and personal activities of daily living, and non-significantly higher scores for leisure participation. These effects appeared to be maintained over time. No effects were seen between the groups for death or minor psychiatric disorders as measured in patients or their carers. Subgroup analyses revealed that the benefits of occupational therapy were greatest when targeted interventions were used. Also, being male or independent at baseline was found to be important predictor of extended activities of daily living scores.

Dipyridamole in Stroke Collaboration (DISC)

Results from randomised controlled trials of dipyridamole, given with and without aspirin, for secondary prevention after stroke or transient ischaemic attack have given conflicting results; therefore, we performed a systematic review and meta-analysis. The risk of subsequent fatal or non-fatal stroke was reduced using the dual treatment of aspirin and dipyridamole as compared to either aspirin alone, dipyridamole alone, or control. Additionally, the dual treatment lowered the risk of non-fatal stroke, and subsequent vascular events defined by a composite outcome (non-fatal stroke, non fatal myocardial infarction, or vascular death). Analyses indicated that

these results were independent of method of formulation of dipyridamole, dose of aspirin, type of qualifying event, and gender of the patients. However, increasing age was found to be an important predictor of subsequent stroke.

These systematic reviews demonstrate that collaborations within the area of stroke medicine can be successful and much data can be shared. The findings from meta-analyses can be informative about the effectiveness of particular treatment and about which patients should be targeted for treatment; and may help steer the direction of future trials.

Although summary data meta-analyses are practically easier to perform, it is important that assessments and explorations of heterogeneity should always be performed. Meta-analyses based on individual patient data may be needed to allow for more in depth investigations of heterogeneity, especially of patient characteristics. However, they themselves are not the panacea to all difficulties since they are subject to particular problems, mainly related to obtaining individual patient data to enable these in depth analyses to be performed.

ACKNOWLEDGEMENTS

The work in this thesis was carried out under the supervision of Professor Philip Bath in the Division of Stroke Medicine, University of Nottingham and Dr Michael Dewey formerly of the Trent Institute in Health Services Research, University of Nottingham. Funding for the work was supported by the NHS Executive Trent and the Stroke Association.

I would like to thank Dr Marion Walker and Dr John Gladman for their involvement in the design and implementation of the 'Community Occupational Therapy in stroke patients' collaboration; Dr Fiona Bath-Hextall for her involvement in the design and implementation of the 'Blood pressure in Acute Stroke' collaboration (BASC); and Professor Philip Bath for his involvement in the design and implementation of the 'Dipyridamole in Stroke' collaboration (DISC).

I would like to give thanks to the collaborators for sharing their data with the collaborations; and to Dr Bill Browne for his advice about the statistical modules used within this thesis.

I would like to give particular thanks to my husband Mark, our children William and Aaron, and my parents David and Vivien, for their support, understanding and encouragement throughout this work.

LIST OF COLLABORATORS

Blood pressure in Acute Stroke Collaboration (BASC)

Bath-Hextall F, Bath P, Sandercock P, Ahmed N, Asplund K, Autret A, Barer D, Davis S, Donnan G, Dyker S, Gray C, Grotta J, Horn J, Hsu C, Huczynski J, Kaste M, Lees K, Limburg M, Lisk R, Lowe G, Muir K, Norris J, Pokrupa R, Rashid P, Squire I, Steiner T, Strand T, Uzuner N, Wahlgren NG, Azcona A, Lataste X, Bayer (Canada)

Community Occupational Therapy for stroke patients collaboration

Walker M, Gladman J, Bath P, Langhorne P, Dewey M, Corr S, Drummond A, Gilbertson L, Jongbloed L, Logan P, Parker C.

Dipyridamole in Stroke Collaboration (DISC)

Bath P, Bousser M-G, Davalos A, Diener H-C, Guiraud-Chaumeil B, Sivenius J, Yatsu F, Dewey ME, Baigent C.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

AN INTRODUCTION TO SYSTEMATIC REVIEWS AND META-ANALYSIS

1.1 Introduction

In the world of healthcare research, there is a vast quantity of subjective and evidence based knowledge. As researchers, we attempt to make sense of the evidence based knowledge which can be difficult especially if the findings are conflicting. How do we decide which of the results do we give more credibility to, and how do we form an overall conclusion of the findings?

This chapter describes how research evidence may be combined in a narrative manner, and using statistical methodology. This chapter also stresses the importance of identifying all of the available evidence based knowledge, and then describes the problems that are associated with reviews and the types of bias that may be present.

1.2 Combining sources of evidence

Within a research area, many clinical trials and studies may have been performed. Some of these studies may show positive intervention effects or associations, others may be inconclusive or show negative effects or associations. This variation may be related to differences in the characteristics of the participants under assessment within the studies, known as sampling variation; or possibly related to the studies using different measures to assess the outcome of interest. Alternatively, some studies may have been unable to detect a significant difference or association because they were too small and hence under-powered.

The overall conclusion from all of these studies may be difficult to interpret as they stand therefore methods have been devised to help combine the primary knowledge, and give a more representative and clearer understanding of the intervention/association.

1.3 Literature and systematic reviews

One technique used for combining the sources of evidence is called a literature review. Historically, literature reviews have been used to pool the evidence in a narrative manner and have been largely unsystematic, possibly resulting in many relevant trials or studies being ignored. In light of the disadvantages of this method, review methods were developed in the 1970s which systematically examined all the current evidence from primary studies in a narrative manner; known as systematic reviews. These methods reduce the chance of studies being missed by identifying more relevant studies, which in turn should lead to a clearer and more balanced view of the current evidence being achieved.

1.4 Statistically combining the current evidence

A technique was designed to critically evaluate and statistically combine results from the primary trials and studies to yield quantifiable intervention estimates. Glass first referred to this systematic method of reviewing current literature and knowledge as '*meta-analysis*' (Glass 1976); Huque describes the term '*meta-analysis*' better as "*...a statistical analysis that combines or*

integrates the results of several independent clinical trials considered by the analyst to be combinable" (Huque 1988).

Meta-analysis has been shown to be a powerful tool, since it allows for quantification of an intervention or association to be estimated, which is not possible in literature or systematic reviews. Maximising the number of studies through a comprehensive literature search will improve statistical power in the analyses, which should in turn improve the estimates yielded for the effect size of an intervention or association.

However, it must also be stressed that meta-analyses may also be used in epidemiological areas where an association may be of interest between two variables. The advantages and disadvantages of using meta-analyses are relevant for a clinical or an epidemiological setting.

The numbers of meta-analyses annually published have grown dramatically (Easterbrook *et al.* 1991) and it has been clearly established that they have the potential to change patient care (Lau *et al.* 1992).

1.5 Identifying eligible studies

A critical part to any systematic review, and hence meta-analysis; is the identification of primary studies. Strict inclusion and exclusion criteria need to be formulated each time a review is performed. If all of the relevant studies are not identified then the result from the systematic review could be biased and not truly represent the intervention effect or association. Many biases may enter in the early

stage of a meta-analysis which could potentially affect the interpretation of the results and hence conclusions from the meta-analysis. Therefore a thorough search programme needs to be well documented at the protocol stage to attempt to identify all of the available knowledge and literature.

This programme must not just rely on identifying studies through an electronic search engine but must also involve searching non English written journals, reference lists, and through contacting authors. The Cochrane Collaboration (The Cochrane Collaboration 2003) and the NHS Centre for Reviews and Dissemination (Deeks *et al.* 1996) have published guidelines which may be followed when conducting systematic reviews. Both of these guidelines are similar in the structuring and content formats, which involve how to identify relevant studies and the search procedures one may use to achieve this. They imply that the search should be exhaustive and attempts should be made to identify every study that has been performed in the particular area of interest by using specified search strategies using search engines such as MEDLINE, EMBASE and Web of Knowledge. They also stress the importance of hand searching reference lists and sifting through grey material such as dissertations, reports and conference proceedings.

1.6 Biases associated with systematic reviews

Systematic reviews are vulnerable to several problems, all of which may contribute to invalidating the generalisability of the results of the meta-analysis. As described earlier in this chapter, biased results

may be yielded when an inadequate search strategy has been performed where not all eligible studies are identified. Other sources of bias maybe related to publication bias, selection bias, missing data bias, English language bias, multiple publication bias and study quality bias.

1.6.1 Publication bias

Electronic databases such as MEDLINE or PUBMED are not sufficient to be the sole sources used to search for studies or trials. These databases only contain a selection of all medical journals. Another limitation is that they only contain published studies, therefore any unpublished data or data published as abstracts will be missed. This may introduce a type of bias known as publication bias into the meta-analysis (Easterbrook *et al.* 1991).

It is generally acknowledged that studies which show a benefit with regards to the efficacy of an intervention, are more likely to get published than studies which fail to detect a benefit (Song *et al.* 2000). In contrast, studies which fail to detect a benefit are more likely to be underpowered from using smaller sample sizes; which may lead to smaller studies being published in lower impact journals which themselves are less likely to be included in the electronic database (Gotzche 1987).

A variety of statistical methods may be used to assess publication bias; the most common methods used include Begg's funnel plot, Begg and Mazumdar's rank correlation test, and Egger's asymmetry test; these methods will be described and discussed in section 2.13.

1.6.2 Inclusion of unpublished data

There is a debate as to whether unpublished studies should be included in a meta-analysis. The Cochrane Collaboration generally includes unpublished material in their systematic reviews, since it is assumed that the results from an unpublished study are comparable to those obtained from published studies; and excluding the study may invalidate the generalisability of the results from the systematic review. Conversely, unpublished material is likely to be of inferior quality and bias may be increased through including unpublished work in the review. However, it is possible for a sensitivity analysis to be performed to assess the effect of including the unpublished studies. The interpretation should be that if a difference is seen when these studies are included then it is questionable whether unpublished data should be excluded.

1.6.3 Missing data bias

Missing data can lead to bias in the results from a meta-analysis. Data which is missing at study level, where the study did not record the variable of interest, is unique to meta-analyses (Sutton *et al.* 1998) and can be problematic since studies may not be included in the adjusted analyses, leading to a situation analogous to publication bias.

Data from studies may be included in a meta-analysis at the individual participant level and analyses may proceed which are similar to those used to analyse multi-centre studies. However, data may be missing at the individual participant level where a person in a

study does not have a recorded value. If this is related to either the study design or the intervention under investigation, concerns are raised since the exclusion of these participants may over-estimate the efficacy of the intervention (Sutton *et al.* 1998).

1.6.4 English language bias

Studies performed in a country where English is not the first language are more likely to be published in non-English language journals if the results are inconclusive whereas studies which show benefit towards an intervention are more likely to be published in an English language journal (Moher *et al.* 1996). This finding is known as English language bias, which can lead to data being missed through inadequate search methods. This type of bias needs to be minimised by searching journals irrespective of their publication language; however, this may not always be achieved since many of the electronic databases are predominately English language based.

1.6.5 Multiple publication bias and selection bias

Two other types of publication bias exist; the first relates to multiple publication bias, which can be a serious problem in meta-analysis. It occurs when a study is included in the meta-analysis more than once because it has been published in different forms. Sometimes it may be unclear as to whether the identified paper is an offshoot of the main paper.

The second type of bias is known as selection bias which occurs when only a selection of the relevant studies are chosen and their

inclusion is dependent upon their results, for example, only studies with beneficial effects are included.

In light of these potential biases which can exist in any meta-analysis, it has been recommended that at least two independent reviewers should decide which studies are included following a thorough literature search; similarly, the data should be extracted from the primary studies to minimise errors (Deeks *et al.* 1996).

1.6.6 Assessment of study quality

Issues relating to the quality of each study included in the meta-analysis is an important factor to assess, since the quality of the studies have been found to have an impact on the results generated from a meta-analysis (Chalmers *et al.* 1981; Schulz *et al.* 1995).

Unfortunately at present, there is no consensus on how to determine the quality of a trial; various checklists have been published (Moher *et al.* 1995; Begg *et al.* 1996) although they do not give consistent results with each other. The scales differ considerably in the components they include to assess quality and in the weights they assign to each component.

An example of this is a scale developed by Jadad and colleagues. This scale tends to gives more weight to how well the results were reported rather than to the methodological quality of the design of each study (Jadad *et al.* 1996). Therefore, this would imply that a well-designed randomised clinical trial that is poorly reported will receive relatively less weight than an inadequately designed study that mentions a statement about withdrawals and dropouts but does

not analyse the data according to the intention-to-treat principle, which is clearly debatable.

It has been also suggested (Schulz *et al.* 1995) that only double-blind randomised controlled trials should be included in systematic reviews (and meta-analyses) to improve the quality of included studies. However, this would imply that the results from these studies are more reliable than single- or un-blinded studies. Excluding these studies may invalidate the results; however, including these studies may produce biased results if the studies are of low quality. Similarly, it has been suggested that meta-analysis should only include trials which have used the intention to treat principle, since withdrawals and dropouts may be related to the treatment or intervention received in the trial.

The Cochrane collaboration have devised recognised criteria to assess the quality of randomised controlled trials (The Cochrane Collaboration 2003). These criteria are based on assessing whether there is a low, moderate or high risk of bias which could invalidate the results from the study. Biases can be grouped into four main areas; selection bias, performance bias, attrition bias, and detection bias.

Selection bias deals with whether there were systematic differences between how the participants were selected for inclusion into the study and what method was used to conceal the allocation of treatment. Adequate methods used for concealment of allocation include centralised randomisation; pre-coded identical containers administered directly to participants; protected onsite computer

system; and sequentially numbered sealed opaque envelopes. However, methods which include alternation, e.g. using case records numbers or dates of birth, are inadequate since the investigators will be aware of which patient received which intervention. Double blinding of the investigator and patients to the intervention assignment is the gold-standard although this can not always be achieved. For example, in the case of investigating the efficacy of a surgical intervention compared to non-surgical intervention, the patient and investigator will be aware of which they received/administered.

Performance bias is related to whether there were differences in the standard of care or treatment between the two groups external to the intervention under study. The levels of care between the two groups should remain comparable so that the efficacy of the intervention can be studied. Attrition bias deals with whether there were differences in the rates of drop-outs between the two groups. Obviously, if the drop out rate is large for those receiving the active intervention, then this has implications for the practical application of the intervention.

Lastly, detection bias deals with how the assessments were conducted and whether they were consistent between the two intervention groups. The outcome assessor should be blinded to the allocation of intervention to ensure that measurement of the outcome is not biased by prior knowledge.

It has been found that the results of a study lacking adequate concealment of allocation and double blinding may result in an over-

estimate the intervention efficacy (The Cochrane Collaboration 2003). However, these criteria rely on the publication giving sufficient details with regards to these areas; where insufficient details are given the investigators should be contacted for more details.

Incorporating only randomised controlled trials in meta-analyses may eliminate the majority of known biases associated with study designs; however, other biases such as performance and attrition bias may be present.

Smaller studies are usually conducted and analysed with less methodological rigour than larger studies and tend to show larger effects than would be seen with the larger studies (Kjaergard *et al.* 2001). Therefore, it has been proposed that the sample sizes of the studies could be categorised and analyses could be presented stratified by the variable to allow for an adjustment for quality (Bath *et al.* 1998).

1.6.7 Misleading meta-analyses

Misleading meta-analyses have been published identifying a beneficial intervention effect which has not been replicated in subsequent clinical trials.

An example is a meta-analysis of clinical trials which assessed the efficacy and safety of magnesium infusions in acute myocardial infarction (Yusuf *et al.* 1993). The meta-analysis concluded that magnesium reduced all cause mortality. From these conclusions a large randomised controlled trial was instigated called the ISIS-4 trial (ISIS-4 (Fourth International Study of Infarct Survival) Collaborative

Group 1995). The beneficial treatment effect seen in the meta-analysis was not apparent in this subsequent clinical trial.

A review of these findings suggests that the results from this misleading meta-analysis were subject to selection bias and possible heterogeneity (Egger 1998). It was thought that the studies included in the meta-analysis were not representative of all available evidence, e.g. publications that were inconclusive or found a detrimental treatment effect were under-represented. Misleading meta-analyses may also result from inappropriately combining heterogeneous studies; and it has been suggested if the studies are too heterogeneous, a meta-analysis should be avoided.

1.7 Aims of the thesis

Although meta-analyses can suffer from several problems, this has not prevented them from being performed widely. Until recently, many meta-analyses have used summary data extracted from publications or, where data were not extractable, supplied by the authors of the study. Meta-analyses were then performed using data at summary level. Occasionally, further more in-depth analyses have been performed to ascertain why there are differences between the results from each study by assessing the effect predictors may have on the results. This was achieved by either grouping the studies into categories, such as high versus low study quality; or by modelling a predictor using simple linear regression techniques, such as average age of patients in the study.

More recently, meta-analyses have been performed which use the raw data from the studies. The analyses used are similar to those performed in multicentre studies, where the data are stratified by study. This type of model may allow for in-depth analyses to be performed to assess which predictors; either at the patient level and/or the study level, influence the efficacy of the intervention using multivariate regression analyses.

However, the majority of the published meta-analyses only present the overall efficacy results for an intervention. Although it is now becoming more common for additional analyses to be performed which investigate why there are differences between the results from the studies, it is still rare for them to assess the level of unexplained variation between the results.

This thesis will describe the types and models that can be used in meta-analyses both at summary and individual patient level (Chapter 2 and 3). These models will be exemplified using data from three meta-analyses of randomised controlled trials in stroke medicine (Chapter 4).

The first meta-analyses is concerned with assessing the efficacy of vasoactive drugs in acute stroke to see whether altering blood pressure in patients with acute stroke is safe and effective in reducing the risk of death, and death or dependency; and to determine the effects of vasoactive drugs on systolic blood pressure (Chapter 5). The next meta-analysis is concerned with assessing the efficacy of occupational therapy in the community setting to see whether giving

occupational therapy in this setting is associated with a range of functional scores, such as extended activities of daily living and leisure participation; and to assess whether the differences between the findings of the individual trials are related to trial and/or patient level factors (Chapter 6). The final meta-analysis is concerned with assessing whether the combination of aspirin and dipyridamole given to stroke patients as a secondary prevention treatment is more effective in reducing vascular events than compared to the mono therapies of dipyridamole alone or aspirin alone; and also to identify which patients should be targeted for the combination treatment (Chapter 7).

Next, this thesis aims to assess whether there are clinically important differences between the results from using methodologies based on summary and individual patient data within these meta-analyses; and whether there are distinct advantages to using individual patient data as compared to summary data methods.

CHAPTER 2

SUMMARY DATA AND INDIVIDUAL PATIENT DATA METHODOLOGIES IN META-ANALYSIS

2.1 Introduction

A variety of mathematical methods may be employed to combine data from different sources depending on the type of outcome data available. These methods range from being very simple to perform through to requiring complicated mathematics. This chapter will provide an overview of simple techniques for combining data; and describe how data from studies can be summarised and combined together to yield an overall result. A variety of statistical methods which combine data from studies using summary and individual patient data methodologies will be described from both classical and Bayesian perspectives. Additionally, several methods for assessing publication bias will also be described and discussed.

2.2 Simple methods for combining studies

Very simple methods exist for combining the conclusions from studies; these include methods such as vote counting and combining p values. Vote counting is a technique where the direction of the result from each study is considered. The magnitude of the result is not taken into account and it is usual for the significance of the result to be ignored as well. It has been commented that this method can lead to extremely misleading results due to a lack of power (Greenland 1987), and can miss important sources of why the studies yield differing results. Therefore it has been suggested that these methods be used only as a prelude to a more detailed analysis (Greenland 1987).

An alternative method involves combining the significance level of the effect estimate (p value) from each study. Both this method and the vote counting method are easy to employ but suffer from common problems. Specifically, these methods do not yield an estimate for the overall intervention effect. Therefore these methods are rarely used unless other more complicated methods are not possible, which is usually related to a lack of available data required for the other techniques.

The more complicated techniques used in conventional meta-analyses have the advantage over these simpler methods in that they yield an overall estimate for the intervention effect. Also, the precision with which this result is estimated may be presented in the form of confidence intervals. Conventionally 95% confidence intervals are presented with the point estimate.

2.3 Single study effect estimation

In conventional meta-analyses the data from each study is summarised into a single summary measure. This thesis will concentrate on dichotomous outcomes and continuous outcomes as they are pertinent to the examples used in the exemplified meta-analyses.

2.3.1 Dichotomous outcome measures

The most commonly used outcome measure for summarising dichotomous data is the odds ratio (OR); other outcome measures that may be used for dichotomous data are the risk difference and the

risk ratio (relative risk). The odds ratio is used in a wide range of studies and trials; additionally it may be used in retrospective and cross-sectional studies which aim to assess associations rather than differences. The odds ratio has mathematical advantages over the other two measures which include symmetry with respect to 'successes' and 'failures', and the values it takes are unrestricted between zero and infinity (Engels *et al.* 2000). Also for more complicated analyses such as logistic regression the odds ratios is the only measure of association that can be used which does not require special assumptions. However, the main disadvantage of the measure is that it is quite complicated to interpret, even using the simplest analyses, as compared to the relative ease of interpretation for the risk difference and risk ratio. However, if the outcome measure is rare then the odds ratio may be interpreted as a risk ratio thereby simplifying the interpretation of the odds ratio. Although the examples used within this thesis will estimate the effects using odds ratios, these will be interpreted as risk ratios since the outcome measures used are relatively rare within each of the studies.

The risk difference (RD) is the simplest measure to use and interpret since an estimate of the percentage of patients who would directly benefit from the intervention or treatment may be calculated. It is appealing to a wide range of researchers since it reflects both the underlying risk of the control group and the reduction in risk associated with the intervention. However, it does not give an estimate that is relative (Engels *et al.* 2000) which is problematic

since it has been found that relative effect measures give more consistent results than absolute measures (The Cochrane Collaboration 2003). Additionally, the RD has been found to give more emphasis to the results from the studies with lower prevalence rates, when used in a meta-analysis.

The risk ratio (RR) gives an estimate of relative benefit of effect; this outcome measure is intuitively appealing since it compares the probabilities from two groups in terms of their proportionate difference (Fleiss *et al.* 1994). It is commonly used in observational studies such as case-control and cross-sectional studies, but is also used in clinical trials due to its simple interpretation. However, caution must be used in its interpretation since its value is asymmetrical meaning that the reciprocal of the RR for benefit is not equal to the relative risk for harm.

The remainder of this section will focus on the methods for the odds ratio due to its advantages.

2.3.1.1 Example of calculating the odds ratio

Data from a single two-armed study may be grouped into a 2x2 table for each dichotomous outcome measure.

	Event	Event free
Intervention Group	a_i	b_i
Control Group	c_i	d_i

Table 2.1 Binary data from the *i*th study

From Table 2.1, if we assume there are two treatment groups. The number of patients having an event is $(a_i + c_i)$ and the number of patients whom are event free is $(b_i + d_i)$. The odds ratio (OR) is the

ratio of the odds of an event from each group and may be calculated using a maximum likelihood estimate (Equation 2.1).

$$\hat{\mathrm{OR}}_i = \frac{a_i d_i}{b_i c_i} \tag{2.1}$$

The estimate of the odds ratio is usually expressed in logarithmic form since this should provide a measure that is normally distributed and symmetrical about its null value (Fleiss *et al.* 1994). An asymptotic estimate of the large sample variance of the logarithm of the odds ratio may be estimated as shown in Equation (2.2).

$$\mathrm{Var}\left(\ln \hat{\mathrm{OR}}_i\right) = \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i} \tag{2.2}$$

A 95% confidence interval may be calculated by assuming that the estimate of the log odds ratio is normally distributed (Equation 2.3).

$$\ln\left(\hat{OR}_i\right) \pm 1.96\sqrt{\mathrm{Var}\left(\ln \hat{OR}_i\right)} \tag{2.3}$$

It has been found that these methods of calculating the odds ratio work well however, sparse data in the 2x2 table can cause instability in the estimates yielded (Sankey *et al.* 1996). Also, where zero events are seen in either the control group or in both the intervention and the control groups; this leads to the study being excluded from the meta-analysis since an estimate can not be calculated. Whitehead and Whitehead comment that it is appropriate for the study to be dropped from the meta-analysis since the data does not provide information on the magnitude of the intervention effects (Whitehead *et al.* 1991). However, this does not take into account the size of the study where

the information yielded from a small study with no events is not equivalent to that from a large study with a zero event rate. Therefore, it has become common for studies with sparse event data or where groups have zero events for an arbitrary value of 0.5 to be added to each of the cells in the 2x2 table. Adding an arbitrary value to each of the cells has been shown to reduce bias in the estimation of the odds ratio (Sankey *et al.* 1996).

2.3.1.2 Other methods for estimating the odds ratios

Adaptations to the maximum likelihood estimate for the odds ratio have been developed to handle different types of problems within the data sets. The Mantel-Haenszel method (Mantel *et al.* 1959) was initially designed to calculate the odds ratio for individual case-control studies; subsequently it has been shown that it can be applied to most data sets (Robins *et al.* 1986a; Robins *et al.* 1986b; Hasselblad *et al.* 1995). However, the Mantel-Haenszel method has been found to not compensate for sparse data within the 2x2 table and in the case of zero cells or very small frequencies this method should not be used.

An adaptation by Peto and colleagues was devised, which is based on a modified likelihood, to overcome the problems associated with the Mantel-Haenszel method where sparse data exist (Peto *et al.* 1977; Yusuf *et al.* 1985). This method however, can produce biased estimates for the intervention effect when the odds ratio is very large in magnitude, or when there is a serious imbalance in the numbers

between the intervention and control groups and hence should not be used in these circumstances (Fleiss 1993).

Other more complicated techniques also exist for calculating the odds ratio which includes a method which uses the maximum likelihood method to yield an unconditional estimate using iterative formulae (Emerson 1994).

2.3.2 Continuous outcome measures

Continuous data from an individual study may be summarised; the most common methods include the absolute difference in means and the standardised difference between means.

2.3.2.1 Absolute difference in means

An estimate of absolute difference in means may be calculated when a common normally distributed outcome measure is used across the studies. The difference in means estimate for each study, $\hat{\delta}_{i_{WEI}}$, is a maximum likelihood estimate of the absolute difference between the means for the control, μ_{c_i}, and intervention, μ_{t_i}, group for each study (Equation 2.4).

$$\hat{\delta}_{i_{WEI}} = \mu_{t_i} - \mu_{c_i} \tag{2.4}$$

The variance for the absolute difference in means may be estimated using Equation (2.5) where σ_i^2 is the individual study variance and n_{c_i} and n_{t_i} are the sample sizes in the control and intervention groups, respectively.

$$\mathrm{Var}\left(\hat{\delta}_{\mathrm{i_{WEI}}}\right) = \sigma_i^2\left(\frac{1}{n_{t_i}} + \frac{1}{n_{c_i}}\right) \tag{2.5}$$

The individual study variance, σ_i^2, may be estimated using a variety of methods. The most common method assumes a common variance and uses the pooled within-group variance, s_i (Shadish *et al.* 1994). Where s_{c_i} and s_{t_i} are the standard deviations for the control and intervention groups, respectively (Equation 2.6).

$$s_i = \sqrt{\frac{(n_{t_i}-1)s_{t_i}^2 + (n_{c_i}-1)s_{c_i}^2}{n_{t_i}+n_{c_i}-2}} \tag{2.6}$$

95% confidence intervals for the effect estimate may be calculated as described in Equation (2.7).

$$\hat{\delta}_{i_{WEI}} \pm 1.96\sqrt{\mathrm{Var}\left(\hat{\delta}_{\mathrm{i_{WEI}}}\right)} \tag{2.7}$$

The absolute difference in means is used when the all the studies have assessed an outcome measure using the same scale and units, and therefore has an advantage that the estimate for the intervention effect may be described in the same units as it was measured and the interpretation of the estimate may be applied directly to the outcome measure.

2.3.2.2 Standardised difference between means

The standardised mean difference may be used when different measures have been used to assess the outcomes of the patients across the studies. The maximum likelihood estimate of the

standardised difference between means, $\hat{\delta}_{i_{STD}}$, is given in Equation (2.8).

$$\hat{\delta}_{i_{STD}} = \frac{\mu_{t_i} - \mu_{c_i}}{\hat{\sigma}_i} \tag{2.8}$$

The standard deviation $\hat{\sigma}_i$ is a maximum likelihood estimate, however, this estimate is known to be biased and therefore alternative sources have been suggested. Glass recommends using the standard deviation from the control group (Glass 1976), however this has also been found to be biased (Hedges *et al.* 1985). If it is reasonable to use a common variance for the standardised mean difference (when the variance between the two groups are similar within a study), Hedges and Olkin recommend using the unbiased estimate for the standard deviation, s_i (Equation 2.6) (Hedges *et al.* 1985).

If the variances in the active and control groups can not be assumed to be similar then Rosenthal suggest transforming the data using logs or square roots; to attempt to make the variances more similar, however this assumes that the original data is available (Rosenthal 1994).

The variance for the standardised difference between means may be estimated using various formulae. If the underlying data can be assumed to be normally distributed then the most robust method is the overall conditional variance, $\mathrm{Var}\left(\hat{\delta}_{i_{STD}}\right)$ (Shadish *et al.* 1994).

$$\mathrm{Var}\left(\hat{\delta}_{i_{STD}}\right) = \frac{n_{t_i} + n_{c_i}}{n_{t_i} n_{c_i}} + \frac{\hat{\delta}^2_{i_{STD}}}{2\left(n_{t_i} + n_{c_i}\right)} \quad (2.9)$$

The standardised mean difference method may be used if the data are approximately normally distributed, therefore transformations of the outcome data may be required to achieve normality.

Interpreting the standardised difference between means is difficult because a common unit has not been used across all of the studies and the effect estimate was transformed to be dimensionless; hence the confidence intervals of the intervention effect are used primarily to assess significance of the effect.

The outcome measure may be transformed to achieve approximate normality if there is evidence of non-normality for the absolute difference in means or the standardised difference between means. It has been suggested that the logarithm of the outcome measure may be used where there is evidence that the outcome measure is not normally distributed (Hasselblad *et al.* 1995), however in practice this is rarely performed.

2.4 Classical approaches to meta-analysis using summary data

The choice of model for the analysis is dependent upon the question that is being investigated. Most meta-analyses attempt to answer how well an intervention works mainly because this is what the individual trials were assessing. In this case, where the question is whether the intervention could ever achieve benefit then a model should be used which evaluates a single underlying effect estimate (Petitti 2001). This

model assumes that the studies included in the meta-analysis are homogeneous in nature designed to answer the same question using patients from the same population (Lau *et al.* 1998).

However, if the question is related to whether the intervention will produce benefit "on average" then a model should be used which allows for the studies to be heterogeneous in their study design and patient populations. This allows for the intervention effect estimate to have extra variability and yield a global effect estimate (Petitti 2001).

2.4.1 Fixed effect methods

A fixed effect model will estimate the single true underlying effect by assuming that the variation in the intervention estimates from the trials are due to sampling variation alone (Fleiss 1993). When combining effect estimates from k individual studies, there will almost certainly be some degree of difference between the point estimates. If the point estimates vary by a small measure, then the studies are homogeneous and it may be appropriate to consider using a fixed effect model (Equation 2.10); where θ is the true underlying intervention effect, $\hat{\theta}_i$ is an estimate of the underlying intervention effect associated with the i'th study where $i = 1,....,k$; and the error terms are normally distributed random variables with a mean zero and variance ξ_i^2; and it is assumed that ξ_i^2 is equal to $\mathrm{Var}(\hat{\theta}_i)$

$$\hat{\theta}_i = \theta + \varepsilon_i \qquad (2.10)$$

$$\varepsilon_i \sim N(0, \xi_i^2)$$

An interaction between the study and the intervention effects could be included in the fixed effect model. This term would allow for the intervention effects to vary between studies and is included in the model as a fixed effect. However, there is usually a lack of power to be able to estimate the effect size of the interaction term especially when a small number of studies are present in the meta-analysis and so is rarely used in practice.

Several methods have been proposed which combine the estimated effect sizes from the individual studies. The most common method used is weighted least squares regression, the weightings for this method are calculated using the inverse variance method (Birge 1932; Cochran 1937).

2.4.1.1 Weighted least squares regression

In an ordinary least squares regression the observations are not weighted, therefore this assumes that a constant weight is given to each study, thereby implying that each study is of equal importance. The method also assumes that a common variance exists across studies, this is referred to as homoscedasticity (Neter *et al.* 1977). In practice this may be unreasonable, since it is inevitable that a meta-analysis will include studies with differences in design, sample size and patient population. Therefore, if the intervention estimates were combined and the average taken, misleading results would arise, since the estimates would have varying degrees of precision due to different sample sizes. The precision to which a study estimates its own variation depends on the sample size of the study, each study

will have its own sampling error (DerSimonian *et al.* 1986). Generally, larger studies will have a more precise estimate of the intervention effect than smaller studies; therefore ordinary least squares regression method will be inadequate to incorporate the non-constant variances across the studies.

The weighted least squares regression allows for non-constant variance across the studies (Neter *et al.* 1977). As the name suggests, this method incorporates a weighting function into the regression model, where a specific weight is assigned to each study which reflects the level of the precision the study provides. The weights assigned to each of the studies can vary depending on which assumptions are made. A conventional method of weighting that is used is called the inverse variance method (Birge 1932; Cochran 1937).

2.4.1.2 Inverse variance method

The inverse-weighted method, sometimes referred to as Woolf's method, was first described in the 1930's (Birge 1932; Cochran 1937) and remains the most commonly used method due to its relative simplicity (Woolf 1955). This method may be used to combine summary measures whilst weighting the results of each trial, denoted by w_i. The weight is conventionally the reciprocal of the variance associated with the individual study, $\mathrm{Var}\left(\hat{\theta}_i\right)$ (Equation 2.11).

$$w_i = \frac{1}{\mathrm{Var}\left(\hat{\theta}_i\right)} \tag{2.11}$$

Describing the weighting in this manner gives more weight to larger studies than smaller studies. This seems intuitively correct since larger studies will generally have a smaller within-study variance, than compared to smaller studies, hence will tend be more precise in their estimation of the intervention effect size.

A pooled estimate of the intervention effect, $\hat{\theta}$, may be calculated using the following Equation (2.12), where $\hat{\theta}_i$ is an estimate of the intervention effect from the i'th study.

$$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i \hat{\theta}_i}{\sum_{i=1}^{k} w_i} \tag{2.12}$$

The large sample asymptotic variance of the pooled estimate for the intervention effect may be estimated as the reciprocal of the sum of the weights calculated from the studies.

$$\mathrm{Var}\left(\hat{\theta}_i\right) = \frac{1}{\sum_{i=1}^{k} w_i} \tag{2.13}$$

It is assumed that the asymptotic pooled variance is normally distributed, however the estimate for the variance may be imprecise if the number of studies included in the meta-analysis is small (Li *et al.* 1994).

Li and colleagues have found that the weights used in the above method are sensitive to unequal variances between intervention groups from each study, and become biased when the sample sizes in each study are not large (Li *et al.* 1994). They proposed a method for calculating the pooled sample variance based on the sample size

for each study which takes into account the total number of patients in the studies. This method was found to be more robust than previous methods with regards to the variation of the sample variances and the sample sizes within the studies, and is not sensitive to any individual result estimate from a particular study. It has also been suggested that the method based on sample size should be used when the sample variances are not homogeneous within a study (Li *et al.* 1994).

Potential problems may result if fixed effects models are used; there are concerns that the simple weighting scheme may overweight the results from the larger studies and not truly reflect the differences between the studies (Pocock *et al.* 1981). Also, the results from the analysis can only be applied to the studies used in the meta-analysis. It may be undesirable to use a fixed effect model if there is a considerable degree of disagreement between the estimates for the intervention effect from the studies. This type of disagreement between the study estimates is called heterogeneity. An assessment of heterogeneity should be performed to ascertain whether there appears to be statistical variation between the studies.

2.4.2 Random effects methods

When there is a considerable difference in the estimates for the intervention effect between the k studies, $(\theta_1.....\theta_k)$, it may be more appropriate to compensate for the variation by assuming that they are a sample of independent observations from $N(\theta,\tau^2)$ and using a

random effect model (DerSimonian *et al.* 1986); where the u_i are the random effect terms with mean zero with a variance of τ^2 for $i = 1,.....,k$; and u_i and ε_i are assumed independently distributed.

$$\hat{\theta}_i = \theta + u_i + \varepsilon_i \qquad (2.14)$$

$$u_i \sim N(0,\tau^2) \text{ and } \varepsilon_i \sim N(0,\xi_i^2)$$

The random effects model allows estimates for the intervention effect to vary from study to study; where the studies are assumed to be from a random sample of studies which follow a specified distribution. Conventionally the normal distribution has been used to accommodate the variation (Sutton *et al.* 1998). In the random effect model, the estimated variances of the intervention effects for the individual studies, $Var(\hat{\theta}_i)$, contain two components, an estimate of the conditional variation, $\hat{\xi}_i^2$, and an estimate of the random variation, $\hat{\tau}^2$ (Equation 2.15).

$$Var(\hat{\theta}_i) = \hat{\tau}^2 + \hat{\xi}_i^2 \qquad (2.15)$$

When performing a random effects analysis where the outcome is a dichotomous measure, the standard error for the intervention effect will be too conservative (Greenland 1987) unless an adjustment is made to fix the residual variance at one. This ensures that the residual heterogeneity is accounted for in the model as an additive effect, and not a multiplicative effect (Thompson *et al.* 1999).

The additive effect for the random effect is incorporated into the analysis through the weights; this allows for the more variation to be

estimated between the studies than would be seen using a fixed effect model. However, this has been criticised since a large value for the additive effect would result in larger studies having a smaller relative weight in a random effects model than a fixed effect model; and in contrast, the relative weights given to the results from the smaller sized studies will be larger (Leonardi-Bee 2000). Since there is the tendency for these studies to yield outlying results; these two issues could overtly influence the pooled intervention estimate and lead to a spurious result.

The concept of allowing for the intervention effect estimates to vary according to a predetermined distribution means that the results can be generalised to other trials. However, if each study estimated exactly the same true effect size, then the estimated random effects variance would be equal to zero. The variation between the effect sizes across all of the studies would be attributed to sampling variance only; and the model would then reduce to the fixed effect model.

As mention previously, an estimate of the heterogeneity, denoted as $\hat{\tau}^2$, is required; a variety of classical techniques have been proposed, these include the method of moments based on weighted and un-weighted least squares regression methods, the maximum likelihood method and the restricted maximum likelihood method.

2.4.2.1 Method of moments, un-weighted and weighted

Two non-iterative methods for estimating the between study heterogeneity in a simple model have been proposed (DerSimonian

et al. 1986); which are based on using the method of moments approach and are relatively simple to calculate. The first uses an un-weighted approach (DerSimonian *et al.* 1986; Shadish *et al.* 1994), and so can be thought of as allowing all of the studies to have equal sampling variances and thus equal weightings. The methodology for calculating the estimate of heterogeneity is divided into three parts. Firstly, an ordinary un-weighted estimate of the variance for the effect size, $s^2\left(\hat{\theta}_{UNW}\right)$, is calculated as shown in Equation (2.16), where there are k studies included in the meta-analysis where $i=1,.....,k$; $\hat{\theta}_i$ is the estimate of the intervention effect, and $\hat{\theta}_{UNW}$ is the un-weighted estimate for the intervention effect, and is specified as $\frac{\sum_{i=1}^{k}\hat{\theta}_i}{k}$.

$$s^2\left(\hat{\theta}_{UNW}\right)=\sum_{i=1}^{k}\frac{\left(\hat{\theta}_i-\hat{\theta}_{UNW}\right)^2}{k-1} \tag{2.16}$$

Equation (2.16) can be rearranged to form Equation (2.17).

$$s^2\left(\hat{\theta}_{UNW}\right)=\left[\sum_{i=1}^{k}\hat{\theta}_i^2-\frac{\left(\sum_{i=1}^{k}\hat{\theta}_i\right)^2}{k}\right](k-1)^{-1} \tag{2.17}$$

The un-weighted sample estimate of the variance for the effect size is an unbiased estimate of the expectation (Equation 2.18).

$$E\left[s^2\left(\hat{\theta}_{UNW}\right)\right]=\tau^2+\frac{\sum_{i=1}^{k}\sigma^2\left(\hat{\theta}_i\mid\theta_i\right)}{k} \tag{2.18}$$

Substituting an estimate of the $\mathrm{Var}(\hat{\theta}_i)$ for $\sigma^2(\hat{\theta}_i \mid \theta_i)$, and rearranging yields an estimate for the between study heterogeneity (Equation 2.19).

$$\hat{\tau}^2 = s^2(\hat{\theta}_{UNW}) - \frac{\sum_{i=1}^{k} \mathrm{Var}(\hat{\theta}_i)}{k} \qquad (2.19)$$

The estimate for $\mathrm{Var}(\hat{\theta}_i)$ is dependent on which effect estimate is chosen. In the case where the log odds ratio is used the large sample variance is commonly used (Equation 2.2). For the case where the mean difference is used, it is based on using a common variance, σ_i^2 (Equation 2.5).

The second method is known as the weighted method of moments (DerSimonian *et al.* 1986; Shadish *et al.* 1994) since it uses the estimate for the unconditional sample variance from a weighted regression and an estimate from Cochran's homogeneity statistic, Q. The expectation for Q is described in Equation (2.20).

$$E[Q] = \left[\sum_{i=1}^{k} w_i - \left(\frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i} \right)^2 \right] \tau^2 + (k-1) \qquad (2.20)$$

Substituting Q for its expectation and rearranging yields an estimate for τ^2 as shown in Equation (2.21). If Q is less than $k-1$ then $\hat{\tau}^2$ is replaced with zero, so that the magnitude of the random effect for the summary estimate will not exceed the magnitude of the fixed effect (Higgins *et al.* 2002).

$$\hat{\tau}^2 = \frac{Q-(k-1)}{\sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}} \qquad (2.21)$$

Both of these estimators have not been fully studied with regards to their advantages (Shadish *et al.* 1994). Even though the weighted method appears to be less complex to calculate than iterative methods, it only produces non-zero estimates for the between study heterogeneity when the value for Q is greater than the critical value of χ^2 on *k-1* degrees of freedom, as detailed under the null hypothesis.

2.4.2.2 Maximum likelihood methodology

The method of maximum likelihood may be used to estimate the between study variability, where an iterative scheme is implemented (Hardy *et al.* 1996). A conventional weighted least squares regression is performed where the estimate for the heterogeneity is set to zero to yield an initial estimate for the pooled intervention effect. Then, $\hat{\theta}$ is fixed and an initial estimate for the heterogeneity is calculated (Equation 2.22).

$$\hat{\tau}^2 = \frac{\sum_{i=1}^{k} w_i^2\left[\left(\hat{\theta}_i - \hat{\theta}\right)^2 - \mathrm{Var}\left(\hat{\theta}_i\right)\right]}{\sum_{i=1}^{k} w_i^2} \qquad (2.22)$$

The initial estimate of $\hat{\tau}^2$ is then used to re-calculate the weights (Equation 2.23).

$$w_i^* = \frac{1}{\left(\mathrm{Var}\left(\hat{\theta}_i\right) + \hat{\tau}^2\right)} \tag{2.23}$$

The cycle is then repeated to generate new estimates for the summary statistic parameter using a weighted least squares regression where the weights w_i^2 are replaced with w_i^{*2}. The values are then inputted into Equation (2.22) and a new estimate of $\hat{\tau}^2$ is calculated. The cycle is repeated until $\hat{\tau}^2$ converges.

This method has been found to yield results which are smaller in magnitude for the between trial heterogeneity as compared to other methods (Turner *et al.* 2000), since this method uses the observations to calculate the log-likelihood function and is not based on the residual terms. The process of convergence may be slow for this estimator and the reliability may be poor in practice (Thompson *et al.* 1999). Therefore the maximum number of iterations specified in the program needs to be large enough to ensure that convergence has been reached and has not stopped prematurely due to the maximum number of iterations being exceeded.

2.4.2.3 Restricted maximum likelihood methodology

Restricted maximum likelihood may be used to estimate heterogeneity (Thompson *et al.* 1999). The scheme used to estimate the heterogeneity is similar to the method used in ML where an iterative cycle is used to achieve convergence. Equation (2.22) is slightly modified to partially allow for the pooled intervention effect and heterogeneity being estimated from the data (Equation 2.24).

$$\hat{\tau}^2 = \frac{\sum_{i=1}^{k} w_i^2 \left(\frac{k}{k-1}\right)\left[\left(\hat{\theta}_i - \hat{\theta}\right)^2 - Var\left(\hat{\theta}_i\right)\right]}{\sum_{i=1}^{k} w_i^2} \qquad (2.24)$$

The REML method has been found to be less biased than ML since REML uses a modified likelihood equation to estimate heterogeneity and the fixed effects (Brown *et al.* 1994). The difference in magnitude for the between trial heterogeneity between the ML and REML methods has been estimated about 10% (Turner *et al.* 2000); in contrast, REML and weighted MM techniques have been found to yield similar results for the heterogeneity, hence also the fixed effects, since the weighted MM also yields an unbiased estimate (DerSimonian *et al.* 1986).

Both the ML and REML approaches are based on the assumption of normality of the random effects. However, the two methods are not sensitive to moderate deviations from normality (Raghunathan *et al.* 1993). Specifying the random effects distribution as non normal has been studied using a simulation study. Both the *t*- and log normal distributions produced similar estimates for the intervention effect however, the confidence intervals were slightly more conservative for small values of heterogeneity (Raghunathan *et al.* 1993).

2.5 *Meta-analysis using individual patient data*

Conventionally, data from a meta-analysis are analysed using summary statistics estimated from each of the studies; however it is possible to use the individual data from each participant within a study.

Meta-analyses which use individual patient data (IPD) have been quoted as the yardstick against which the quality of other systematic reviews of randomised controlled trials should be measured (Chalmers *et al.* 1993). The techniques used when performing IPD meta-analyses allow for many of the problems associated with using data from published articles, and a few of the problems associated with using summary data to be overcome. These include allowing for detailed data checking to be performed on each study and for analyses of data to be performed using consistent methods across all of the studies.

However, there are some disadvantages associated with using IPD. These mainly include problems with obtaining the raw data from the investigators who are unwilling or unable to supply the data. It has been suspected that some trialists may not want to share their data in a meta-analysis because it may dilute their results of a positive intervention effect (Sutton *et al.* 1999). It has also been acknowledged that high costs and time are involved in building a database (Stewart *et al.* 1995). It has been estimated that a meta-analysis based on IPD costs at least five time more than a meta-analysis based on summary data (Steinberg *et al.* 1997).

At present there appears to be little evidence that the gains from performing a meta-analysis based on individual patient data are worthwhile and justified (Sutton *et al.* 1999). Although since the late eighties, much collaboration has been achieved between investigators and much individual patient data has been shared to try

to answer pressing questions that could not be answered alone from using meta-analysis with summary data.

2.6 Classical approaches to meta-analysis using individual patient data

Several approaches have been advocated to analyse data in a meta-analysis of individual patient data. A simple but naïve method has been used where the data are analysed without adjusting for study effects (Man-Son *et al.* 1995). This simple method does not allow for the patients to be clustered within their study and hence the variability between the studies is not adequately described.

An alterative simple method has involved adjusting for confounding factors using regression techniques within each study, followed by a conventional meta-analysis on the adjusted summary estimates (Nicolucci *et al.* 1996). This method is inefficient but does allow for problems associated with missing variables within studies to be overcome, since each study may be adjusted for the covariates recorded, and the covariates may vary between studies.

Traditional mixed effect models (Searle 1971) or multilevel models (Goldstein *et al.* 2000) may be used to analyse IPD meta-analyses. Within the individual patient data meta-analysis, the study may be specified as either a fixed effect or a random effect; where the choice of method depends on the question to be answered (see Section 2.4).

2.6.1 Fixed effect methods

The methodology used for analysing individual patient data meta-analysis depends on the type of outcome data and in the case where

either dichotomous or continuous data is used; the analysis performed is similar to a conventional regression analysis.

2.6.1.1 Dichotomous outcome measures

The outcome for a patient is denoted as y_{ij} and is from a random variable Y_{ij} which has a binomial distribution with parameter π_{ij} and a denominator of 1. If π_{ij} is the probability of an event for patient j in study i where $i=(1,....,k)$, then y_{ij}=1 if the event occur and 0 if they are event-free (Equation 2.25).

$$\mathrm{Y}_{ij} \sim Bin(\pi_{ij}, \pi_{ij}(1-\pi_{ij})) \qquad (2.25)$$

Using a logit link function leads to a linear regression model where the parameter α is the intercept, β_{0i} is the study effect which is constrained so that β_{0k} is equal to zero, z_{ij} is a dummy variable for intervention assignment, and β_1 is the pooled log odds ratio of an event on intervention as compared to control (Equation 2.26).

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha + \beta_{0i} + \beta_1 z_{ij} \qquad (2.26)$$

A logistic regression analysis is performed to provide a maximum likelihood estimate of the intervention effect. Unlike the methods used when the summary outcome is based on dichotomous data where an assumption of normality is required for the log odds ratio; this method allows for the data in its binomial form to be used directly to estimate the log odds ratio.

Whitehead has compared using individual patient data as compared to combining summaries, and has found that the results from the two methods were similar and any differences in the estimates of the parameters were due to the normality assumption of the log odds ratio from the summary data model (Whitehead 2002).

Turner and colleagues have investigated the advantages of analysing the log odds ratio using summary and individual patient data analysis methods using two examples (Turner *et al.* 2000). In their first example, they found similar odds ratios and identical standard errors for the intervention effects were yielded from the summary and individual patient data methods.

The slight difference in the estimates for the intervention effect may be related to the summary data method inadequately estimating the parameters due to a lack of events in some of the trials, which required adding an arbitrary value of 0.5 to the cells within the trials (Turner *et al.* 2000). This was exemplified in Turner and colleagues second example, where smaller differences between the intervention estimates and its standard errors for the two methodologies were seen, since the included trials within this meta-analysis had larger sample sizes and adequate numbers of events within each intervention group (Turner *et al.* 2000).

2.6.1.2 Continuous outcome measures

A linear model may be used to analyse data from an individual patient data meta-analysis based on continuous outcome data (Whitehead 2002). Within this model, the outcome relating to patient j from study

i, where $i = (1,....,k)$; may be denoted by y_{ij}. The outcome y_{ij} is from a random variable Y_{ij} with a Normal distribution (Equation 2.27).

$$Y_{ij} \sim N(\mu_{ij}, \sigma^2) \tag{2.27}$$

A linear model may be described where α is the intercept, β_{0i} is study effect and is constrained so that β_{0k} is equal to zero, z_{ij} is a dummy variable relating to the intervention assignment, and β_1 is the intervention effect for the intervention group as compared to the control group (Equation 2.28). A common variance for the error terms is assumed across the studies (Whitehead 2002).

$$y_{ij} = \alpha + \beta_{0i} + \beta_1 z_{ij} + \varepsilon_{ij} \tag{2.28}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

A linear regression analysis is then performed to estimate the parameter for the intervention effect.

Whitehead has compared the estimates and standard errors obtained from modelling IPD and combining summaries (Whitehead 2002). The methods were found to give identical results for the intervention estimate. However, the standard error for the intervention estimate varied between the two methods where a larger estimate was seen from the individual patient data method as compared to the summary data method. The difference in the estimates is attributed to how the variance for each of the trials was calculated. In the summary data model the variances are allowed to vary across the trials, however the IPD model used in Equation (2.28) assumes a common variance

across all of studies. The difference in how the models are specified impact of the results since the estimate for the standard error of the intervention effect depends upon the estimates for the common variances (Whitehead 2002).

2.6.2 Random effect methods

If there are concerns using a fixed effect model where heterogeneity may be present then a random effect model could be considered. Recently, articles have been published which detail the implementation of multilevel models in individual patient data meta-analyses using dichotomous, continuous or ordinal outcome measures (Turner *et al.* 2000; Higgins *et al.* 2001; Whitehead *et al.* 2001; Whitehead 2002) although traditional mixed models can be used. Unlike the fixed effect model that can be performed using basic software, a random effect model requires the use of more expert software. The analyses of classical and Bayesian IPD models used in the examples in this thesis (Chapters 5, 6 and 7) have been performed using SAS for Windows (SAS Institute Inc) and WinBUGS (Spiegelhalter *et al.* 2003), respectively.

2.6.2.1 Random treatment by trial effects model

For comparability with the models described in section 2.4.2, fixed effects may be specified for the intervention and study effects and a random effect for interaction between the intervention effect and study effect, thereby allowing the intervention to vary randomly across the studies.

2.6.2.1.1 *Dichotomous outcome variables*

The model used for analysing dichotomous outcome measures is an extension to the fixed effect model (Equation 2.25 and 2.26) where a random effect term, u_{1i}, is included in the model for the interaction between intervention and study (Equation 2.29).

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha + \beta_{0i} + \beta_1 z_{ij} + u_{1i} z_{ij} \qquad (2.29)$$

$$u_{1i} \sim N\left(0, \tau^2\right)$$

Turner and colleagues have investigated the differences between using summary and IPD methods to analyse the log odds ratio in a random effects model (Turner *et al.* 2000). They found that using individual patient data method yielded larger estimates for the heterogeneity, and hence impacted on the estimates for the intervention effect and its standard error. These difference maybe also due to the summary data method performing poorly for the particular data set, possibly due to the corrections made for trials with no events in the intervention groups (Turner *et al.* 2000).

2.6.2.1.2 *Continuous outcome variables*

The model used for analysing continuous outcome measures is an extension to the fixed effect model described (Equation 2.28) where a random effect term, u_{1i}, is included in the model for the interaction between intervention and study (Equation 2.30). A common variance is assumed within each study so that the variances are not allowed to

vary across the study, and the ε_{ij} and u_{1i} are assumed to be uncorrelated.

$$y_{ij} = \alpha + \beta_{0i} + \beta_1 z_{ij} + u_{1i} z_{ij} + \varepsilon_{ij} \qquad (2.30)$$

$$u_{1i} \sim N(0, \tau^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

τ^2 needs to be estimated from the random effect models; ML and REML techniques have been proposed. The former has been found to yield estimates which are downwardly biased; therefore the restricted maximum likelihood method is preferred (Whitehead 2002) and used within the examples in this thesis. The approaches used in estimating the heterogeneity are similar to those presented in Section 2.4.2 however; the full likelihood for the individual patient data is used instead of calculating the likelihood based on summary data.

For the continuous outcome models, the Kenward and Roger approximation may be used to artificially inflate the standard errors of the variance components to allow for extra variation since this model assumes that the heterogeneity and variance for the error terms are known and not estimated from the data.

2.7 More than two treatment groups

This chapter has described the classical methodologies for combining data from trials with two treatment groups. However, trials may be conducted where there are multiple treatment groups, such as a new

treatment may be compared to not only a placebo, but also a standard treatment.

Conventionally when summary data methodologies have been used, meta-analyses which include more than two treatment groups have either been collapsed into two treatment arms, or separate meta-analyses have been performed for each pair wise comparison. However, the former approach is questionable since it may be clinically inappropriate to combine, for example, a standard treatment with a placebo; and the latter approach also raises questions about independence and multiple testing since the active treatment will be used in the analyses more than once. It would be more efficient to use a method which simultaneously estimates the pair wise comparisons using either a factorial design or contrast statements; individual patient data methods allow for this.

An advantage of using the traditional mixed effects model as compared to the multilevel models is that they provide a useful framework to analyse data were there are multiple treatment groups (Whitehead 2002).

Using the IPD models to analyse trials with more than two groups in a meta-analysis is advantageous over using the simple summary data methods, since they are not subject to the problems associated with independence and multiple testing (Whitehead 2002). Additionally, only a selection of the studies need contain more than two intervention groups.

2.8 *Bayesian approaches to meta-analysis*

A Bayesian framework may be applied to meta-analysis. Unlike the classical analysis, the Bayesian approach allows for statements and predictions to be made based on external knowledge concerning the efficacy of the intervention of interest. This information is introduced into the model in the form of a prior distribution. Bayes theorem is used to update the prior by combining it mathematically with the data from the individual studies, known as the likelihood, to form a posterior distribution. The posterior distribution may be summarised by its mean, which is the pooled estimate for the intervention effect; its standard error, and 95% credibility intervals.

2.8.1 Prior distributions

Prior external knowledge is an important aspect of the Bayesian framework. The information for the prior distribution may be sought from many sources. Data may be generated from previous relevant reviews, observational studies, or from expert opinion. Data from these sources can yield a variety of prior distributions and since the choice may considerably affect the results generated from the meta-analysis under investigation; caution needs to be taken when deciding on the prior distribution; however, there is no such thing as a 'correct' prior and consequently this has led to the methodology being heavily criticised. Therefore, the choice of subjective prior should be investigated in a sensitivity analysis. In the case where the reliability of the prior distribution is poor, or if prior distribution dominates the likelihood, then an inaccurate posterior distribution may be generated.

Alternatively, if the prior information is too specific, the prior may dominate the likelihood function from the individual trials data and overly influence the results of the meta-analysis.

Where there does not appear to be a consensus on the prior distribution, it may be advantageous to use what is known as a 'non-informative' or 'reference' prior distribution (Spiegelhalter *et al.* 2003). These prior distributions are usually uniform distributions over a wide range of values. This allows for the posterior distribution to have the same form as the likelihood function thereby allowing the likelihood function to dominate the prior distribution and hence the posterior distribution (Spiegelhalter *et al.* 2003). However, these 'non-informative' prior distributions need to be used with caution since they imply that all of the values within their bounds are equally as likely as each other (Fisher 1996). Therefore, whichever prior distribution is used, a thorough sensitivity analysis of the conclusions to the choice of prior distribution should be employed (Sutton *et al.* 1998).

2.9 Bayesian approaches to meta-analysis using summary data

In a fixed effect analysis, the Bayesian framework will yield estimates for the overall intervention effect and its associated standard errors, which are comparable to the results yielded from a classical analysis when non-informative prior distributions are specified in the model (Carlin 1992). Therefore, the Bayesian analyses performed in this thesis were only based on random effect models with non-informative prior distributions. Two methods were considered based on Bayesian

methodology to estimate the between study heterogeneity; these were Empirical Bayes (EB) and a full Bayesian (FB) framework.

2.9.1 Empirical Bayesian methodology

The EB methodology can be thought of as part Bayesian part classical; it was originally used to perform Bayesian analyses before there was software available which was capable of performing computationally intensive iteration and integrations that are required for a full Bayesian framework. However, empirical Bayes does not make use of subjective *a priori* beliefs to derive numerical values for the prior distributions (Sutton *et al.* 1998); instead it uses the data from the individual studies to generate prior beliefs about the overall intervention effect estimate and the between study heterogeneity estimate (Carlin 1992). However, the same limitations apply to this method as to the previous classical summary data methodologies for ML and REML since this method does not take into account that the heterogeneity estimate is calculated from the data from the individual studies.

A formula for estimating the heterogeneity based on an EB approach has been developed (Morris 1983; Berkey *et al.* 1995). The method is based on an iterative scheme similar to the REML method used in the classical approaches. The cycle begins where an initial value for θ is estimated using a weighted least squares regression model. This value is used to calculate an initial estimate for the heterogeneity (Equation 2.31).

$$\hat{\tau}^2 = \frac{\sum_{i=1}^{k} w_i \left(\frac{k}{k-1} \right) \left[\left(\hat{\theta}_i - \hat{\theta} \right)^2 - \mathrm{Var}\left(\hat{\theta}_i \right) \right]}{\sum_{i=1}^{k} w_i} \qquad (2.31)$$

The weights for a weighted least squares regression are then re-calculated (Equation 2.23) and a second estimate of $\hat{\theta}$ is yielded. The process is continued until convergence of the estimate for heterogeneity is achieved.

2.9.2 Full Bayesian framework

In a full Bayesian framework, the posterior distributions for the intervention effect and the between study variability are estimated using a series of complicated integrations (Smith *et al.* 1993). Since the integrations can not be computed in closed form due to their relative complexities and lack of exact analytical solution, statistical methodology has been developed to perform the integrations using iterations based on simulations. The most common method used is a particular form of the Markov chain Monte Carlo (MCMC) method called Gibbs sampling (Geman *et al.* 1984). Gibbs sampling generates samples from the conditional posterior densities, which should converge to the desired marginal posterior densities (Smith *et al.* 1993).

The Gibbs Sampling method may be performed in a freely available package called WinBUGS (Spiegelhalter *et al.* 2003). Within WinBUGS, the model and prior distributions are specified together with the data and initial starting values for the simulations; sampling distributions are formed from using the model and data, which are

then used to perform Gibbs sampling. An attractive feature of the method is that it can handle missing values in the data since the model consists of a joint distribution over all the observed and missing data (Spiegelhalter *et al.* 2003). The data is conditioned on to obtain a posterior distribution for the unknown parameters of interest; and marginalising over this posterior distribution is carried out using Gibbs sampling where algorithms are used to simulate values for the parameter so that inferences about them can be made (Spiegelhalter *et al.* 2003).

2.9.2.1 Continuous outcome measures

Assuming the intervention effect from each study is represented as $\hat{\theta}_i$, conventionally $\hat{\theta}_i$ follows a Normal distribution with mean θ_i and variance $\xi_i^{2^*}$ where $i = 1,....,k$ (Sutton *et al.* 1998) Equation (2.32).

$$\hat{\theta}_i \sim N\left(\theta_i, \xi_i^{2^*}\right) \tag{2.32}$$

$$\xi_i^{2^*} = \sigma_i^2 \left(\frac{n_{t_i} n_{c_i}}{n_{t_i} + n_{c_i}} \right)^{-1}$$

Where the prior distribution are specified for parameters as follows:

$$\theta_i \sim N\left(\theta, \tau^2\right)$$

And vague prior distributions are used for:

$$\theta \sim N\left(0, 10^6\right)$$

$$\tau^2 \sim IG(0.001, 0.001)$$

2.9.2.2 Alternative methods for dichotomous outcome measures

The Bayesian framework allows for the dichotomous data from a 2x2 contingency table to be modelled directly (Table 2.1 in Section 2.3.1.1). The observed number of events in each intervention group of the study is assumed to follow a binomial distribution, where $r_{.i}$ is the number of observed events and $n_{.i}$ is the total number of patients, and the unknown risk parameters are $\pi_{.i}$, within each intervention group (Equation 2.33), where subscript t and subscript c refer to the active intervention and the control groups, respectively, for study $i = 1,.....,k$.

$$r_{t_i} \sim \text{Bin}(\pi_{t_i}, n_{t_i})$$
$$r_{c_i} \sim \text{Bin}(\pi_{c_i}, n_{c_i}) \qquad (2.33)$$

The odds for each group may be calculated (Smith *et al.* 1995) using Equation (2.34) where θ_i is the log odds ratio for trial i.

$$\log\left(\frac{\pi_{t_i}}{1-\pi_{t_i}}\right) = \lambda_i + \theta_i \qquad (2.34)$$

$$\log\left(\frac{\pi_{c_i}}{1-\pi_{c_i}}\right) = \lambda_i$$

The log odds ratio, θ_i, is the difference between the logarithms of the odds for each group as shown in Equation (2.35).

$$\log\left(\frac{\pi_{t_i}}{1-\pi_{t_i}}\right) - \log\left(\frac{\pi_{c_i}}{1-\pi_{c_i}}\right) = \theta_i \qquad (2.35)$$

Heterogeneity is incorporated into the model so that the log odds ratio from each study is allowed to vary around the overall log odds ratio for the intervention effect (Equation 2.36).

$$\theta_i \sim N(\theta, \tau^2) \tag{2.36}$$

Where vague prior distributions are specified for the parameters as follows:

$$\theta \sim N(0, 10^4)$$

$$\tau^2 \sim IG(0.001, 0.001)$$

The distribution is conventionally specified as a Normal distribution and the estimation procedure follows the same format as used for continuous outcome measures in Equation (2.32).

It has been suggested that there is substantial gain from modelling the data directly using a binomial distribution since it avoids the need for simplifying approximations and the assumption of linearity of the log odds ratio (Warn *et al.* 2002). Also, this method uses binomial distributions for the underlying distributions of the data and so is not subject to the requirement of a continuity correction being applied in the case of any sparse data cells.

2.10 Bayesian approaches to meta-analysis using individual patient data

A Bayesian framework may be used to model the data from a meta-analysis based on individual patient data, however additional prior distributions need to be specified for the parameters. Dichotomous and continuous outcome measures are described below for random

effect models only, since the Bayesian framework would yield estimates from a fixed effect model which are comparable to those yielded from a classical fixed effect IPD model. The models which may be used for dichotomous and continuous outcome variables are described below.

2.10.1 Dichotomous outcome measures

Assuming that the outcome for a patient is denoted as y_{ij} which is from a random variable Y_{ij} (Equation 2.25); and assuming Y_{ij} has a binomial distribution with parameter π_{ij} and a denominator of 1. Then π_{ij} is the probability of an event for patient j in study i where $i = 1,.....,k$; and y_{ij} =1 if the event occur and 0 if they are event-free. Comparing this model to Equation (2.29), $\alpha = 0$, so that β_{0i} is the intervention effect in the control group for study i, and β_1 is the pooled intervention effect of an event on intervention as compared to control and u_{1i} is the intervention effect in study i (Equation 2.37).

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0i} + u_{1i} z_{ij} \qquad (2.37)$$

$$u_{1i} \sim N\left(\beta_1, \tau^2\right)$$

Where vague prior distributions are specified for parameters as follows:

$$\beta_{0i} \sim N\left(0, 10^4\right)$$

$$\beta_1 \sim N(0,10^4)$$

$$\tau^2 \sim IG(0.001,0.001)$$

Whitehead has compared the results from Bayesian random effect models of IPD and summary data (Whitehead 2002). Similar estimates for the log odds ratio of the intervention effect were found between the two methods. The standard error for the intervention effect was larger from the IPD model as compared to the summary data model which could be attributed to a larger estimate for the between trial heterogeneity being estimated from the IPD model.

Whitehead has also compared the results between Bayesian and classical individual patient data random effect models, and has found that similar estimates for the treatment effect were seen between the models. Since the models specified for the Bayesian and classical IPD models are identical then the parameters estimated from the models should also be identical. However, difference may arise due to not being able to specify a truly non-informative prior for each parameter in the Bayesian model. The estimation process used in the Bayesian models are based on using simulations (Monte Carlo methods) and so may be subject to error in its final estimation of the parameters. Additionally, unlike the classical models which assume that the between trial heterogeneity and individual study variances are known, the Bayesian model allows for extra variation in its estimates to account for the uncertainty associated with estimating these components from the data.

2.10.2 Continuous outcome measures

The random effect IPD model used for analysing continuous outcome measures in a Bayesian framework is analogous to that from a classical setting (see Equation 2.30); however, $\alpha = 0$ so that β_{0i} is the intervention effect in the control group for study i. A common variance, σ^2, is assumed for the error terms so that it is the same across trials. β_1 is the pooled intervention effect and u_{1i} is the intervention effect in study i.

$$y_{ij} \sim N(\lambda_{ij}, \sigma^2) \qquad (2.38)$$

$$\lambda_{ij} = \beta_{0i} + u_{1i} z_{ij}$$

$$u_{1i} \sim N(\beta_1, \tau^2)$$

Where vague prior distributions are specified for parameters as follows:

$$\beta_{0i} \sim N(0, 10^4)$$

$$\beta_1 \sim N(0, 10^4)$$

$$\sigma^2 \sim IG(0.001, 0.001)$$

$$\tau^2 \sim IG(0.001, 0.001)$$

Whitehead has compared the results from Bayesian random effect models based on summary and IPD (Whitehead 2002). Similar estimates for the intervention effect were seen between the models since both of the data sets are assumed to follow a normal

distribution; however, a slightly larger standard error was seen from the IPD model which could be attributed to a larger estimate for the heterogeneity due to a common within trial variance, σ^2, being estimated.

Higgins and colleagues have compared the results from IPD meta-analyses based on Bayesian and classical frameworks (Higgins *et al.* 2001). They found that estimate for the intervention effect from the Bayesian analysis was noticeably smaller and its associated standard error was larger in magnitude than compared to the classical analysis, additionally the estimate for the heterogeneity was considerably smaller from the Bayesian analysis. The difference in the intervention effect was thought due to the Bayesian model being drawn towards the estimate for the largest trial and was specific to the example used and not generalisable (Higgins *et al.* 2001).

Although the standard errors for the intervention effect estimated from the classical IPD random effect model can be artificially inflated using the Kenward and Roger approximation, this method is an approximation where as in the Bayesian model the parameters are specified as random variables and hence the Bayesian method estimates more variation for the between study heterogeneity then its classical counterpart.

2.11 Methods for Assessing Convergence in Bayesian models

Iterative simulation methodology used in the analysis of Bayesian statistics is a valuable method for summarising posterior distributions;

however, there are disadvantages associated with the methodology, particularly related to the assessment of convergence of the chains. Many tools have been developed to assess whether convergence has been achieved, these include methodology developed by Geweke (Geweke 1992), Gelman and Rubin (Gelman *et al.* 1992), Raferty and Lewis (Raferty *et al.* 1992). These tools should be used in parallel with each other since no one method is superior for assessing convergence.

2.11.1 Gelman and Rubin diagnostic

Gelman and Rubin have developed a method for assessing convergence where multiple chains or sequences are run simultaneously with different starting values for the parameter of interest (Gelman *et al.* 1992). The simulations are run for a particular length of chain where the first q iterations represent the burn-in period and hence discarded, the remaining n iterations are focussed on for evaluation of convergence for each of the m chains. For each parameter of interest a between-sequence and within-sequence variance are calculated. The between-sequence variance is the variance between the m sequence means for the parameter of interest, and the average of the m within-sequence variances for each chain is calculated and represented by W. A scale factor, $\hat{R}$, is then estimated from the ratio of the current variance estimate and the average of the within-sequence variances. The scale factor is the scale by which the current distribution for the parameter of interest might be reduced by if the simulations sequences were allowed to run

to infinity. When the value of $\hat{R}$ approaches one then the chain is thought to have converged for that particular parameter of interest. Potential problems with this method, which are mainly related to conditions where the parameter of interest has a multimodal distribution (Gelman *et al.* 1992). This is thought to be a problem since the tool will assess convergence of the whole chain and if the modes for the distribution are wide apart then the simulations will tend towards one of the modes and the tool will assess convergence at this one mode only. Even though this method can not correct for multimodal distributions, it should draw attention to this problem by repeatedly bouncing from one mode to another.

2.11.2 Raferty and Lewis diagnostic

Raftery and Lewis have proposed a method for assessing convergence of a single chain (Raferty *et al.* 1992). This method first assesses the length of burn-in required for the chain of iterations, values are specified by the user for the percentile that is to be assessed, degree of accuracy for the estimate in this percentile, and the required probability for attaining the degree of accuracy specified. The tool then calculates the total number of iterations that would need to be performed to reach convergence. The method reports an independence factor, *I*, which is a measurement of the dependence between the iterations in the single chain. *I* is the ratio of the total number of iterations that the model should be run for to achieve convergence, divided by the number of iterations calculated for the burn-in. Raferty and Lewis suggest that if *I* has a value greater than 5

then there are some high within-chains correlation and convergence has probably failed (Raferty *et al.* 1992).

2.11.3 Geweke diagnostic

A standard time series method called 'Geweke's diagnostic' may be used to assess if the mean of the variable of interest has converged (Geweke 1992). The diagnostic splits the chain generated into two segments, usually the first ten percent and the last fifty percent. If the chain has converged, then the figures in the two segments should be similar. A statistical calculation is performed that checks the similarity of the two figures, where the difference of the two figures is divided by the asymptotic standard error of their difference is the convergence diagnostic, *z*. It is assumed that if the chain has converged, then as the length of the chain tends to infinity, the sampling distribution should tend to a standard normal distribution. Hence, if any of the values for *z* fall in the extreme ends of the tails of a standard normal distribution then the chain has not converged.

There are problems associated with this method; the first being related to the segments used to check convergence. It has been recommended by Geweke that the first 10% and last 50% of segments should be used and compared; assuming that convergence has been achieved within the last 50% of the chain, then values from these two segments will only be similar if the chain has already converged within the first 10% of the iterations (Naylor 1992), which is unlikely. Additionally, the interpretation of the convergence diagnostic, *z*, is unusual. The null hypothesis set up is that there is no difference

between the mean of the estimates in the two segments. However, convergence is achieved when there is sufficient evidence to not reject the null hypothesis (Naylor 1992).

2.12 Random trial effects in meta-analysis using individual patient data

In the previous classical and Bayesian individual patient data models we have assumed that the trial effects are fixed. A random effect for the trial effect could be specified in addition to the random effect for the interaction between intervention and trial effects, or as an alternative to having fixed interaction term between intervention and trial effects. Incorporating studies as random effects has been controversial.

In a meta-analysis which consists of trials with various numbers of centres, it may be advantageous to model the trial effects as random. It has been argued that missing data relating to intervention differences may be recovered by having the trials as random effects when the sizes of the intervention groups within the trials vary (Brown *et al.* 1994).

Conversely, it has suggested that it may be inappropriate to model study effects as random since this would imply that the results yielded from a particular set of studies are drawn at random from an underlying population of studies (Turner *et al.* 2000; Higgins *et al.* 2001).

2.13 Assessing publication bias

The results yielded from a meta-analysis are reflective of the data included in the analysis; therefore it is essential that the data are at least representative of the whole population of studies that have been conducted in the particular medical area. Publication bias may still be present in a meta-analysis even if a thorough search of the literature is performed; therefore it is necessary to visually and statistically inspect the data to assess if publication bias is present.

Techniques have been devised to assess this, including the funnel plot (Light *et al.* 1984; Begg *et al.* 1994), Begg and Mazumdar's rank correlation test (Begg *et al.* 1994), an asymmetry test by Egger and colleagues (Egger *et al.* 1997) and more recently the 'Trim and Fill' method (Duval *et al.* 2000).

2.13.1 Begg's funnel plot

The funnel plot is the most commonly used procedure due to its relative simplicity to use and interpret. The funnel plot is created usually by plotting the standard error for the effect size against the effect size of the intervention for each trial (Figure 2.1). Trials that have smaller sample sizes will be located towards the bottom of the plot where the standard errors are greater. If asymmetries between the left and right sides of the funnel plot are seen, this may indicate that there is publication bias present. Figure 2.1 shows asymmetry where there is an absence of smaller sized trials towards the left-

hand side of the plot indicating that smaller size trials which show a detrimental intervention effect could have been missed.

Pseudo confidence intervals can be generated for the estimate and plotted on the figure; these interval lines may aid the investigator to determine whether appears to be evidence of asymmetry in the plot (Figure 2.1).

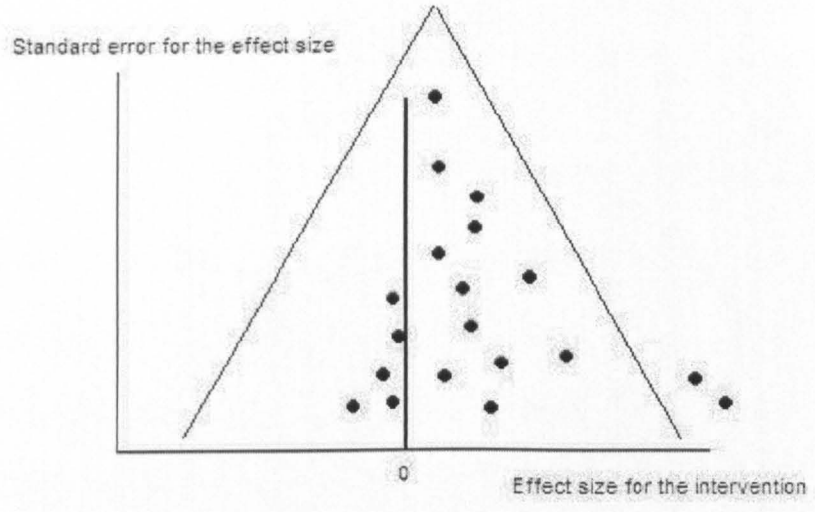

Figure 2.1 Example of Begg's Funnel Plot showing publication bias using hypothetical data

2.13.2 Begg and Mazumdar's rank correlation test

This test has been developed to statistically test for the presence of publication bias and is a direct statistical analogue to the funnel plot. The test examines whether there is a correlation between the effect estimates and their variances, since it has been found that publication bias tends to produce such an effect (Begg *et al.* 1994); and produces a p value. However, the test has variable power being dependent on the number of studies included in the meta-analysis. The test was

found to be powerful for meta-analyses containing 75 studies; but had a marked reduction in power when only 25 studies were considered. Since it is relatively common for meta-analyses to have much less than 25 studies, this test may be inadequate at finding evidence of publication bias in many meta-analyses, therefore caution is needed in interpreting the results in these circumstances.

2.13.3 Egger's asymmetry test

This test is statistically based on estimating a regression line for relationship between the effect size (odds ratio or standardised difference) and the precision of the study (usually defined as the reciprocal of the study variance). The results are often presented visually as shown in Figure 2.2. Evidence of publication bias is detected if the intercept of the regression line deviates significantly from zero. In the example shown there does not appear to be evidence of publication bias since the intercept of the regression line is close to zero.

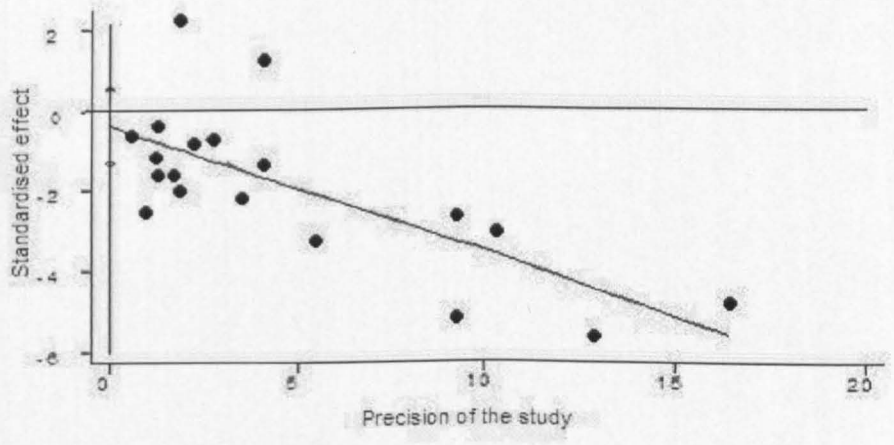

Figure 2.2 Example of Egger's Asymmetry Plot using hypothetical data

2.13.4 "Trim and Fill" methods

A relatively new statistical procedure called the "Trim and Fill" method is a statistical and visual procedure to assess publication bias. The method was developed by Duval and Tweedie and is based on a non-parametric ranking technique (Duval *et al.* 2000). This method essentially assesses the effect of adjusting for potential publication bias by including the missing data in the meta-analysis.

Initially, a funnel plot is used to visually detect if asymmetry is present. If asymmetry is seen, an estimate of how many studies would be required to make the points on the plot symmetrical is performed. Next, the data from the asymmetrical part of the plot is 'trimmed' and the underlying effect estimate for the remaining trials is calculated. Then the trimmed studies and their missing counterparts are included in a further analysis to calculate a new 'filled' estimate for the overall intervention and confidence intervals. The adjusted result may be presented in the form of a sensitivity analysis rather than as a 'better' estimate since the adjusted result could be misleading if the asymmetry in the plot is due to factors other than publication bias (Song *et al.* 2000).

2.13.5 Summary to publication bias

Publication bias may be assessed using various visual and statistical methods. However, when performing these statistical tests it is wise to remember that these tests generally have low power in detecting publication bias since the power of the test depends on the number of studies included in the meta-analysis.

The test by Duval and Tweedie (Duval *et al.* 2000) is more sophisticated than the two previous methods since it allows for the investigator to not only assess publication bias in the studies identified but to also allow for adjustment in the treatment estimates to see the effect of including hypothetical data (Sutton *et al.* 2000). However, it must be noted that asymmetry may be detected from the funnel plot which it is unrelated to publication bias, since the asymmetry may be related to poor methodological quality of the studies included in the meta-analysis (Petticrew *et al.* 1999).

2.14 Summary of chapter

This chapter has highlighted the models that can be used in meta-analyses using summary and individual patient data methodologies, both in classical and Bayesian frameworks. The issues surrounding the whether fixed or random effects models should be used and the estimators for the between trial heterogeneity have been described and compared. Statistical methods for assessing publication bias have also been described. However, assessments of heterogeneity need to be made to aid with determining whether the fixed effect models are valid for use or whether it would be more appropriate to use a random effect model.

CHAPTER 3

HETEROGENEITY IN META-ANALYSES USING SUMMARY AND INDIVIDUAL PATIENT DATA METHODOLOGIES

3.1 Introduction

The previous chapter outlined two models, the fixed effect model and the random effect model. The fixed effect model assumed that the studies are identical in design and patient population, and ignored any differences between the studies with regards to the intervention efficacies. In contrast, the random effect model accounted for any variation (heterogeneity) between the studies through incorporating it into the model and estimating its magnitude. However, neither of these methods explored why the studies gave varying results. It has been suggested that in a meta-analysis, it is important to perform a full investigation into exploring why these differences exist (Thompson 1994).

This chapter concentrates on how the differences between the studies can be explored using graphical and more formal statistical tests, and then describes two statistical methods called subgroup analysis and meta-regression, which allow for covariates to be included in both a fixed effect model and a random effect model, using summary study and individual patient data. Within the random effect model, a variety of estimation methods are described which quantify any residual variation between the studies using classical and Bayesian methodologies.

3.2 Types of heterogeneity in meta-analyses

Even if a perfect replication of studies were performed to reduce variation of the internal factors, the intervention effect estimates from

the studies will vary due to random fluctuation. However, if these differences appear to vary considerably more than random fluctuation then it may be necessary to explore this variation.

Heterogeneity can be categorised into four major areas relating to variations and difference between patient, intervention, co-intervention, and outcome (Glasziou *et al.* 2002). These are factors which may produce a true variation in the effectiveness of the intervention. Patient level factors such as age or severity of disease could influence the effectiveness of the intervention. Also, differences in the intensity or dose of the intervention or the effect of co-interventions may alter the effectiveness of the intervention under study. Differences in the outcome such as when it is measured or indeed what outcome measure is chosen could have dramatic effects on the efficacy of the intervention.

However improper randomisation procedures and non-adherence with interventions may also be causes of heterogeneity (Glasziou *et al.* 2002). Since these may produce apparent differences in the intervention effect estimates which lead the investigators to assume that some other factor is causing the variation, whereas these differences may be due to the quality of the study. Indeed it has been shown that improper concealment of allocation for the intervention can produce inflated intervention effect estimates (Schulz *et al.* 1995).

3.3 Assessments of heterogeneity

Although it has been generally agreed that heterogeneity should be accounted for in the model, at present there is no consensus on how

it should be performed (Thompson 1994; Lau *et al.* 1998). Various methods exist for assessing and quantifying these differences between the study estimates. These range from simple graphical assessments to complicated formal statistical tests and estimation methods. The simplest method to use is the forest plot.

3.3.1 Graphical Assessment of Heterogeneity

In a forest plot the study estimates are plotted with their corresponding confidence interval limits (Figure 3.1). The size of the point estimate symbol on the graph relates to the precision of the estimate. If high precision is present, the standard error for the point estimate will be small, hence the symbol is large; and vice versa. From Figure 3.1 there appears to be some heterogeneity between the study estimates, since the results are varied and the overall estimate is not included in the 95% CI for study F.

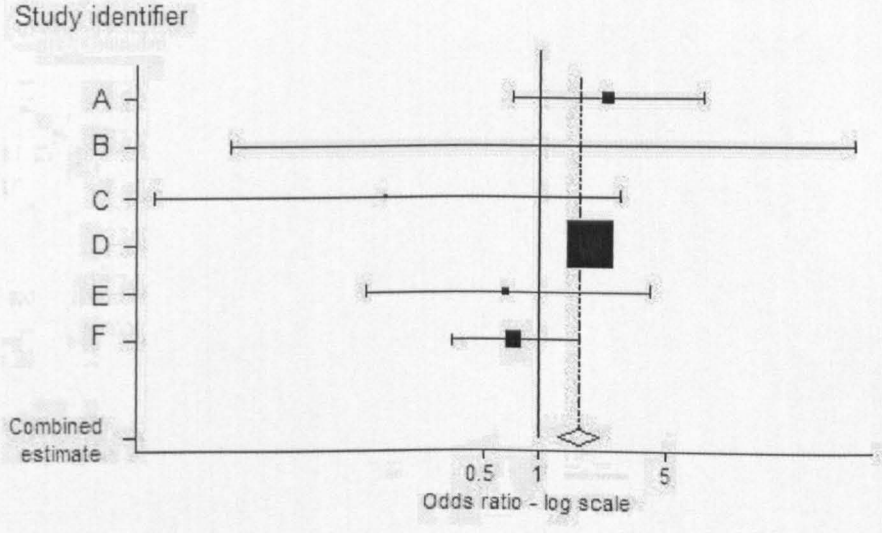

Figure 3.1 Example for assessing heterogeneity using a forest plot

3.3.2 Classical approaches for assessing heterogeneity for summary data

A variety of statistical tests can be performed to identify if there is evidence of statistical heterogeneity between the study point estimates, including Cochran's homogeneity test (Cochran 1954) and Higgins and Thompson I^2 statistic (Higgins *et al.* 2003).

3.3.2.1 Cochran's homogeneity test

Cochran first proposed a test for homogeneity in 1954 and this has been used conventionally to statistically test for between study heterogeneity (Cochran 1954). If we assume k studies are included in the meta-analysis, then under the null hypothesis, H_0, the underlying intervention effect, denoted by θ_i for each study, are the same, θ.

$$H_0 : \theta_1 = \theta_2 = ... = \theta_k = \theta \tag{3.1}$$

The alternative to the null hypothesis is that at least one of the intervention effects from the studies differs from the others. Under the null hypothesis the statistic Q follows a χ^2 distribution with its degrees of freedom dependent up on the number of studies included in the meta-analysis.

$$Q = \sum_{i=1}^{k} w_i \left(\hat{\theta}_i - \hat{\theta}\right)^2 \tag{3.2}$$

Where w_i is the weight associated with the i'th study, and $\hat{\theta}_i$ is estimate for the intervention effect from each study and $\hat{\theta}$ is the estimate for the pooled intervention effect. The conventional weight

given to each study is the reciprocal of the variance for each trial (Cochran 1937) (see Equation 2.11).

The main disadvantage associated with this test is related to the power the test has to be able to detect heterogeneity between the studies (Fleiss 1986; Whitehead *et al.* 1991; Thompson 1994). Insufficient power may be seen when a small number of studies are being combined in the meta-analysis. When small numbers of studies are being combined, Fleiss recommends using a 10% significance level to improve detection (Fleiss 1986). It should also be noted that there is the potential to detect heterogeneity if large sample sized studies are used in the meta-analysis, even if the intervention estimates for each study appear homogeneous (Hardy *et al.* 1998). Also, this test is based on the assumption that the variances for each study in the meta-analysis are known, when in fact they have been estimated from the data and therefore has been criticised as being of limited value (Hardy *et al.* 1998).

Therefore due to the disadvantages associated with this test it has been recommended that the Q statistic should not be used as the only tool for identifying heterogeneity (Hardy *et al.* 1998); but used in conjunction with other techniques such as the forest plot.

3.3.2.2 Higgins and Thompson, I^2

A recent method has been devised which quantifies heterogeneity as a proportion of the total variability in the model. This method appears not to be subject to the problems of Cochran's homogeneity test (Higgins *et al.* 2003). In the below equation (3.3), H is the estimate of

the heterogeneity from the χ^2 test statistic as devised by Cochran and k is the number of studies in the meta-analysis.

$$H = \sqrt{\frac{Q}{k-1}} \tag{3.3}$$

Higgins and Thompson propose that a measure called I^2, which is the 'proportion of total variability explained by heterogeneity' to determine if heterogeneity is present (Equation 3.4).

$$I^2 = \min\left(\frac{H^2 - 1}{H^2}, 0\right) \tag{3.4}$$

In the case of negative values of I^2 being calculated, values of I^2 are set to zero so that the value will lie between 0% and 100%. Values of 25%, 50% and 75% have been suggested as categorisations for low, moderate, and high measures of heterogeneity, respectively (Higgins *et al.* 2003). This method has the added advantage that the values generated may be compared across meta-analyses. However, since the nature of systematic reviews is to bring together studies which are naturally diverse in their design protocols and patient populations, many meta-analyses will find quantifiable levels of heterogeneity. It is also important to identify the reasons behind the existence of heterogeneity in addition to quantifying it. Even though this method is simple to calculate and does not appear to be subject to the disadvantages of Cochran's method, it is still dependent on the assumption that the within study variances are known for each study and are equal between the intervention and control groups within each study; additionally, the test does not produce a pooled point estimate adjusted for the heterogeneity.

3.4 Adjusting for covariates using summary data

Conventional meta-analyses have tended to focus their methodology on presenting results based solely on the averaged outcomes of the available studies. When there is a large quantity of unexplained variation between the study estimates a full exploration of heterogeneity needs to be performed (Thompson 1994).

Covariates could be included in the model which were thought to either explain the variability between estimates or influence the efficacy of the intervention (Rubin 1990). These could be based on either patient characteristics such as age, gender or study characteristics such as sample size, and the quality of randomisation procedure (Berkey *et al.* 1995). Subgroup analysis and meta-regression are two such methods that allows for the covariates to be assessed in a model.

3.4.1 Subgroup analysis using summary data

If the covariate of interest may be categorised, then subgroup analysis may be performed to assess its effect on the overall effect estimate. Subgroup analyses investigate the patients' characteristic by considering a subset of studies from the pooled studies, therefore one needs to be cautious of the problems associated with misclassification (Gelber *et al.* 1987).

Similar methodologies are used for each subset of studies for fixed and random effect models as described previously. Forest plots are commonly used for each subset to assess visually whether there are any differences between the estimates for subset. Additionally, Deeks

and colleagues have described how Cochran's homogeneity statistic, Q may be used to assess whether using the subgroups has explained any residual heterogeneity (Deeks *et al.* 2001). This method involves calculating the Q statistic for all of the studies (Q_{all}), and subtracting the sum of the Q values from the m subset of studies (Q_m). For example, if two subsets were considered then the new value of Q called Q_{new}, would be as shown in equation (3.5).

$$Q_{new} = Q_{all} - (Q_1 + Q_2) \tag{3.5}$$

Q_{new} is then compared to a χ^2 distribution with $m-1$ degrees of freedom to test for a difference amongst the subgroups considered.

3.4.2 Meta-regression analysis using summary data

Meta-regression is a more flexible method which can be used to test for differences, continuous and categorical covariates may be included in the model. Meta-regression is based on using simple regression models to assess the relationship between the outcome and the explanatory variables; however, weightings are used in the model to allow for the size of the trial to be taken into account.

In a meta-analysis of summary data, the number of included trials represents how many observations are modelled in the regression, therefore it has been recommended that meta-regressions are not performed with less than 10 studies in the meta-analysis (The Cochrane Collaboration 2003). Also, the number of covariates that are included in the model needs to be decided with caution since if

only a relatively small number of studies are included in the meta-analysis, only a limited number of covariates may be incorporated into the regression analysis to prevent the residual degrees of freedom being exhausted.

Two types of meta-regression exist, the fixed effect extension and the random effect extension. The covariates are introduced as fixed effects in both cases. It is thought to be appropriate to use the fixed effect extension when the additional variation between the intervention estimates can be explained by the covariates introduced in the model (Whitehead *et al.* 1991) however, in practice they may be rarely achieved and so it has been recommended that random effect models are used to allow for the covariates only explaining part of the heterogeneity (Sutton *et al.* 1998).

This remainder of this chapter will concentrate on incorporating only one additional predictor in the model however the models can be extended to allow for multiple covariates.

3.4.2.1 Random effects meta-regression using Classical and Bayesian methodologies

The model described in Equation (2.14) may be extended so that a trial level covariate, x_{1i} is now included in the model where $i = 1,....,k$ (Equation 3.6), where u_i is the random effect term, and the terms u_i and ε_i are assumed to be independently distributed. θ is now replaced by β_0 and represents the pooled intervention effect when x_{1i} = 0. The covariates are assumed fixed and known without error.

$$\hat{\theta}_i = \beta_0 + \beta_1 x_{1i} + u_i + \varepsilon_i \tag{3.6}$$

$$u_i \sim N(0, \tau^2) \text{ and } \varepsilon_i \sim N(0, \xi_i^2)$$

Estimation of the parameters is achieved using a weighted least squares regression model, where the weights, w_i, used are given by the inverse of the study's variance making the assumption that:

$$w_i = \frac{1}{Var(\hat{\theta}_i)} = \frac{1}{\xi_i^2 + \tau^2} \tag{3.7}$$

In addition to the individual study variance being estimated from the data (see section 2.3); heterogeneity and the unknown regression parameters need to be estimated too, but to estimate the unknown regression parameters first need to be estimated; and vice versa. Similar methodologies to as described in sections 2.4.2 have been extended to allow for covariates to be included in the model so that the above parameters may be estimated. These include the weighted and approximate MM, ML, and REML.

3.4.2.1.1 Approximate and weighted method of moments

An approximate MM estimator can be used to estimate τ^2 from a meta-regression model which incorporates one covariate (Raudenbush 1994). An ordinary least squares meta-regression is performed to estimate the residual sum of squares (RSS), then an initial value is sought for $\hat{\tau}^2$ using Equation (3.8).

$$\hat{\tau}^2 = \left[\frac{\mathrm{RSS}}{k-2}\right] - \frac{\sum_{i=1}^{k} \mathrm{Var}\left(\hat{\theta}_i\right)}{k} \qquad (3.8)$$

A weighted least squares regression model is then used to estimate the intervention effect and the covariate effect using optimal weights (Equation 3.7). This method is relatively simple to calculate but requires the use of an approximation to replace the sampling variance, since this method is an extension of a method that assumes the sampling variances from each study are the same across all of the studies (Raudenbush 1994; Sutton *et al.* 1998).

As an alternative to the above method, a more complex method may be used which estimates the heterogeneity using MM (Thompson *et al.* 1999). Firstly, a weighted least squares regression is performed to obtain the heterogeneity statistic, Q, and initial estimates for the intervention and covariate effects. Next, the method of moments estimator is used to estimate the between study heterogeneity (Equation 3.9) where $F(w,x)$ is a function of the weight and covariate for each study (Thompson *et al.* 1999).

$$\hat{\tau}^2 = \frac{Q-(k-2)}{F(w,x)} \text{ if } Q > k-2 \text{ or 0 otherwise} \qquad (3.9)$$

Finally, a weighted least squares regression is performed using the optimal weights to obtain final estimates for the intervention and the covariate effect. This method is more complex than both the approximate method and weighted method of moments without a

covariate, since it requires the use of matrices to estimate the between study heterogeneity (Thompson *et al.* 1999).

A similar method has been proposed by Raudenbush however this method assumes that the variances within each study are balanced (Raudenbush 1994), so that they are same across the studies. This assumption cannot be met in most cases (Sutton *et al.* 1998) and so is not described in details in this thesis.

3.4.2.1.2 Maximum likelihood

The ML methodology used in Section 2.4.2.2 may be used where a fixed effect weighted least squares meta-regression is performed to obtain the initial values for $\hat{\beta}_0$ and $\hat{\beta}_1$; where $\hat{\beta}_{0i}$ is the estimate for the intervention effect in study i. The iterative model used in section 2.4.2.2 may be extended to include a covariate (Equation 3.10) (Hardy *et al.* 1996).

$$\hat{\tau}^2 = \frac{\sum_{i=1}^{k} w_i^2 \left[\left(\hat{\beta}_{0i} - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} \right)^2 - \mathrm{Var}\left(\hat{\theta}_i \right) \right]}{\sum_{i=1}^{k} w_i^2} \qquad (3.10)$$

The cycle is repeated using new weights as described in Equation (3.10) until convergence of $\hat{\tau}^2$ is achieved.

3.4.2.1.3 Restricted maximum likelihood

The REML iterative cycle described in section (2.4.2.3) may be used to obtain initial estimates of the unknown regression coefficients (Thompson *et al.* 1999). Then an initial estimate of the between study heterogeneity may be obtained using Equation (3.11).

$$\hat{\tau}^2 = \frac{\sum_{i=1}^{k} w_i^2 \left(\frac{k}{k-2}\right)\left[\left(\hat{\beta}_{0i} - \hat{\beta}_0 - \hat{\beta}_1 x_{1i}\right)^2 - Var\left(\hat{\theta}_i\right)\right]}{\sum_{i=1}^{k} w_i^2} \quad (3.11)$$

This estimate may then be used to calculate optimal weights. The cycle is repeated until convergence of $\hat{\tau}^2$ is achieved.

3.4.2.3.1 Empirical Bayes

EB methodology may be used to obtain estimates for the between study heterogeneity. Initial estimates of the intervention effect and the covariate may be yielded and an initial estimate of the heterogeneity may be calculated using Equation (3.12) (Berkey *et al.* 1995).

$$\hat{\tau}^2 = \frac{\sum_{i=1}^{k} w_i \left(\frac{k}{k-2}\right)\left[\left(\hat{\beta}_{0i} - \hat{\beta}_0 - \hat{\beta}_1 x_{1i}\right)^2 - Var\left(\hat{\theta}_i\right)\right]}{\sum_{i=1}^{k} w_i} \quad (3.12)$$

New weights are then estimated using optimal weights and the cycle is repeated until convergence of $\hat{\tau}^2$ is achieved.

3.4.2.1.5 Full Bayesian framework

A FB framework can be used which is similar to the methods described in Section (2.9.2) for continuous outcome data and for binary outcome data. Both of these methods require the use of priors for the unknown parameters; which may be non-informative (Smith *et al.* 1995), and an assessment of convergence for all of the estimate parameters (see Section 2.11).

For continuous summary outcome data, the methodology used to perform a FB meta-regression analysis is an extension to the method

used in Equation (2.32) where the pooled intervention effect θ is replaced by β_0 and represents the pooled effect when the covariate, x_{i1} = 0. The prior distribution for β_{0i} is shown as in Equation (3.13).

$$\beta_{0i} \sim N(\mu_i, \tau^2)$$
$$\mu_i = \beta_0 + \beta_1 x_{1i} \qquad (3.13)$$

For dichotomous outcome data, the Bayesian framework allows for the data used in a *2x2* contingency table to be modelled directly (see Equation 2.32); x_{1i} may be included in the model as shown in Equation (3.14) (Smith *et al.* 1995).

$$\log\left(\frac{\pi_{t_i}}{1-\pi_{t_i}}\right) = \lambda_i + \beta_{0i} + \beta_1 x_{1i} \qquad (3.14)$$
$$\log\left(\frac{\pi_{c_i}}{1-\pi_{c_i}}\right) = \lambda_i$$

β_0 and β_1 may be estimated as the difference between the logarithms of the odds for each group (Equation 3.15).

$$\log\left(\frac{\pi_{t_i}}{1-\pi_{t_i}}\right) - \log\left(\frac{\pi_{c_i}}{1-\pi_{c_i}}\right) = \beta_{0i} + \beta_1 x_{1i} \qquad (3.15)$$

For the above models additional vague prior distributions are specified.

$$\beta_1 \sim N(0, 10^4)$$
$$\beta_0 \sim N(0, 10^4)$$
$$\tau^2 \sim IG(0.001, 0.001)$$

3.5 Heterogeneity in meta-analysis using individual patient data

A statistical test for heterogeneity between the k trial estimates, where $i = 1,....,k$ may be estimated by including a fixed effect interaction term between intervention effects and study effects (Whitehead 2002). The model used for the main effects is comparable to the models described in Section 2.6.2 where fixed effects are specified for the intervention and study effects, where β_{0k} is constrained to zero. The interaction term allows for the intervention effects to vary across the trials.

For dichotomous outcomes the model presented in Equation (2.26) may be extended to Equation (3.16). A test for the interaction may be performed by assessing the change in deviance between this model and the model presented in Equation (2.26) may be compared to a χ^2 distribution with $k-1$ degrees of freedom.

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha + \beta_{0i} + \beta_{1i} z_{ij} \tag{3.16}$$

For continuous outcomes the model presented in Equation (2.28) may be extended to Equation (3.17). The estimate for the interaction term may be compared to a F distribution with $k-1, n-2k$ degrees of freedom n is the total number of patients in the meta-analysis.

$$y_{ij} = \alpha + \beta_{0i} + \beta_{1i} z_{ij} + \varepsilon_{ij} \tag{3.17}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

However, there is usually a lack of power to test the interaction terms in the above two models and so in practice the variability associated with the intervention effect across the trials is often ignored (Brown *et al.* 1994).

If the final model chosen for the analysis includes the interaction term, then caution must be observed in interpreting the estimates for the parameters since the estimates for the intervention effect have been estimated assuming that equal weight is given to the results from each study regardless of the size (Brown *et al.* 1994). Also, if the sizes of the studies vary greatly within the meta-analysis then the results yielded from this model may differ considerably from a model where the interaction term is omitted. Additionally, since the interaction term is fitted as a fixed effect then if small studies were seen to give spurious results, the results from this model may be misleading since it would be assumed that the variation in the results across the trials would be due to random variation only.

3.5.1 Adjusting for covariates using individual patient data

A covariate may be included in the individual patient data random effect model. The covariate can either relate to a patient predictor, such as age of the patient; or to a study level predictor, such as method of administration for the intervention. However, care needs to be used in the interpretation of the covariate since they may describe both within-study and between-study relationships (Higgins *et al.* 2001). Additionally, if the patient level covariates included in the

model are the same for each patient within a study then they are essentially interpreted as a study level covariate.

Random effect models are presented below where a covariate is included in the model using individual patient data methods either to allow for an imbalance between the intervention groups or to assess potential sources of heterogeneity. The models are described in a classical setting however; the same models may also be used within a Bayesian framework using IPD. If a Bayesian framework is considered then it would be necessary for prior distributions to be specified for all of the parameters associated with the covariate.

3.5.1.1 Meta-regression

A random effect meta-regression model may be used based on individual patient data to assess the impact of a trial level covariate on the outcome measure (Whitehead 2002). The models described in Equations (2.29) and (2.30) may be extended to include a fixed effect trial level covariate x_{2i} (Equation 3.18 shown for continuous outcome measure).

$$y_{ij} = \alpha + \beta_{0i} + \beta_1 z_{ij} + \beta_2 x_{2i} + \beta_3 x_{2i} z_{ij} + u_{1i} z_{ij} + \varepsilon_{ij} \quad (3.18)$$

$$u_{1i} \sim N(0, \tau^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

However β_{0i} and $\beta_2 x_{2i}$ are not separately identifiable (Whitehead 2002), therefore a single fixed effect trial term can be written as

shown in Equation (3.19) (shown for continuous outcome measure); where ε_{ij} and u_{1i} are assumed uncorrelated.

$$y_{ij} = \alpha + \beta_{0i} + \beta_1 z_{ij} + \beta_3 x_{2i} z_{ij} + u_{1i} z_{ij} + \varepsilon_{ij} \quad (3.19)$$

$$u_{1i} \sim N(0, \tau^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

3.5.1.2 Imbalance in prognostic factors

An imbalance in prognostic factors between the intervention and control groups may arise in a particular study; therefore it may be advantageous to model the imbalance in an individual patient data random effect model. The models presented below are for continuous outcome measures, however, the models can be easily adapted to use for dichotomous outcome measures.

The random effect IPD model described in Equation (2.30) (shown for continuous outcome measures) may be extended to include a fixed effect term for the patient level covariate, x_{2ij} to assess the effect on an imbalance between the intervention groups (Equation 3.20) where ε_{ij} and u_{ij} are assumed uncorrelated.

$$y_{ij} = \alpha + \beta_{0i} + \beta_1 z_{ij} + \beta_2 x_{2ij} + u_{1i} z_{ij} + \varepsilon_{ij} \quad (3.20)$$

$$u_{1i} \sim N(0, \tau^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

3.5.1.3 Potential sources of heterogeneity

Potential sources of heterogeneity may be investigated to determine whether there are patient level factors which affect the size of the intervention effect. This may be achieved through modelling an interaction term between the intervention and patient level covariate using a random effect IPD model. The models presented below are for continuous outcome measures, however, the models can be easily adapted to use for dichotomous outcome measures.

Extending the model presented in Equation (3.20), the effect for the interaction term between the patient level covariate and the intervention is included as fixed effect and represented as $x_{2ij}z_{ij}$ (Equation 3.21). The ε_{ij} and u_{1i} terms are assumed to be uncorrelated.

$$y_{ij} = \alpha + \beta_{0i} + \beta_1 z_{ij} + \beta_2 x_{2ij} + \beta_3 x_{2ij} z_{ij} + u_{1i} z_{ij} + \varepsilon_{ij} \quad (3.21)$$

$$u_{1i} \sim N(0, \tau^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

Thompson and Sharp have investigated the effects of including a study level covariate in a random effect meta-analysis based on classical summary and IPD using the log odds ratio (Thompson *et al.* 1999); for the summary data methods they used MM, ML and REML to estimate the heterogeneity. They found the methods were comparable in the interpretations of the results however, they also found that the estimates for the intervention and covariate parameters

and their associated standard errors depended upon the estimate yielded for the heterogeneity. Slightly smaller estimates for heterogeneity were seen from the ML and REML methods as compared to modelling the IPD; conversely the MM gave a larger estimate.

3.5.2 Problems with exploring heterogeneity

Subgroup analysis is an attractive assessment to use since it allows for the studies to be grouped in subsets, however, problems may arise where there are insufficient numbers of studies in each subset, or where studies are misclassified into the subset.

Meta-regression analysis should be only treated as exploratory since distinct disadvantages are recognised; there is the potential for an association to be found between a covariate and the outcome event purely by chance or due to other confounding factors (Sutton *et al.* 1998). Also, this type of analysis is prone to aggregation bias if the covariate is measured at individual level but summarised and included in the meta-analysis at trial level (Greenland 1987). Aggregation bias occurs where the relationship between the covariate means for the summary data and the summary events do not directly reflect the relationships between covariate means for the patients and the patients events (Sutton *et al.* 1998). Additionally all of the estimation procedures used above for meta-regression suffer from the same inherent problems as detailed for each method in Sections (2.4.2) and (2.9.2).

Additionally, the classical methods for estimating point estimates for the between trial heterogeneity described in this section do not take into account variation in the estimate. Further methods have been proposed to calculate approximate confidence intervals for these heterogeneity estimates based on either moment estimates (Larholt *et al.* Unpublished) or maximum likelihood methods; however, both methods have been found to be restrictive in either requiring the individual sampling variances from each trial to be less than the estimate for heterogeneity, or are based on asymptotic methodologies (Biggerstaff *et al.* 1997). Unlike the classical method which assume that the between trial heterogeneity is known rather than estimated from the data, full Bayesian approach allows for the variation associated with the estimating heterogeneity from the data to be incorporated into the analysis. Additionally, credibility intervals for the between trial heterogeneity may be estimated from the model. Thompson and Sharp have recently compared the methods described here to explain heterogeneity in two meta-analyses where one trial level covariate was included and the outcome measure was the log odds ratio (Thompson *et al.* 1999). They found it was important to take into account residual heterogeneity not explained by the inclusion of a trial level covariate, and ignoring the residual heterogeneity was found to under-estimate the standard errors of the regression coefficients, and thereby overstating the importance of the covariate. The choice over which estimator for the between trial heterogeneity was less clear; for the classical methods they found the

maximum likelihood estimator were more asymptotically efficient but tended to under-estimate the between trial heterogeneity as compared to the REML method. Also, the weighted method of moment's technique which included the covariate was found to be relatively more complex to use to estimate heterogeneity than the simple method of moments technique without the covariate.

Both the empirical Bayes and a full Bayesian framework were considered in the Bayesian analyses; empirical Bayes was found to yield very large estimates for the between trial heterogeneity as compared to the classical methods and indeed the full Bayesian framework. The full Bayesian framework was found to take into account the imprecision of estimating the between trial heterogeneity from the data; however, this advantage of the method was found to have little impact on the results in practice. Therefore the authors concluded than they recommend using REML for most practical applications in practice (Thompson *et al.* 1999).

An obvious advantage of using individual patient data in a meta-analysis is that differences between the studies may be investigated in more depth than would be achieved from using summary data. However, this relies on all of the studies measuring the same patient level covariates. Also, this type of information can not usually be extracted from a publication and so the data needs to have been kept preferably in an electronic format.

Where data is missing at study level, it has been proposed that values of zero are assigned to the patients within the particular study

so that the difference in the characteristic between the intervention and control group is zero (Higgins *et al.* 2001); however, this assumes that the characteristic was balanced between the groups. Due to the nature of the methodologies used in a Bayesian framework, the Bayesian software allows for missing data at study and/or patient level to be included in the analyses by assuming a specific distribution for the variable, and is therefore advantageous over the classical methods.

Within the random effect IPD model, the effect for the covariate is included in the model as a fixed effect however; there is a choice over how the interaction between the covariate and the intervention effect is specified. A logical choice would be to specify a random effect term since the intervention has been specified as a random effect too however; problems arise if there are an insufficient number of studies being included in the meta-analysis due to deficient data being available to model the extra variance components.

3.6 Summary to chapter

This chapter has described the types of methodologies that can be used to visually and statistically assess the presence of heterogeneity in meta-analyses of summary data and IPD. Also, methods for including covariates in summary and IPD models have been presented and discussed using classical and Bayesian methodologies.

Three meta-analyses in the area of stroke medicine will be used in this thesis to describe and exemplify the models used in this and the previous chapters. The next chapter gives an overview to the area of stroke medicine, including the incidence, aetiology, risk factors and outcome measures that are associated with stroke.

CHAPTER 4

AN INTRODUCTION TO STROKE

4.1 Introduction

The World Health Organisation has defined stroke as 'a syndrome of rapidly developing clinical signs of focal (or global) disturbance of cerebral function, with symptoms lasting 24 hours or longer or leading to death, with no apparent cause other than that of vascular origin' (Hatona 1976). Stroke is the fourth most common cause of death and the primary cause of adult disability in the UK (Department of Health 2003). Despite this only a small proportion of funding is spent on stroke research compared with the quantity spent on cancer research (Rothwell 2001). However, recently the Government have considered stroke to be an important public health issue and stroke is now a core part of the National Service Framework for the elderly (Department of Health 2001).

4.2 Incidence and cost of stroke

Each year 1,825,000 strokes occur in the UK, USA and European Union (Sudlow *et al.* 1997), with approximately 100,000 first strokes occurring in Britain alone (King's Fund Consensus Statement 1988). It has been estimated that about 25% of men and 20% of women can expect to suffer a stroke if they live to be 85 years of age (Bonita 1992). However, two thirds of strokes are not fatal therefore the cost of stroke care is great when the lengths of hospital stays, possibly followed by community support or nursing home care are considered. About a third of stroke survivors are functionally dependent one-year

post stroke, and in the UK alone it has been estimated that there are about 250,000 disabled stroke survivors (Stroke Care 1998).

In terms of overall cost, stroke care consumes between 4-5% of health service expenditure. Although this proportion may remain the same, actual expenditure on stroke care is likely to rise in real terms over the next 20 years because of the effects of the ageing population (The Stroke Association 1996).

4.3 Aetiology and symptoms of stroke

A stroke follows when part of the brain becomes damaged as a result of a problem with the blood supply. The location of the stroke is seen in the lack of blood supply around a certain area in the brain and this determines which symptoms or complications are seen. The clinical diagnosis of stroke may be confirmed through a computed tomography (CT) scan, magnetic resonance imaging (MRI) scan, or autopsy. CT and MRI techniques allow for similar presenting conditions, such as tumour or infection, to be differentiated from stroke.

Around 80 per cent of strokes occur where there is a blockage (or occlusion) to an artery from a blood clot; this is known as an ischaemic stroke (Rudd *et al.* 2000). The remaining 20 per cent of strokes are due to either a bleed in the brain; an intracerebral haemorrhage, or a bleed onto the surface of the brain; a subarachnoid haemorrhage (Rudd *et al.* 2000). Patients may suffer a mini-stroke or transient ischaemic attack (TIA). The main difference between a mini-stroke and a full stroke is that the deficits or

symptoms seen following a mini-stroke are resolved within 24 hours of onset, with no lasting complications.

Most strokes present suddenly and without warning. The most common symptom is weakness to one side of the body, which may involve the face or limbs (Dennis 1988). Symptoms arise mainly from the stroke itself or through swelling (oedema). A variety of neurological symptoms may be present after stroke onset (Warlow *et al.* 1996), these include cortical symptoms such as difficulty swallowing, balance problems, difficulty understanding or expressing spoken language, dyslexia, or difficulty writing; sensory symptoms such as loss of vision in one side of the vision field, or total visual loss; and behavioural or cognitive symptoms such as confusion and forgetfulness. Other presentations include incontinence and loss of consciousness (Warlow *et al.* 1996).

The symptoms a patient has after stroke onset, such as weakness or speech problems, can help to determine which area of the brain has been damaged. Several classification methods exist, the most commonly used was developed by Bamford and colleagues (Bamford *et al.* 1991). This contains four subtypes of stroke; lacunar syndromes (LACS); posterior circulation syndromes (POCS); total anterior circulation syndrome (TACS); and partial anterior circulation syndrome (PACS).

Table 4.1 shows how the subtype of stroke may be determined, where the numbers represent the following deficits:

1a. One side of weakness and/or sensory deficit affecting face

1b. One side of weakness and/or sensory deficit affecting arm

1c. One side of weakness and/or sensory deficit affecting leg

2. Cortical symptoms/signs such as dysphagia, neglect

3. Loss of vision in one side of the vision field

4. Signs such as ataxia, discontugated eye movements

Subtype	Combination
LACS	Any two of 1
TACS	1 and 2 and 3
POCS	3 and/or 4
PACS	Any two of 1, 2, 3 Or any one of 1

Table 4.1 Determining the subtype of stroke using the Bamford classification

The subtypes of stroke are associated with the degree of damage to the brain and hence may be used to predict outcome. Patients diagnosed with a lacunar stroke (LACS) will generally have a milder stroke since only weakness or sensory deficits are seen. In contrast patients with a total anterior circulation syndrome (TACS) will have a more severe stroke because in addition to those seen for LACS, other symptoms relating to loss of vision and cortical symptoms are seen, and hence more of the brain is affected. This rating is also reflected in the outcome for these patients, where approximately 2 % of LACS patients die within 30 days of stroke as compared to 39% of TACS patients (Bamford *et al.* 1990). Patients diagnosed with partial

anterior circulation syndromes (PACS) and posterior circulation syndromes (POCS) have been found to have mortality rates of 4% and 7%, respectively, at 30 days post stroke (Bamford *et al.* 1990).

4.4 Risk factors for stroke

The main risk factors for stroke are increasing age, high blood pressure, exercise inactivity, cigarette smoking, high cholesterol levels, cardiovascular problems (such as angina or heart attack), atrial fibrillation (irregular heart rhythm) and diabetes (Warlow *et al.* 1996). A higher risk of stroke is also associated with genetic inheritance, especially if the relative had a stroke whilst they were relatively young (<50 years). This is thought to be related to not only genetic factors predisposing an individual, but may be also attributed to a family history of high cholesterol level and diabetes. Lower social class and ethnicity have been found to increase the risk of stoke where people from an African-Caribbean, African or southern Asian ancestry are at a higher risk of stroke than people with White ethnicity.

The British Hypertension Society has suggested that optimal levels for blood pressure, irrespective of age, should be <140mm Hg for systolic blood pressure and <85mm Hg for diastolic blood pressure (Williams *et al.* 2004). Blood pressure is modified by many factors such as obesity, salt intake, and lack of exercise (Rudd *et al.* 2000).

4.5 *Prognostic factors for outcome*

About a third of the patients which have a stroke will die shortly after stroke onset. Another third of patients will make a complete or near-complete recovery given time. The remaining third of patients will remain functionally dependent following a stroke. A patients' outcome following stroke has been linked with factors such as smoking, old age, high blood pressure, cardiovascular problems and diabetes. In addition to these, other factors which are related to a poor outcome following stroke include urinary incontinence, history of a previous stroke, pre-morbid disability, impaired level of consciousness after stroke onset, total anterior circulation syndrome (TACS), large stroke lesion, presence of cognitive impairment, severe motor deficit, and visuospatial dysfunction (Hier *et al.* 1991).

4.6 *Treatments following stroke*

Various drug treatments have been developed for treating stroke patients, these include antiplatelet drugs and anticoagulant drugs, which make blood less sticky and reduce clotting; and antihypertensive drugs to lower blood pressure. Rehabilitation is used as a treatment towards improving outcome after stroke, and utilises the expertise of the multidisciplinary team, comprising of an occupational therapist, physiotherapist, speech and language therapist and psychologist. They help to aid physical recovery and encourage independence by managing physical, emotional and social effects of stroke.

4.7 Stroke outcome measures

Many outcome measures are important in the area of stroke medicine. Since one third of patients will require help with functional abilities following a stroke, mortality is not the only important measure of effectiveness of an intervention for stroke patients. Disability measures were introduced into stroke trials in the 1960's and the use of the measures has gradually increased. Any measure that is used to assess outcome should be shown to be valid and reliable; however, many stroke trials have used non-validated outcome measures (Roberts *et al.* 1998).

Recently, the World Health Organisation (WHO) has updated the classification for the International Classification of Functioning, Disability and Health, or the ICF (http://www.who.int/classification/icf/intros/icf-Eng-Intro.pdf). The ICF is now separated into 2 sections, each containing two components. Section 1 deals with the broad area of functioning and disability, and is subdivided into two components; body functions and structures, and activities and participation. The body functions and structures component assesses the changes in physiological functions and anatomical structures. In contrast, the activities and participation component assess the capacity and performance of executing tasks in the standard and current environment. The second section to the ICF concentrates on aspects of contextual factors of an individual's life and living. This section is subdivided into two components; environmental factors, and personal factors, which deal with the

external and internal influences on functioning and disability, respectively.

Other outcome measures which fulfil the areas the ICF identifies include rating scales for handicap levels, quality of life, mood levels, satisfaction levels, health care costs and disability, the length of stay in hospital, and the discharge destination after hospital. These measures are widely acknowledged outcome assessments for stroke patients (Gompertz *et al.* 1993). Wade and colleagues have collated stroke scales and comment that the scales used to investigate the efficacy of a particular intervention should be valid, reliable and sensitive to clinically relevant changes. In addition to this, they should also be simple to administer, and easily communicated to non-specialists and consumers, including patients and relatives (Wade 1986).

A variety of stroke outcome measures will be used in the examples in this thesis which include the activities of daily living, instrumental activities of daily living, leisure participation and activity, minor psychiatric measures, recurrent stroke, and death and dependency outcome measures.

4.7.1 Activities of daily living

Self-care tasks such as continence, dressing, personal hygiene and basic mobility, are referred to as personal activities of daily living (ADL). The Barthel Index was developed in 1965 by Mahoney and Barthel (Mahoney *et al.* 1965) and later modified by Granger (Granger *et al.* 1979). The scale measures the patients' performance

in 10 activities of daily living items. The items are divided into two components; the first component is related to activities of self-care and includes assessing dependence in feeding, grooming, bathing, dressing, bowel and bladder continence and toilet use. The second component relates to mobility levels and assesses ambulation, transferring, and stair climbing. The maximum score is 100 if the 5-point increment is used, indicating that the patient is fully independent in physical functioning; the lowest score is zero and relates to a totally dependent bed-ridden state. Due to the ordinal nature of the outcome measure, conventionally cut-off scores have been used to define different categories. A cut-off of 60 on the Barthel Index corresponds to a shift from dependence to assisted dependence (Granger *et al.* 1979). The Barthel Index has been shown to be a valid and reliable measure to use for stroke patients as a postal self-reported questionnaire (Gompertz *et al.* 1994) and for use over the telephone (Wade 1992). Other values have been used to dichotomise the data from the Barthel Index, including a cut of 80 has been recommended in accordance with the European Stroke Database (http://www.ncl.ac.uk/stroke-research-unit/posters/bsrcsmx.htm).

Another ADL measure is the self-care section of the Rivermead Index (Whiting *et al.* 1980). The Rivermead ADL is very similar to the Barthel Index and is directed specifically towards the stroke population although it does not assess continence of bowels or bladder. The Rivermead ADL scale is based on the patient's performance and is scored on a 3-point scale; dependent,

independent but requires verbal supervision; or independent. These ADL measures concentrate upon actual observed behaviour and not potential abilities (Wade 1992), and do not measure why patients fail to achieve certain tasks or achieve independence.

4.7.2 Instrumental activities of daily living

Activities of daily living scales such as the Barthel Index have some weaknesses which are mainly related to ceiling and flooring effects, where patients either make a complete recovery or remain severely disabled. Therefore, measures are needed which assess the higher levels of activity (Duncan *et al.* 2000).

Instrumental activities of daily living tasks include more complex activities required to live in the community such as walking outside, cooking, household management, and the ability to engage in social activities (Gladman *et al.* 1993). The Extended Activities of Daily Living scale is an example of an instrumental ADL scale and was developed for use in stroke patients. The extended ADL scale is a 22-item questionnaire and consists of four uni-dimensional sub-scales assessing mobility (6 items), kitchen (5 items), domestic (5 items) and leisure abilities (6 items). The questions were designed to place emphasis on whether the patient did the activity, and not on if they could do it, and therefore looks at activity rather than capability (Nouri *et al.* 1987). The scale was designed to be used as a postal questionnaire, and has been shown to be reliable and valid and therefore it is attractive as an outcome measure.

4.7.3 Leisure participation and activity

Leisure interests and hobbies used to be viewed as a way that the patients fill in their time and not thought to be of interest as an outcome measure. However, literature has suggested that leisure participation is related to life satisfaction (Allen *et al.* 1984), and forms a positive aspect of life. Leisure participation has been shown to decline with age, and after disabling conditions, such as stroke (Drummond 1990). The Nottingham Leisure Questionnaire (NLQ) was developed to measure leisure participation in stroke patients, and to monitor effects of interventions (Drummond *et al.* 1994). The NLQ is a 37-item questionnaire where each item represents an activity, the frequency that the activity is performed is also recorded using five possible categories (very regularly, regularly, occasionally, infrequently, never). It has been shown to be reliable and valid when administered by a therapist (Drummond *et al.* 1994) but has not been tested for use as a postal questionnaire. However, recently the 37-item questionnaire has been shortened to a more compact 30-item questionnaire (Drummond *et al.* 2001), by removing some questions which had either a low prevalence or did not represent positive activities, such as 'day-dreaming' or 'just sitting'. Also, the response categories that represented the frequency of activity were collapsed down to three (regularly, occasionally, never). The shortened version was found to be valid for use as a postal questionnaire and had good reliability rates when administered in this manner (Drummond *et al.* 2001).

4.7.4 Common mental disorders

Many patients will suffer depression after stroke and therefore this should clearly feature in any outcome assessment. The General Health Questionnaire (Goldberg 1972) is a self-administered screening questionnaire. The questionnaire measures depression as well as somatic symptoms, anxiety and insomnia, and social dysfunction by determining a critical number of key symptoms rather than a particular symptom. This questionnaire has been modified and validated for use in stroke patients (Ebrahim 1985).

4.7.5 Recurrent stroke

The estimate risk of a recurrent stroke within 5 years of the initial stroke is 17% (Hillen *et al.* 2003), and was found to be related to stroke patients having ischaemic heart disease, atrial fibrillation, or diabetes. Increased risks of recurrence were associated with older patients. Similar rates of recurrence were seen across subtypes of stroke (Hillen *et al.* 2003). Also, stroke patients which are admitted to hospital with high blood pressure were found to have an increased risk of recurrence within 14 days of the initial stroke (Leonardi-Bee *et al.* 2002).

4.7.6 Combined death and functional scales

A common outcome measure for stroke trials is the combined outcome of death or dependency, or death or disability. This is thought to reflect a 'poor outcome' where not only mortality is considered, but also the level of dependency or disability of the stroke

survivors. Dependency is usually measured using a validated outcome measure of activities of daily living, such as the Barthel Index; and disability is commonly measured using a scale such as the modified Rankin scale (van Swieten *et al.* 1988). The Rankin scale is a validated outcome measure which is based on a six point scale which ranges from no disability, through increasing levels of disability, and finally death; and measures independence rather than performance of specific tasks (van Swieten *et al.* 1988).

An introduction to the data sets

The thesis will focus on analysing data from three meta-analyses of randomised controlled clinical trials. Each meta-analysis has been designed to answer a specific key question in the area of stroke management and hence all of the patients included in each of the meta-analyses have had at least one cerebrovascular event, such as a stroke or transient ischaemic attack.

The first meta-analysis (Chapter 5) is concerned with patients who have very recently had a stroke. The healthcare question the meta-analysis addresses is whether altering a patients' blood pressure in the acute phase of stroke, using medication, is beneficial in increasing the patients' chance of survival at long term follow-up.

The second meta-analysis (Chapter 6) concentrates on patients who have been discharged from hospital following a stroke. The healthcare question this meta-analysis is targeting is whether occupational therapy in the community setting is effective in increasing the patients' abilities at performing everyday household and leisure tasks in stroke patients.

The last meta-analysis (Chapter 7) involves patients who have previously suffered a stroke or transient ischaemic attack. The healthcare question this meta-analysis investigates is whether a combination of antiplatelet agents, dipyridamole and aspirin, will reduce the risk of subsequent stroke.

CHAPTER 5

ALTERING BLOOD PRESSURE IN ACUTE STROKE: A SYSTEMATIC REVIEW AND META-ANALYSIS OF RANDOMISED CONTROLLED TRIALS

5.1 Introduction

High blood pressure (BP), defined by the World Health Organisation as a BP>140/85 mmHg, is present in 75-80% of patients with acute ischaemic and haemorrhagic stroke (Wallace *et al.* 1981; Britton *et al.* 1986; Leonardi-Bee *et al.* 2002). The mechanisms underlying hypertension in stroke remain unclear but include pre-existing hypertension, activation of neuroendocrine systems (sympathetic, glucocorticoid, mineralocorticoid), increased cardiac output, the stress of hospitalisation, and the Cushing reflex (reactive increases in systemic blood pressure in response to a raised intracranial pressure) (Carlberg *et al.* 1991; Harper *et al.* 1994; Boreas *et al.* 2001).

Many studies, and a meta-analysis of them, have found that high BP, whether measured casually or using 24 hour ambulatory readings, is associated with a poor outcome, judged as an increase in death or combined death and disability or dependency (Carlberg *et al.* 1993; Dandapani *et al.* 1995; Robinson *et al.* 1997; Leonardi-Bee *et al.* 2002; Willmot *et al.* 2004). High BP is also associated with a greater risk of recurrence, and possibly of developing fatal cerebral oedema, after ischaemic stroke (Leonardi-Bee *et al.* 2002).

The frequency of high BP in acute stroke and its independent association with a poor outcome, suggest that BP should be lowered. However, impaired cerebral autoregulation is present during the acute phase of stroke (Meyer *et al.* 1973), such that regional cerebral blood flow becomes passively dependent on arterial blood pressure. Hence, it has been hypothesised that lowering BP might further reduce

perfusion and worsen outcome leading to considerable debate as to whether BP should be lowered acutely (Spence *et al.* 1985; Yatsu *et al.* 1985; International Society of Hypertension Writing Group 2003). Other factors also need to be considered, including drug class, and timing of administration (Bath *et al.* 1997). Although no large and definitive randomised controlled trials assessing the management of BP in acute stroke have been completed, many studies have investigated the effect of vasoactive drugs in acute stroke and an integrated analysis of these might further elucidate the relationship between altering BP and outcome. Therefore the 'Blood pressure in Acute Stroke Collaboration' (BASC) project was initiated to investigate relationships of vasoactive drugs in acute stroke using a series of systematic reviews (Bath *et al.* 1997).

5.1.1 Blood pressure in Acute Stroke Collaboration

The objective of BASC was to determine whether altering blood pressure in patients with acute stroke is safe and effective in reducing the risk of death, and death or dependency. To determine the effects of vasoactive drugs on blood pressure three phases were undertaken.

The first phase was to identify and assessed the relationships between vasoactive drugs and outcome where the primary aim of the individual trials was to alter blood pressure in patients with acute stroke (Blood pressure in Acute Stroke Collaboration (BASC) 2001(b)). Five small randomised controlled trials were identified which contained six drug classes. Of these only calcium channel blockers

were found to significantly impact on lowering blood pressure; however no significant effect on outcome was seen in any of the trials.

The second phase of BASC was to identify and assess the effects of vasoactive drugs in the acute treatment of stroke where the primary aim of the trial was not necessarily to assess the drug's effect on blood pressure (The Blood pressure in Acute Stroke Collaboration (BASC) 2001(a)). Sixty five completed randomised controlled trials were identified and data were obtained from 32 of these. Calcium channel blockers, ACE-inhibitors, beta blockers, glyceryl trinitrate and prostacyclin were found to reduce blood pressure within 2 days as compared to control. However, magnesium, naftidrofuryl and piracetam did not appear to alter early blood pressure. Death was found to be increased in trials which assessed either streptokinase, beta blockers or piracetam. However, no noticeable effects were seen for any of the drug classes for death or dependency.

5.2 Design of study

The third phase of BASC is to perform a systematic review and meta-analysis of data from randomised controlled trials of vasoactive drugs in patients with acute stroke, where measurements of systolic blood pressure were recorded. It was decided to use systolic blood pressure rather than diastolic or mean arterial blood pressure because systolic is thought to be more precise in measuring blood pressure. This is because some automated measurements of blood pressure, such as the OMRON, use the systolic blood pressure

measurement to calculate the diastolic measurement, and hence its derivatives, such as mean arterial blood pressure and pulse pressure.

5.2.1 Measures of interest

The primary outcome measure is death at the end of trial. Secondary outcome measures include death or dependency/disability at the end of trial.

5.3 Study selection and search strategy

For a trial to be eligible for inclusion in this systematic review it has to fulfil the following criteria: (i) randomisation, (ii) controlled, (iii) vasoactive drug which would be expected, on pharmacological grounds, to lower blood pressure, (iv) patients with ischaemic or haemorrhagic stroke within two weeks of onset, (v) baseline systolic blood pressure measurements. Vasoactive drugs with hypotenstive properties included alpha-receptor antagonists, angiotensin converting enzymes (ACE) inhibitors, angiotensin II receptor antagonists, beta-receptor antagonists calcium channel blockers (CCB), dipyridamole, diuretics, magnesium, naftidrofuryl, nitrates, papaverine, pentoxifylline, prostacyclin, serotonin receptor antagonists, sympathomimetics, theophylline, thromboxane antagonists, vinpocetine, and their derivatives.

Trials were excluded from the review if they were controlled but confounded trials where active treatments were compared (without a control group), or if they included patients with subarachnoid haemorrhage.

A comprehensive literature search, using a strategy developed by the Cochrane Stroke Group (Sandercock *et al.* 2003) was performed to identify trials using The Cochrane Library (2003 Issue 1), MEDLINE (1966-March 2003), EMBASE (1980-March 2003) and Web of Knowledge (1981-March 2003). Searches were also made of the reference list of identified trials, reviews of hypertension in acute stroke and the UK National Research Register. The project was also advertised at meetings and in journals (Blood pressure in Acute Stroke Collaboration 2003). No restrictions on the language of the publication were made.

5.4 Data collection and management

Identified trials were assessed independently by the secretariat of the collaboration. Contact authors were approached about joining BASC and sharing their individual patient data. If contact with collaborators was not made, repeat invitations were sent out; strenuous efforts were made to find trialists who had moved or to identify another senior author.

Thirty-six eligible trials were identified involving in 8,058 acute stroke patients (Figure 5.1). Outcome data from all of these trials were either abstracted from the original publication (Martin *et al.* 1985; Gelmers 1988; Herrschaft 1988; Paci *et al.* 1989; Bogousslavsky *et al.* 1990; Martinez-Villa *et al.* 1990; Murphy *et al.* 1990; Mohr *et al.* 1992; Kramer *et al.* 1994; Wimalaratna *et al.* 1994; De Deyn *et al.* 1997; Lamsudin *et al.* 1997; Bogousslavsky *et al.* 2002; Lees *et al.* Personal communication) or through the sharing of individual patient data

(Wester *et al.* 1984; Huczynski *et al.* 1985; Uzuner *et al.* 1985; Pokrupa *et al.* 1986; Steiner *et al.* 1986; Barer *et al.* 1988; Azcona *et al.* 1990; Gray *et al.* 1990; Limburg *et al.* 1990; Autret *et al.* 1992; Kaste *et al.* 1994; Norris *et al.* 1994; Wahlgren *et al.* 1994; Muir *et al.* 1995; Squire *et al.* 1996; Steiner *et al.* 1996; Dyker *et al.* 1997; Muir *et al.* 1998; Bath *et al.* 2001; Rashid *et al.* 2003; Barer *et al.* Unpublished; Lowe *et al.* Unpublished).

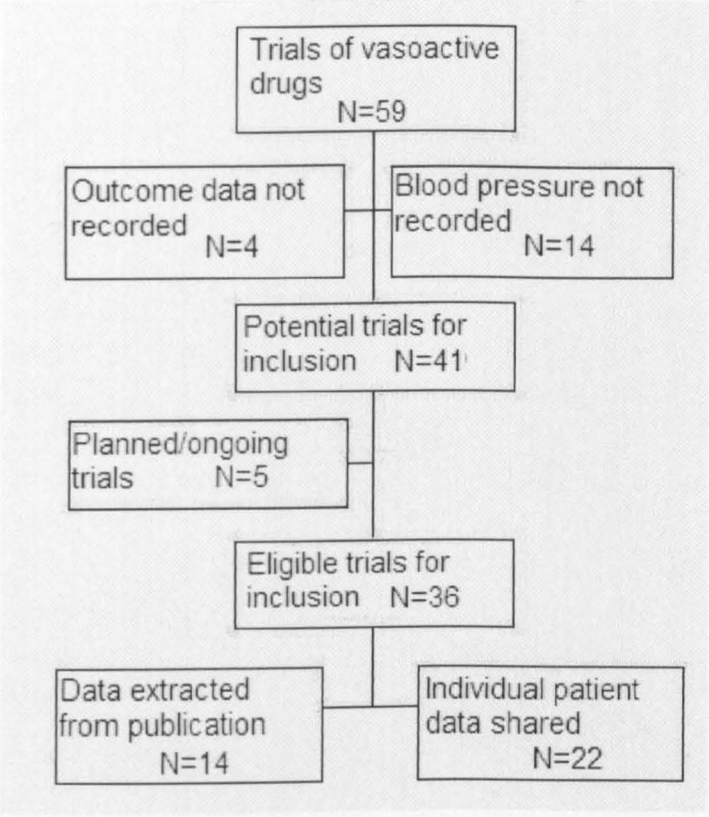

Figure 5.1 Flow chart for trial identification and selection

However, data on blood pressure were only available for 34 of the trials; the data from the other two trials had been previously lost (Martin *et al.* 1985) or the authors could not be contacted (Gelmers 1988).

Trial, year of publication	Subjects	Vasoactive treatment	Individual data
ASCLEPIOS, 1990	230	Isradipine	✓
Autret, 1992	312	Naftidrofuryl	✓
Bath, 2001	37	Glyceryl trinitrate	✓
BEST pilot, 1988	55	Atenolol/propananol	✓
BEST, 1988	302	Atenolol/propananol	✓
Bogousslavsky, 1990	52	Nimodipine	×
Bogousslavsky, 2002	293	Fibroblast	×
Dyker, 1997	28	Perindopril	✓
Gelmers, 1988	186	Nimodipine	×
Gray, 1990	100	Naftidrofuryl	✓
Herrschaft, 1988	40	Piracetam	×
Huczynski, 1985	30	Prostacyclin	✓
IMAGES, pilot 1998	51	Magnesium sulphate	×
INWEST, 1994	281	Nimodipine	✓
Kaste, 1994	355	Nimodipine	✓
Kramer, 1994	482	Nimodipine	×
Lamsudin, 1997	150	Nimodipine	×
Lees, 1995	60	Magnesium sulphate	✓
Limburg, 1990	26	Flunarizine	✓
Lowe, 1992	112	Nimodipine	✓
Martin, 1985	32	Prostacylin	×
Martinez-Vila, 1990	164	Nimodipine	×
Mohr, 1992	1064	Nimodipine	×
Muir, 1995	25	Magnesium sulphate	✓
Norris, 1994	189	Nimodipine	✓
Paci, 1989	41	Nimodipine	×
PASS, 1997	927	Piracetam	×
Pokrupa, 1986	23	Prostacyclin	✓
PRISTINE, 1996	620	Naftidrofuryl	✓
Rashid, 2003	90	Glyceryl trinitrate	✓
Squire, 1996	147	Lifarizine	✓
Steiner, 1986	100	Naftidrofuryl	✓
Strand, 1984	26	Magnesium sulphate	✓
TRUST, 1990	1215	Nimodipine	×
Uzuner, 1995	88	Nimodipine	✓
Wimalaratna, 1994	125	Nimodipine	×

Table 5.1 Trial characteristics for the identified trials

Five further other trials were excluded because they were ongoing at the time of analysis (Bath 2001; Lees *et al.* 2001; COSSACS 2003; Robinson *et al.* 2003; Willmot *et al.* Unpublished).

Shared individual patient data and abstracted data were checked for consistency, re-coded to ensure uniformity and merged into a single database using SAS version 8.02 (SAS Institute Inc). Combined death and disability/dependency were defined *a priori* as: Barthel Index<60 or adapted 7 point Rankin Scale>3 (Steiner *et al.* 1986). Some trials used their own scale and a dichotomous variable was created according to the grand median score for that trial.

5.5 *Assessment of publication bias and quality*

Begg's funnel plot was used to assess visually if there was evidence of publication bias within the meta-analysis. The plot suggested that there was a lack of smaller sized studies which showed detrimental effects. Additionally, two of the point estimates were located outside the pseudo 95% confidence interval lines; hence there may be a suggestion of publication bias in this review (Figure 5.2).

Begg and Mazumdar rank correlation test found no evidence of publication bias (p=0.307, 36 trials), similarly the trim and fill method found that no trimming needed to be performed; however, Egger's asymmetry test found evidence of publication bias (p=0.026, 36 trials).

Methodological quality for each trial was assessed on the basis of the method of randomisation employed, whether the allocation to treatment was concealed, the completeness of follow-up achieved, and whether

the outcome assessment was blinded to the allocation of treatment using recognised criteria (The Cochrane Collaboration 2003).

Thirty-one of the 36 trials were deemed to be of high quality, the remaining five were found to be of a moderate quality due to either inadequate allocation of concealment for treatment assignment or unclear methods of randomisation (Table 5.2).

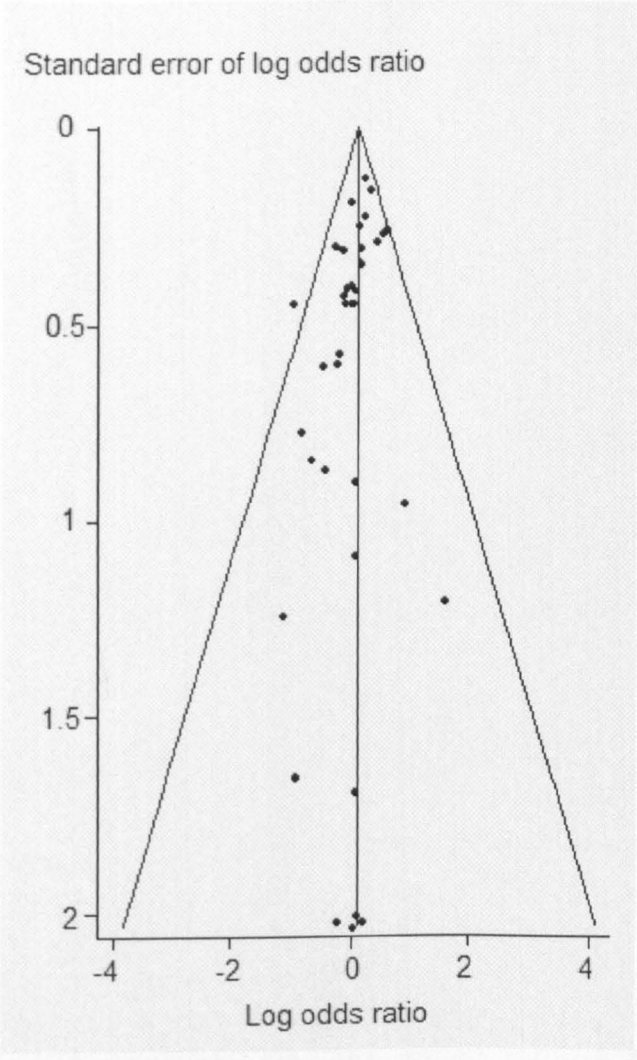

Figure 5.2 Begg's funnel plot for death at the end of trial

Trial, year of publication	Allocation to concealment	Adequate randomisation	Placebo controlled
ASCLEPIOS, 1990	A	Yes	Yes
Autret, 1992	B	Yes	Yes
Bath, 2001	A	Yes	Yes
BEST pilot, 1988	B	Yes	No
BEST, 1988	A	Yes	Yes
Bogousslavsky, 1990	A	Yes	Yes
Bogousslavsky, 2002	A	Yes	Yes
Dyker, 1997	A	Yes	Yes
Gelmers, 1988	A	Yes	Yes
Gray, 1990	A	Yes	Yes
Herrschaft, 1988	A	Yes	Yes
Huczynski, 1985	A	Unclear	Yes
IMAGES, pilot 1998	A	Yes	Yes
INWEST, 1994	A	Yes	Yes
Kaste, 1994	A	Yes	Yes
Kramer, 1994	A	Yes	Yes
Lamsudin, 1997	A	Yes	Yes
Lees, 1995	A	Yes	Yes
Limburg, 1990	A	Yes	Yes
Lowe, 1992	A	Yes	No
Martin, 1985	A	Yes	Yes
Martinez-Vila, 1990	A	Yes	Yes
Mohr, 1992	A	Yes	Yes
Muir, 1995	A	Yes	Yes
Norris, 1994	A	Yes	Yes
Paci, 1989	A	Yes	Yes
PASS, 1997	A	Yes	Yes
Pokrupa, 1986	A	Yes	Yes
PRISTINE, 1996	A	Yes	Yes
Rashid, 2003	B	Yes	No
Squire, 1996	A	Yes	Yes
Steiner, 1986	A	Yes	Yes
Strand, 1984	A	Yes	Yes
TRUST, 1990	A	Yes	Yes
Uzuner, 1995	B	Yes	No
Wimalartna, 1994	A	Yes	Yes

A Low risk of bias, B Moderate risk of bias, C High risk of bias

Table 5.2 Assessment of quality for the 36 trials

5.6 *Trial level demographics*

Trials were assessed to determine whether patients were similar between the trials. The demographics of the patients across the trials appeared to be relatively well balanced (Table 5.3). The trial by Herrschaft and colleagues (Herrschaft 1988) recruited slightly younger patients (mean 57 years) than the other trials whereas the IMAGES pilot trial recruited slightly older patients (mean 75 years) (Lees *et al.* Personal communication). The majority of the trials recruited more males than female patients. The two extremes for the percentages of males ranged from 35% males (Limburg *et al.* 1990) to 73% males (Bogousslavsky *et al.* 1990). The mean baseline systolic blood pressures from the trials were above 140 mm Hg indicating that the patients generally had higher than optimal blood pressures.

Trial, year of publication	Age (years) [mean, (SD)]	Gender, male (%)	SBP (mmHg) [mean (SD)]
ASCLEPIOS, 1990	69 (10)	146 (62)	159 (27)
Autret, 1992	68 (8)	189 (61)	162 (28)
Bath, 2001	74 (9)	18 (49)	161 (22)
BEST pilot, 1988	71 (10)	28 (51)	149 (27)
BEST, 1988	69 (11)	159 (52)	156 (27)
Bogousslavsky, 1990	65 (-)	38 (73)	147 (-)
Bogousslavsky, 2002	70 (12)	175 (61)	157 (-)
Dyker, 1997	70 (10)	17 (61)	172 (22)
Gelmers, 1988	70 (-)	111 (60)	- (-)
Gray, 1990	67 (9)	46 (46)	154 (25)
Herrschaft, 1988	57 (-)	27 (61)	165 (-)
Huczynski, 1985	61 (12)	16 (53)	159 (31)
IMAGES, pilot 1998	75 (9)	19 (37)	160 (-)
INWEST, 1994	72 (10)	136 (46)	160 (28)
Kaste, 1994	58 (9)	236 (66)	156 (26)
Kramer, 1994	63 (-)	258 (54)	154 (-)
Lamsudin, 1997	- (-)	96 (64)	168 (-)
Lees, 1995	68 (13)	30 (50)	155 (27)
Limburg, 1990	67 (14)	9 (35)	160 (32)
Lowe, 1992	68 (9)	66 (59)	157 (27)
Martin, 1985	68 (-)	- (-)	- (-)
Martinez-Vila, 1990	72 (-)	70 (57)	152 (-)
Mohr, 1992	66 (-)	617 (58)	151 (-)
Muir, 1995	70 (10)	16 (64)	160 (26)
Norris, 1994	72 (10)	104 (55)	147 (25)
Paci, 1989	63 (-)	28 (69)	153 (-)
PASS, 1997	71 (-)	479 (52)	160 (-)
Pokrupa, 1986	63 (13)	11 (48)	150 (20)
PRISTINE, 1996	72 (9)	321 (52)	161 (26)
Rashid, 2003	72 (12)	41 (46)	152 (20)
Squire, 1996	69 (13)	82 (56)	157 (24)
Steiner, 1986	69 (7)	54 (54)	147 (25)
Strand, 1984	74 (11)	14 (54)	162 (32)
TRUST, 1990	73 (-)	555 (46)	157 (-)
Uzuner, 1995	63 (12)	44 (50)	145 (24)
Wimalaratna, 1994	70 (-)	- (-)	159 (-)

- Data not available from publication

Table 5.3 Patient demographics at baseline by trial

5.7 *Outcome assessments and measures*

All of the trials recorded death at the end of trial; however for the secondary outcome measure death or dependency at the end of trial, this was only available for 21 of the trials (Huczynski *et al.* 1985; Uzuner *et al.* 1985; Steiner *et al.* 1986; Barer *et al.* 1988; Herrschaft 1988; Azcona *et al.* 1990; Limburg *et al.* 1990; Murphy *et al.* 1990; Norris *et al.* 1994; Wahlgren *et al.* 1994; Muir *et al.* 1995; Squire *et al.* 1996; Steiner *et al.* 1996; De Deyn *et al.* 1997; Muir *et al.* 1998; Bath *et al.* 2001; Bogousslavsky *et al.* 2002; Rashid *et al.* 2003; Lees *et al.* Personal communication; Barer *et al.* Unpublished; Lowe *et al.* Unpublished).

5.7.1 Type of intervention

The drugs given in the trials may be categorised by class based on mechanism of action. Nine classes of vasoactive drugs were involved in the 26 trials; these were calcium channel blockers, beta blockers, naftidrofuryl, nitric oxide donor, magnesium, prostacyclin, piracetam, fiblast growth factor, and perindopril (Table 5.4). The majority of trials involved assessing the efficacy of a calcium channel blocker. The vasoactive drugs were either given orally, intravenously or transdermally using patches.

5.7.2 Timing of assessments

The timings for the outcome assessments varied between the trials. The median time to the end of trial was 12 weeks (range 2 weeks to 52 weeks) (Table 5.4).

Trial, year of publication	Drug Class	Recruitment window (hours)	Timing of assessment (weeks)
ASCLEPIOS, 1990	CCB	12	12
Autret, 1992	Naftidrofuryl	72	24
Bath, 2001	Nitric oxide donor	120	12
BEST pilot, 1988	Beta blocker	48	24
BEST, 1988	Beta blocker	48	24
Bogousslavsky, 1990	CCB	48	4
Bogousslavsky, 2002	Fiblast	6	12
Dyker, 1997	Perindopril	168	2
Gelmers, 1988	CCB	24	4
Gray, 1990	Naftidrofuryl	48	24
Herrschaft, 1988	Piracetam	120	4
Huczynski, 1985	Prostacyclin	72	4
IMAGES, pilot 1998	Magnesium	12	4
INWEST, 1994	CCB	24	24
Kaste, 1994	CCB	48	52
Kramer, 1994	CCB	48	4
Lamsudin, 1997	CCB	24	2
Lees, 1995	Magnesium	12	12
Limburg, 1990	CCB	24	24
Lowe, 1992	CCB	48	24
Martin, 1985	Prostacyclin	36	2
Martinez-Vila, 1990	CCB	48	4
Mohr, 1992	CCB	48	24
Muir, 1995	Magnesium	24	12
Norris, 1994	CCB	42	52
Paci, 1989	CCB	12	4
PASS, 1997	Piracetam	12	12
Pokrupa, 1986	Prostacyclin	48	4
PRISTINE, 1996	Naftidrofuryl	48	52
Rashid, 2003	Nitric oxide donor	72	12
Squire, 1996	CCB	12	13
Steiner, 1986	Naftidrofuryl	168	52
Strand, 1984	Magnesium	36	24
TRUST, 1990	CCB	24	24
Uzuner, 1995	CCB	24	2
Wimalaratna, 1994	CCB	24	24

Table 5.4 Trial characteristics for the 36 trials

5.7.3 Time between stroke onset and recruitment into study

All trials recruited patients within a pre-specified time limit from the onset of stroke. The median maximum time from stroke onset to recruitment was 45 hours (range 6-168 hours) (Table 5.4).

5.8 Analysis of the merged data set using summary data methods

The combined data set contained individual patient information from 36 trials and comprised of 8,058 stroke patients. Patients received either a vasoactive drug (n=4,494), or placebo/control (n=3,564).

5.8.1 Data analysis for death at the end of trial

The proportions of patients which died at the end of trial in each treatment group within each trial were used as the primary outcome measure. The odds ratios and variances for each trial were calculated as detailed in section 2.3.1.1.

5.8.2 Results for death at the end of trial

In a conventional fixed effect analysis as described in section 2.4.1, patients randomised to a vasoactive drug had non-significantly increased risk of death as compared to patients which received control/placebo (OR 1.10, 95% CI 0.98, 1.23).

5.8.3 Results for secondary outcomes

From a conventional fixed effect analysis as described in section 2.4.1, patients randomised to a vasoactive drug had a non-significant increase

in the risk of death and dependency at the end of trial (OR 1.11, 95% CI 0.98, 1.24).

5.9 Assessment of heterogeneity using summary data

The conventional methods presented above assume that each of the trials is estimating a single underlying treatment effect. An assessment of heterogeneity needs to be performed to evaluate whether this assumption is justified. Visual and statistical assessments used for death at the end of trial.

5.9.1 Graphical assessment of heterogeneity

The forest plot indicates that there is some heterogeneity between the trial estimates when compared overall (Figure 5.3). Therefore, to formally assess the presence of heterogeneity between the estimates, statistical testing of heterogeneity was performed.

5.9.2 Statistical assessments of heterogeneity using classical methods

To formally assess whether there was evidence of heterogeneity between the trials the Cochran's homogeneity Q test and Higgins and Thompson I^2 were performed. Also, other analyses were used to quantify the heterogeneity between the trial estimates using un-weighted and weighted MM, ML, REML, EB and FB (Table 5.5). For the FB model, assessments of convergence were adequate using a burn-in of 10,000 and a sample chain of 40,000 iterations.

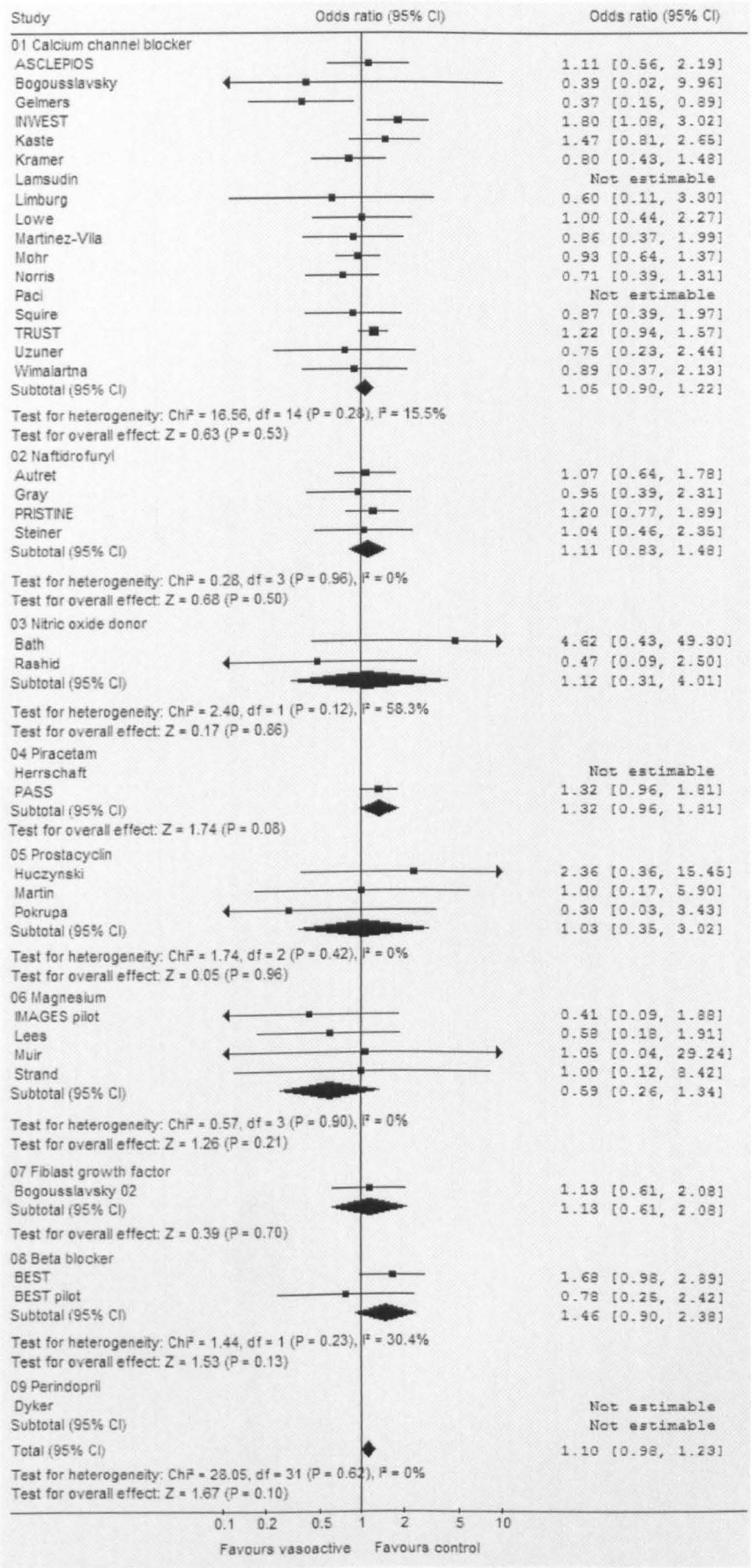

Figure 5.3 Forest plot for death at the end of trial

Heterogeneity Test	χ^2	P value	I^2
Cochran's Q test	28.05	0.62	
I^2			0
Estimation methods	Odds Ratio	95% CI	$\hat{\tau}^2$
MM, un-weighted	1.102	0.983, 1.236	0
MM, weighted	1.102	0.983, 1.236	0
ML	1.102	0.983, 1.236	0
REML	1.102	0.983, 1.236	0
EB	1.102	0.983, 1.236	0
FB	1.072	0.923, 1.229‡	0.139

Odds ratio is the odds ratio of mortality on vasoactive drug relative to control. ‡ 95% credibility interval

Table 5.5 Heterogeneity test and analysis results for death at the end of trial based on summary data

Cochran's homogeneity test did not find evidence of heterogeneity between the trial estimates (p=0.62). I^2 was estimated as zero for the overall comparison (Table 5.5) indicating that none of the variability between the trial estimates could be attributed to heterogeneity.

From Table 5.5, the estimates for heterogeneity from the classical and EB methods were all zero, and hence the estimates for the treatment effect were identical as to those from the fixed effect models; indicating that there was a small but non-significant 10% increase in the risk of death at the end of trial associated with vasoactive drug use.

The FB approach found a small quantity of heterogeneity, hence the standard errors were slightly larger in magnitude as compared to the other models; however, the treatment effect remained non-significant at the 5% level.

5.10 *Exploring heterogeneity*

It appears that the difference in the estimates for the treatment effects between the trials may be due to sampling variation only and not due to heterogeneity. Although these tests did not find evidence of heterogeneity, it may still be advantageous to determine whether any prognostic factors either relating to patient characteristics, such as baseline blood pressure, or trials factors, such as drug class, can explain some of the residual variation between the trials. Subgroup analyses were used to assess the influence of trial factors, and meta-regression analyses were used to assess the influence of patient characteristics.

5.10.1 Subgroup analyses

Subgroup analyses were performed to assess the relationship between death at the end of trial and two trial characteristics; class of the vasoactive drug used, and the time from stroke onset to recruitment.

5.10.1.1 Class of vasoactive drugs

Trends towards increases in the risk of death were seen for all of the drug classes (Figure 5.3), except magnesium where a non-significant 41% decrease in the risk of death was associated with a vasoactive drug as compared to control/placebo (OR 0.59, 95% CI 0.26, 1.34).

The Q value for the change in heterogeneity was 0.751; indicating that there does not appear to be any evidence of a difference between the drug classes for death at the end of trial.

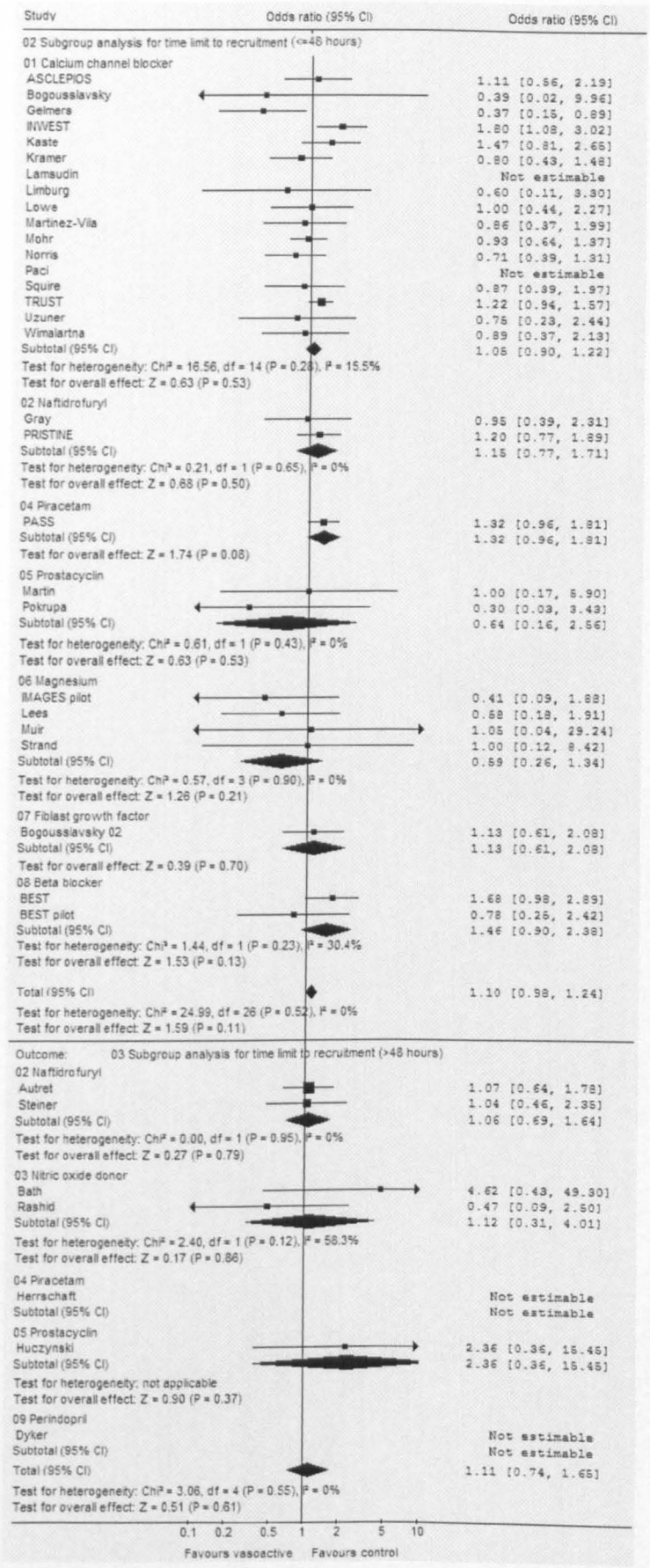

Figure 5.4 Subgroup analyses for death by the time limit for stroke onset

5.10.1.2 Time from stroke onset to recruitment

The inclusion criteria for each trial involved recruiting patients within a time limit, and this limit varied between the trials. Trials were grouped into either ≤48 hours or >48 hours (Figure 5.4). The estimates for death at the end of trial were very similar between the two groups (≤48 hours OR 1.10, 95% CI 0.98, 1.24; >48 hours OR 1.11, 95% CI 0.74, 1.65). The p value for the change in heterogeneity was 1.00; also indicating no differences in the risk of death between the two time limit to recruitment groups.

5.10.2 Meta-regression analyses

The patient characteristics of interest were systolic BP at baseline and the early change in systolic BP. Early change in BP was defined as either the difference between 24 hours BP and baseline BP, or the difference between 48 hour BP and baseline BP (in the cases where 24 hour BP was not recorded in the trials). Baseline BP and early change in BP were assessed using the absolute values for each of the factors, additionally the difference in the absolute values between the vasoactive and control/placebo groups for each trial were used for baseline BP to assess the effect of mismatched baseline BP's between the treatment groups.

For the analyses, baseline BP was centred at 160 mm Hg, the difference in BP was centred at 0 mm Hg, and the percentage change in early BP was centred at 0%.

Method	Odds ratio[§]	95% CI	SBP0[Ω]	95% CI	$\hat{\tau}^2$
MM, approx	1.267	1.089, 1.473	1.035	1.007, 1.065	0
MM, weighted	1.267	1.089, 1.473	1.035	1.007, 1.065	0
ML	1.267	1.089, 1.473	1.035	1.007, 1.065	0
REML	1.267	1.089, 1.473	1.035	1.007, 1.065	0
EB	1.267	1.089, 1.473	1.035	1.007, 1.065	0
FB	1.265	1.047, 1.514	1.038	1.005, 1.072	0.131
Method	Odds Ratio[§]	95% CI	†SBP0[Ω]	95% CI	$\hat{\tau}^2$
MM, approx	1.129	1.005, 1.268	1.018	0.987, 1.050	0
MM, weighted	1.129	1.005, 1.268	1.018	0.987, 1.050	0
ML	1.129	1.005, 1.268	1.018	0.987, 1.050	0
REML	1.129	1.005, 1.268	1.018	0.987, 1.050	0
EB	1.129	1.005, 1.268	1.018	0.987, 1.050	0
FB	1.107	0.954, 1.276	1.019	0.984, 1.055	0.144
Method	Odds Ratio[§]	95% CI	ΔSBP%[Ω]	95% CI	$\hat{\tau}^2$
MM, approx	0.863	0.582, 1.279	0.967	0.920, 1.017	0
MM, weighted	0.863	0.582, 1.279	0.967	0.920, 1.017	0
ML	0.863	0.582, 1.279	0.967	0.920, 1.017	0
REML	0.863	0.582, 1.279	0.967	0.920, 1.017	0
EB	0.863	0.582, 1.279	0.967	0.920, 1.017	0
FB	0.843	0.513, 1.368	0.968	0.911, 1.029	0.193

§ Odds of death on vasoactive drug relative to control at the centred covariate value, Ω Multiplicative increase in mortality OR when covariate is increased by 1 unit, SBP0 baseline systolic BP, †SBP0 difference in SBP between treatment groups, ΔSBP% early percentage change in SBP, ‡ 95% credibility intervals

Table 5.6 Death at the end of trial using summary data meta-regression methods

Although heterogeneity was only found in the FB model, a series of random effect meta-regression models were used to explore the relationships between death at the end of trial and the covariates (see section 3.4). The methods used to estimate heterogeneity were the weighted and approximate MM, ML, REML, EB and FB. For the FB model, assessments of convergence of were found adequate when using a burn-in of 10,000 and a sample chain of 40,000 iterations.

Similar estimates for the treatment and covariates effects were seen from all of the models for each of the covariates. Slight differences between the magnitudes of the treatment and covariate effects between the classical and Bayesian models are related to the Bayesian models estimating heterogeneity where slightly smaller odds ratios for the treatment effect were seen from these models as compared to the classical models (Table 5.6).

In the baseline systolic BP adjusted model, the covariate was centred to 160mm Hg. A significant relationship was seen between baseline systolic BP and the treatment effect which implied that there was a 4% increase in the estimated odds ratio per every 10 mm Hg increase in systolic BP (95% CI 1%, 7%) (Table 5.6). However, it did not appear that the difference in SBP between the treatment groups or the early percentage change in the vasoactive group was related to the treatment effect at the end of trial (Table 5.6).

5.11 *Individual patient data methods*

In a classical fixed effect IPD model, patients randomised to a vasoactive drug were found to have a trend towards a higher risk of death at the end of trial as compared to those randomised to control (OR 1.10, 95% CI 0.98, 1.24) (Table 5.7).

	Odds ratio (OR)	95% CI	$\hat{\tau}^2$
Fixed effect model			
Classical approach	1.103	0.983, 1.237	-
Random effect model			
Classical approach	1.096	0.972, 1.237	0.002
Bayesian approach	1.097	0.922, 1.256‡	0.032

OR is the odds of mortality on vasoactive drug relative to control, ‡ credibility intervals.

Table 5.7 Death at the end of trial using individual patient data methodologies

5.12 *Assessment of heterogeneity using individual patient data methods*

Assessments of heterogeneity between the trials were explored using an interaction term between trial and treatment group in the above fixed effect model. From this interaction term model, it appeared that there was no significant difference between the trials (p=0.643). For consistency of analyses, heterogeneity was estimated using both classical and Bayesian methodologies based on random effect IPD models. For the FB models, assessments of convergence were made using criteria as described in section 2.5.3, and adequate assessments were achieved when a burn-in of 5,000 iterations and a sample chain of 50,000 iterations were used.

Similar results were seen from the classical and Bayesian IPD random effect models for the treatment effect which implied that vasoactive drugs were non-significantly associated with a 10% increase in the risk of death at the end of trial as compared to control/placebo. Heterogeneity was estimated from both the classical and Bayesian models, however, the magnitude of the heterogeneity did not impact greatly on the results since they were similar to the fixed effect classical IPD model (Table 5.7).

5.13 Exploring heterogeneity using individual patient data methods

Both classical and Bayesian methods found quantifiable amounts of heterogeneity between the trials. Therefore trial level covariates, such as drug class and time to recruitment; and patient level covariates, such as baseline systolic BP and the early change in systolic BP, were included in the random effect models to assess their influence on death at the end of trial. For the FB models, assessments of convergence were made using criteria as described in section 2.5.3, and adequate assessments were achieved when a burn-in of 5,000 iterations and a sample chain of 50,000 iterations were used.

5.13.1 Trial level covariates using individual patient data

Drug class was modelled using a classical random effect IPD meta-regression model as described in section 3.5.1.3 to assess its impact on death at the end of trial. No differences in death at the end of trial were seen between the classes of drug considered in this review (p =0.916).

5.13.2 Patient level covariates using individual patient data

Classical and Bayesian models were performed to assess the impact of time from stroke onset to recruitment, baseline systolic BP and early change in systolic BP on death at the end of trial using random effect IPD models. For the FB models, adequate assessments of convergence were seen when a burn-in of 10,000 iterations and sample chains of 50,000 iterations were used.

Initially, an interaction term between the covariate and treatment effect terms was included in the models to ascertain whether the treatment effect varied across the range of covariates. Two of the models found no evidence on an interaction therefore the interaction was subsequently dropped from these models (baseline SBP interaction p=0.249, SBP early change interaction p=0.940). However, some evidence of a possible interaction was seen with time from stroke onset to recruitment and hence these two separate models were used where time from stroke onset to recruitment was split into ≤48 hours and >48 hours (p=0.060).

In the covariate adjusted models presented in Table 5.8, the results for the treatment effect and covariate effects between the classical and Bayesian models were very similar.

In the baseline systolic BP adjusted model, patient randomised to a vasoactive drug had a significantly higher risk of death at the end of trial by 14% as compared to those randomised to control/placebo. Additionally, the results indicated that there was a 4% increase in the

risk of death at the end of trial for every 10mm Hg increase in baseline SBP (Table 5.8).

In the early change in systolic BP model, the covariate didn't appear to be influential in the risk of death at the end of trial.

The random effect models were then adjusted for time from stroke onset to recruitment into the trial as a continuous measure, however the models were split into ≤48 hours and >48 hours from stroke onset. Patients that were recruited within 48 hours were 2% significantly less likely to have a poor outcome as the time to recruitment increased for every hour (95% CI 1%, 3%). However, this effect was not observed in patients that were recruited after 48 hours since stroke onset (Table 5.8).

IPD Models	Odds ratio§	95% CI	Estimate$^{\Omega}$	95% CI	$\hat{\tau}^2$
	Vasoactive drug vs. control		SBP0		
REML	1.138	1.012, 1.279	1.004	1.001, 1.007	0
FB	1.139	1.006, 1.289‡,	1.004	1.001, 1.007‡	0.085
	Vasoactive drug vs. control		ΔSBP%		
REML	1.125	0.922, 1.372	1.004	0.997, 1.010	0.014
FB	1.093	0.876, 1.361‡	1.004	0.997, 1.010‡	0.059
	Vasoactive drug vs. control		Time to recruitment (≤48 hours)		
REML	1.053	0.916, 1.209	0.981	0.970, 0.992	0.005
FB	1.020	0.856, 1.195‡	0.976	0.965, 0.987‡	0.034
	Vasoactive drug vs. control		Time to recruitment (>48 hours)		
REML	0.574	0.050, 7.110	0.979	0.917, 1.045	1.994
FB	0.782	0.052, 10.67‡	0.974	0.913, 1.051‡	2.654

§ Odds of death on vasoactive drug relative to control at the centred covariate value, Ω the estimate is the multiplicative increase in the odds of mortality for each 1 unit increase in the covariate, for those receiving a vasoactive drug or for those receiving control, the covariates are SBP0 baseline systolic BP, ΔSBP% early change in systolic BP, ‡ 95% credibility intervals

Table 5.8 Death the end of trial using individual patient data adjusted for patient level covariates

5.14 Discussion of findings

The discussion of the findings from this systematic review are presented in six sections which relate to a summary of the overall findings, limitations of the data used in the review, comparison of the statistical methods, discussion of the results, practical implications of the results and suggestions for future research.

5.14.1 Summary of overall findings

The management of blood pressure during the acute phase of stroke remains an enigma and will continue to be widely debated. In the absence of completed large and definitive randomised controlled trials assessing this question, we have performed a systematic review of existing randomised controlled trials.

Thirty-six randomised controlled trials of vasoactive drugs in acute stroke where blood pressure assessments had been made were identified from a comprehensive search strategy which was performed up to March 2003. Individual patient data were available from 24 of these trials; data from the other 12 trials had been previously discarded or we were unable to contact the authors of the trial. However, tabulated data from these 12 trials were extracted from the publications and merged with the individual patient data from the other trials.

The data base comprised of data from 8,058 stroke patients, and the patients from these trials were randomised to either a vasoactive drug or control/placebo. Baseline characteristics of the patients appeared

to be well balanced between the treatment groups from the individual trials.

At the end of trial, patients randomised to a vasoactive drug had slightly higher risk of death and death or dependency as compared to patients randomised to control/placebo; however these effects were not statistically significant at the 5% level.

5.14.2 Limitations of the data used

5.14.2.1 Number of studies included in the meta-analysis

The search strategy used in this systematic review identified thirty-six randomised controlled trials of a vasoactive drug given to stroke patients in the acute phase of stroke where blood pressure measurements had been assessed. Outcome data for these trials were available from all of the trials either through sharing individual patient data or extracted from publications. However, systolic blood pressure measurements at baseline were available from only 34 of the 36 trials; the data from the other two trials could not be extracted from the publications and the authors from the original articles could not be contacted.

We decided to assess the impact of vasoactive drugs which were had potential blood pressure lowering effects; therefore, studies which involved drugs which increase blood pressure, such as diaspirin cross-linked haemoglobin (DCLHb), where excluded. This was because they are thought to increase the risk of death and death or dependency in acute stroke patients (Saxena *et al.* 1997; Saxena *et al.* 1999; The Blood pressure in Acute Stroke Collaboration (BASC)

2001(a)) due to them actively increasing systolic blood pressure, which increases the risk of recurrence, possibly cerebral oedema (Leonardi-Bee *et al.* 2002) or haemorrhage to the infarct (Rordorf *et al.* 1997).

5.14.2.2 Timings of the outcomes

It was decided *a priori* to use the end of trial as the primary outcome assessment time. This was to allow for a consistent end point to be used across the trials and thought preferable to using an arbitrary end point, such as 12 weeks, since there was a lack of data at a consistent time point across the trials.

5.14.2.3 Extent of the shared individual patient data

Although the original trials collected considerable amounts of data on prognostic factors and outcome during the baseline and at the end of trial, the data shared with the collaboration were not consistent across the trials. All of the trials had data for death at the end of trial and time from stroke onset to recruitment, therefore the unadjusted analyses and the analyses adjusted for time from stroke onset to recruitment should have yielded results which have minimal bias. However, bias could be present in the results involving BP measurements. Although baseline systolic BP was not available from two of the trials, the data from these trials only comprised of 3% of the total data, hence would probably have little effect on the overall results. Conversely, early change in blood pressure could only be calculated for 24 of the trials, therefore this could have had a major impact on the results when

assessing the relationship between early change in BP and death at the end of trial.

Finally, we were unable to adjust for other potential prognostic factors, such as diabetes or atrial fibrillation (see section 4.4); since these data were not available for most of the identified studies. Hence, some sources of heterogeneity will have not been explored in the analyses.

5.14.2.4 Outcome assessments

The primary and secondary outcome assessments used in this systematic review were selected *a priori* before the first two initial phases of the review were implemented following discussions with the collaborators of the Blood pressure in Acute Stroke Collaboration (BASC).

A trial was deemed eligible for inclusion into the systematic review if it recruited stroke patients within 2 weeks of stroke onset. This criterion was used to ensure that the vasoactive treatments would be given in the acute phase of stroke. Death at the end of trial was chosen as the primary outcome measure because it was the primary end point for the majority of the trials and thought to be the best measure of efficacy for an acute stroke treatment. Death or dependency at the end of trial was included in this review as the secondary outcome measure. This was to allow for a more detailed analysis of functional ability on the patients which had survived till the end of trial. However, due to the lack of data for this outcome where data were available

from only 21 of the 36 trials; further in-depth analyses could not be performed. Also, data on early death (within 30 days of treatment), intermediate events, such as recurrence, or time to an event, could not be assessed in this review due to a lack of data from the trials.

5.14.2.5 Publication bias

Systematic reviews are susceptible to several biases, including publication bias. In this review evidence of publication bias was found from using Egger's Asymmetry test (Egger *et al.* 1997). From Egger's Asymmetry plot, there appeared to be some evidence of a lack of smaller trials which found detrimental effects, however; it was not apparent from either Begg and Mazumdar's rank correlation test (Begg *et al.* 1994) or the non-parametric trim and fill method (Duval *et al.* 2000).

Publication bias relating to an absence of the smaller trials which showed a detrimental effect through not being published is a common finding. This is because the earlier trials are more likely to be smaller in size, hence more likely to be published if they showed a beneficial effect. Also it is acknowledged that subsequent larger trials are more likely to have been conducted if the earlier smaller trials showed beneficial effects. Although there may be a lack of these smaller trials, the overall pooled estimate for the treatment effect from the 36 trials showed a trend towards an increase in the risk of death and hence the absent smaller trials would endorse the overall results and possibly make the detrimental effect statistically significant at the 5% level.

Evidence of publication bias may be related to other reasons, such as poor methodological quality of the included trials (Petticrew *et al.* 1999; Sutton *et al.* 2000). Further analysis from using a sensitivity analysis indicated that excluding trials with poor methodological quality appeared to reduce some of the publication bias, however some publication bias remained which still indicated a lack of small studies which found detrimental effects associated with using a vasoactive drug in acute stroke patients (Egger's test for Asymmetry p=0.04). Additionally, excluding these poor methodological trials resulted in a borderline significant increase in the risk of death at the end of trial in patients randomised to a vasoactive treatment as compare to those randomised to control (OR 1.11, 95% CI 0.99, 1.26; 31 trials).

5.14.3 Comparison of the statistical methods

5.14.3.1 Assessments of heterogeneity

The assessment of heterogeneity on the results was initially investigated using a forest plot, which revealed that there appeared to be little heterogeneity between the trials. Statistical assessments of heterogeneity were performed based on summary data methods, which consistently found no heterogeneity between the results, from using Cochran's homogeneity test and I^2. No evidence of heterogeneity was seen from using an individual patient data model by including a fixed effect interaction term between the trial and treatment. This systematic review contained data from 36 trials and hence all of these methods should have had sufficient power, hence it

appears that the differences between the trials may be due to sampling variation alone.

This was a surprising finding due to the assortment of drug classes that were included in the review which are known to have varying mechanisms for action in lowering blood pressure. Therefore, it was interesting to find that biological heterogeneity does not preclude statistical heterogeneity in this meta-analysis (Glasziou *et al.* 2002).

5.14.3.2 Impact of quantifying heterogeneity

A range of estimation methods were used in a series of random effect models based on summary data to assess whether there were any differences between the results from the models. All of the classical and EB models estimated no heterogeneity between the trials. The FB model estimated a small amount of heterogeneity however the magnitude had little impact on the results of the treatment effect, hence all of the models performed adequately in this meta-analysis.

Fixed and random effect IPD models were then used to assess the impact of vasoactive drugs on death at the end of trial. In a fixed effect model, patients randomised to a vasoactive drug had a 10% non-significant increase in the risk of death as compared to patients which received control/placebo. Due to the random effect models estimating very small quantities of heterogeneity the results from these models were very similar as to the fixed effect model.

5.14.3.3 Impact of exploring heterogeneity

Various estimations models were used in a series of meta-regression models to assess whether there were any treatment effect modifiers, and whether there were any differences between the results from the estimation methods.

All of the methods yielded similar results for the treatment and covariate effects for each covariate separately. A significant effect was seen which implied that baseline systolic BP had an effect on the treatment effect, where there was a 4% increase in the estimated odds ratio per 10 mm Hg increase in baseline systolic BP. No significant effects were seen on treatment for early change in systolic BP or the difference in baseline systolic BP between the groups. The clinical implications for the effect of baseline systolic BP on the treatment effect estimate will be discussed in section 5.14.4.2.

From the adjusted IPD models, although no evidence of effect modifiers were seen for baseline systolic BP or early change in systolic BP; there appeared to be some evidence of an effect with the time to recruitment from stroke onset. This finding appeared to be contrary to the results from the subgroup analysis using summary data methodologies. Therefore, subsequent analyses of time to recruitment were split into two groups for the IPD analyses; ≤48 hours and >48 hours. In patients randomised within 48 hours of stroke onset, patients appeared to be at a lower risk of death at the end of trial for each additional hour from their stroke onset. The clinical

implications for the effect of time to recruitment will be discussed in section 5.14.4.3.

Additionally, although baseline systolic blood pressure did not appear to be an effect modifier, it appeared to be related to death at the end of trial; where a 4% increase in the risk of death was seen per 10 mm Hg increase in systolic BP irrespective of intervention assignment. Also, in this model patients randomised to a vasoactive drug were significantly at a 14% higher risk of death at the end of trial as compared to patients randomised to control/placebo.

5.14.4 Discussion of the results

This systematic review provides substantial evidence for the management of acute stroke with regards to the efficacy of vasoactive drugs given to stroke patients in the acute phase of stroke. The aim of this review was to estimate the extent that vasoactive drugs impact on death at the end of trial and to assess its relationship with systolic BP.

5.14.4.1 Impact of vasoactive drug on death, and death or dependency

The principal finding from this systematic review was that vasoactive drugs given in the acute phase of stroke was non-significantly associated with a 10% increase in the risk of death and 11% increase in the risk of death or dependency at the end of trial. When the models were adjusted for baseline systolic blood pressure, treatment with a vasoactive drug was found to be significantly associated with a

14% increase in the risk of death at the end of trial. This suggests that patients may be at a higher risk of a poor outcome if they are treated in lowering blood pressure acutely after stroke onset.

Cerebral autoregulation is a concept that implies that cerebral blood flow is maintained a constant levels even when there are changes in systemic perfusion, such as those caused by a stroke (Feldmann *et al.* 1999). Autoregulation also ensures that there is a constant supply of oxygen, glucose, and other essential nutrients and that capillary pressure is kept at an optimal level (Feldmann *et al.* 1999). However, cerebral autoregulation is lost during stroke and therefore lowering blood pressure in hypertensive patients could reduce vital penumbral blood flow, thus leading to poor outcome from neurological deterioration secondary to reduced cerebral perfusion (Rashid 2003).

The British Hypertension Society recommends that antihypertensive treatment should be initiated in people with sustained systolic blood pressure ≥ 160 mm Hg (Williams *et al.* 2004); which was also the mean systolic blood pressure of the patients included in this review. However, the results from this review would suggest that these patients may be at a higher risk of mortality and morbidity if blood pressure lowering treatments are initiated too acutely after stroke onset.

Two other systematic reviews have assessed the effect of lowering blood pressure and outcome. The first systematic review of randomised controlled trials assessed the impact of blood pressure lowering or treatment of hypertension in patients with non-acute

stroke (>14 days post-ictus). We found that lowering blood pressure was associated with significant reductions in stroke, non-fatal stroke, myocardial infarction and total vascular events; whilst overall mortality was not altered (Rashid *et al.* 2003).

The second systematic review of published reviews found that antihypertensive treatments showed efficacy in reducing the risk of morbidity and mortality from first ever stroke (Lawes *et al.* 2004). Bearing in mind the results from this review, the findings from these reviews can't be extrapolated to assume that similar benefits would be seen in acute stroke patients and therefore it is suggested that long term benefit of secondary stroke prevention by treating hypertension acutely may be inappropriate and detrimental to mortality and morbidity.

There appeared to be little variation in the risk of a poor outcome between the trials included in this review, and the findings from this review compare well to those of previous meta-analyses using aggregate data (The Blood pressure in Acute Stroke Collaboration (BASC) 2001(a); Blood pressure in Acute Stroke Collaboration (BASC) 2001(b)).

5.14.4.2 Impact of baseline systolic blood pressure

From the meta-regression analyses based on summary data, increasing baseline systolic blood pressure was significantly related to an increase in the relative odds ratio for the treatment effect. This effect appears to be contrary to popular belief that a patient with a

higher blood pressure receiving a vasoactive drug will have a higher risk of a poor outcome; and does not appear to be replicated in the individual patient data analyses since no interaction between systolic blood pressure and treatment was seen.

However, potential explanations for the finding from the meta-regression analysis exist; the result may be related to the fact that high blood pressure is associated with an increased risk of fatal cerebral oedema and recurrent stroke (Leonardi-Bee *et al.* 2002) and so the effect seen is really a reflection of these intermediate effects; however these effects could not be assessed in this review because few trials recorded them.

Alternatively, previous research has indicated that there is the possibility with meta-regression analyses of finding false-positive results which are not replicated from analyses based on individual patient data (Sutton *et al.* 1998; Lambert *et al.* 2002; Higgins *et al.* 2004). Sutton and colleagues suggest that meta-regression analyses should only be treated as exploratory since there are distinct disadvantages (Sutton *et al.* 1998). These are due to the potential for an association to be found between the summary level covariate and treatment effect purely by chance or due to other confounding factors or due to aggregation bias (Greenland 1987), where the relationship between the covariate means for the summary data and the treatment effect do not directly reflect the relationships within an individual (Sutton *et al.* 1998).

This review also highlighted that baseline systolic blood pressure was an important predictor of outcome when an individual patient data model was considered; where a 4% increase in the risk of death was seen for every 10mm Hg increase in systolic blood pressure irrespective of treatment assignment.

High blood pressure is commonly seen in patients presenting with a stroke (Wallace *et al.* 1981; Britton *et al.* 1986; Leonardi-Bee *et al.* 2002) and is thought to be the direct result of the cerebrovascular events itself, either relating to a severe neuroendocrine stress response or due to hypoperfusion of the brain tissue (Carlberg *et al.* 1991). However, is has also been postulated that the increase in blood pressure may be due a transient phenomenon due to acute mental stress associated with being admitted to hospital or accident and emergency (Carlberg *et al.* 1991). Raised blood pressures usually fall spontaneously within a few days (Wallace *et al.* 1981; Morfis *et al.* 1997); and within 10 days post stroke two thirds of patients will be normotensive (Carlberg *et al.* 1991).

The relationship and magnitude of this effect has also been found in a large randomised controlled trial involving 17,398 acute stroke patients (Leonardi-Bee *et al.* 2002). This large trial was able to look at the relationships between blood pressure and intermediate events, and found that recurrent ischaemic stroke within 14 days post ictus and presumed cerebral oedema were independently associated with high systolic blood pressure. Therefore, one could extrapolate these findings from this randomised controlled trial to our review and

suggest that the relationship between high systolic blood pressure and mortality may be due to an increase in intermediate events such as recurrence or cerebral oedema.

Increased blood pressure at baseline does not appear to be subject to a previous hypertension, since only a third of patients presenting with stroke have a history of hypertension (Oppenheimer *et al.* 1992), and other studies have found no relationship between previous hypertension and either early death, or death and dependency (O'Connell *et al.* 1994).

5.14.4.3 Impact of time from stroke onset to recruitment

Patients randomised to a vasoactive drug early and within 48 hours post-ictus were more likely to have a poor outcome. This finding is probably confounded by severity, where more severe patients are admitted to hospital earlier and hence recruited into a trial earlier. Data on prognostic factors such as severity were not available from enough trials to allow this issue to be further examined. Similarly, the effect of timing may have been confounded by drug class since patients in calcium channel blocker trials were always treated early.

This finding was seen only in the individual patient data and not present in the summary data model where it was assessed as a dichotomous cut at 48 hours. Therefore, it appears that the individual patient data model may have been more sensitive in assessing the relationship between time to recruitment and outcome since the actual time of the event was used where individual patient data

allowed and the analyses did not solely rely on the time limit for recruitment for an individual trial.

5.14.4.4 Generalisability of the findings

Research spanning several decades has been performed to help identify factors which influence outcomes, and reach conclusions about the optimal management of acute stroke patients. Blood pressure management in the acute phase of stroke is one of the areas where there has been a lot of interest and research.

This systematic review and meta-analysis conducted an extensive search strategy to identify all eligible randomised controlled trials which have monitored blood pressure levels and assessed the efficacy of vasoactive drugs. Although it is likely that all of the available data has been identified, one of the assessments for publication bias indicated asymmetry where small trials which showed a detrimental effect of vasoactive drugs were under-represented in this review. The inclusion of such trials would have not significantly altered the overall finding that vasoactive drug given in the acute phase of stroke worsened outcome at the end of trial, but instead endorse these results. There are currently five ongoing randomised controlled trials that need to be included in any future analyses. However, the results from this review should be generalisable since they form a collective body of evidence from thirty-six randomised controlled trials involving blood pressure monitoring and assessing vasoactive drugs in stroke patients.

The findings from this meta-analysis provide a balanced interpretation of the available evidence and support the results from earlier studies that high systolic blood pressure at baseline is associated with a poor outcome after stroke (Dandapani *et al.* 1995; Robinson *et al.* 1997; Leonardi-Bee *et al.* 2002). However, our research provides more information about the relationship between a variety of vasoactive drugs and outcome and shows the first results between patient characteristics and outcomes.

5.14.5 Practical implications

This review has demonstrated that vasoactive drugs given to patients in the acute phase of stroke can be detrimental to their outcome, especially when baseline systolic blood pressure is taken into consideration. Therefore this review suggests that caution is used when administering blood pressure lowering agents to acute stroke patients.

Actively lowering blood pressure too acutely after stroke onset may be harmful in ischaemic stroke patients because they may decrease blood perfusion in the ischaemic border zone (Wood 1984). However, there is evidence that there is an increased risk of brain oedema if an elevated blood pressure is not treated (Wallace *et al.* 1981). Therefore, it is important to identify those patients which require urgent treatment of their hypertension and it has been recommended that stroke patients should only be treated with vasoactive drugs when very high blood pressures are seen acutely (systolic blood pressure >200 mm Hg) (Lavin 1986) or where the raised blood

pressure (SBP ≥160 mm Hg) is persistent over the two week since stroke onset (Williams *et al.* 2004).

In depth analyses were performed in this review which highlighted that certain subgroups of patients may be at the greatest risk of a poor outcome from receiving a vasoactive drug. Analyses indicated that increased risks in mortality were associated with increasing systolic blood pressure at baseline, and a shorter time from stroke onset within 48 hours post ictus. Surprisingly, the early change in systolic blood pressure did not appear to mediate the treatment effect or outcome at the end of trial. However, these factors were identified through post hoc analyses and the findings need to be validated from other data of a randomised controlled trial to truly assess whether these patients are at increased risks of mortality.

Also, since baseline blood pressure appears to be influential on the efficacy of the vasoactive drug, this highlights the importance of using adequate randomisation techniques such as minimisation or stratification on baseline SBP to ensure that the estimates from the trial are not overly influenced by mismatching of baseline blood pressure. Many of the trials within this review had a mismatch of baseline blood pressure between the treatment groups.

5.14.6 Future research

In spite of the negative finding of this study, namely that active lowering of BP appears to be associated with a worse outcome, data from large and definitive trials of lowering BP are underway (Bath

2001; Lees *et al.* 2001; COSSACS 2003; Robinson *et al.* 2003; Willmot *et al.* Unpublished).

It is unclear whether the relationship between systolic blood pressure and death at the end of trial identified from this review, is a causal association; and it must be remembered that the results from this review relating to blood pressure measurements could be biased and imprecise since most trials were not intending to alter blood pressure so its measurement was often poor and incomplete. This is because most of the studies included in this review were not aiming to alter BP and assess its effect on outcome, and therefore were probably under-powered, and not able to adequately assess important issues such as timing and dosing of treatment, efficacy in different types of stroke (ischaemic, haemorrhagic, cortical, lacunar), continuation of previous antihypertensive drugs, and baseline blood pressure. Future trials will need to study these issues as well as effects on cerebral perfusion, recurrence and cerebral oedema. Most of the existing studies in this review have data from drug classes which are now known to have no beneficial effects in acute stroke, such as calcium channel blockers (Horn *et al.* 2001), and naftidrofuryl (Stenier *et al.* 1996). In contrast, little data exist for other drug classes where a stronger rationale exists for their potential efficacy, such as nitrates (Willmot *et al.* 2003) and angiotensin receptor antagonists.

CHAPTER 6

COMMUNITY OCCUPATIONAL THERAPY FOR STROKE PATIENTS: A SYSTEMATIC REVIEW AND META-ANALYSIS OF RANDOMISED CONTROLLED TRIALS

6.1 Introduction

Rehabilitation is defined by the WHO (Aho *et al.* 1980) as 'the combined and co-ordinated use of medical, social, educational and vocational measures for training or retaining the individual to the highest level of functional ability'. Occupational therapy is an essential component of the rehabilitation package, and offers a wide range of interventions designed to promote recovery through purposeful activity by focusing on disability (Walker *et al.* 1999). Occupational therapists encourage patients to practise all forms of activities of daily living (ADL) such as dressing and feeding. However, occupational therapists also promote independence in extended activities of daily living (EADL), such as household management and leisure pursuits.

After a stroke, patients often lose the abilities to perform ADL and EADL tasks. Occupational therapists are one of the multidisciplinary team which may provide practice to regain these abilities in the hospital and community after discharge from hospital. However, at present it is unclear whether community based rehabilitation services such as those offered by occupational therapists are beneficial to the patient. There is an increase in emphasis for these services to enhance early discharge from hospital therefore it has become more important for these community based services to grow. Therefore, an evaluation of these community based services is of fundamental importance to the delivery of efficient evidence based stroke care.

Several clinical trials have been performed to assess the efficacy of occupational therapy in stroke patients after discharge from hospital.

A meta-analysis of occupational therapy using data from the literature has already been published (Steuljens *et al.* 2003). This large review which aimed to determine whether occupational therapy interventions improve outcome of stroke patients, identified a small but significant effect size for the efficacy of comprehensive occupational therapy on self-care, EADL, and social participation. However, trials of occupational therapy have disparate findings with regards to whether occupational therapy was effective in increasing ADL and EADL abilities. The variation in findings may be due to the type and intensity of intervention. Internal trial factors relating to the design and protocol, such as length of follow up; or patient's factors, such as age, may have an influence on the effect of intervention by making a patient more or less receptive to the therapy. Therefore it is important for the review to take into account these factors.

6.2 Design of study

A systematic review based on a meta-analysis of individual patient data from randomised controlled trials of community occupational therapy in stroke patients discharged from hospital.

6.2.1 Measures of interest

The primary outcome measure is the EADL scale at the end of intervention. This was chosen because EADL are the aim focus of community intervention. It is thought that self care ADL will have been attended to during the patients hospital stay in the first few weeks after stroke. Other secondary outcome measures included the leisure

questionnaire, general health questionnaire, activities of daily living, both at the end of intervention and end of trial; and the extended activities of daily living scale, and death, at the end of trial.

6.3 *Study selection and search strategy*

For a trial to be included in the meta-analysis it had to fulfil the requirements of randomisation, at least single blinding, and either have follow-up assessments completed using a postal questionnaire, or using a blinded independent assessor. The intervention had to be based in the community and not in hospital.

A comprehensive search strategy was performed using search engines using the keywords 'occupational therapy', 'stroke', community rehabilitation', rehabilitation', 'activities of daily living', and 'leisure therapy'. These databases included the Cochrane Library (Issue 4, 2003), MEDLINE (1966-Nov 2003), EMBASE (1980-Nov 2003), CINAHL (1982-Nov 2003), PsycINFO (1967-Nov 2003), AMED (1985-Nov 2003), Wilson Social Sciences Abstracts (1984-Nov 2003), Science Citation Index and Social Sciences Citation Index (1981-Nov 2003). The Cochrane Stroke Group Trials Register (last searched November 2003) was also searched. Reference lists were searched from publications and thorough contact with other known researchers. Abstracts from national and international occupational therapy conferences were also hand searched. No restrictions on language were made. The library at the College of Occupational Therapy was also searched for relevant theses and dissertations.

Nine trials were identified which fulfilled the inclusion criteria (Turton *et al.* 1990; Jongbloed *et al.* 1991; Corr *et al.* 1995; Drummond *et al.* 1995; Walker *et al.* 1996; Logan *et al.* 1997; Walker *et al.* 1999; Gilbertson *et al.* 2000; Parker *et al.* 2001) (Figure 6.1). Disparate findings were seen between the published trial results, where even though all of the trials have shown beneficial effects of community occupational therapy, only half of the trials found results which reached statistical significance at the 5% level. The largest trial to date (TOTAL) found that occupational therapy provided little benefit in stroke patients (Table 6.1).

Trial, year of publication	Number of patients	Primary outcome scale	Result of primary outcome
Turton, 1990	22	Peg test	Positive
Walker, 1996	30	NSDA	Positive
Jongbloed, 1991	40	KAI	Neutral
Drummond, 1996	65	NLQ	Positive
Corr, 1995	110	Barthel, NEADL	Neutral
Logan, 1997	111	NEADL	Positive
Gilbertson, 1999	138	NEADL	Neutral
Walker, 1999	185	NEADL	Positive
TOTAL, 2001	466	GHQ, NEADL, NLQ	Neutral

Ordered by size of trial (ascending). KAI is the Katz Adjustment Index, NLQ is the Nottingham Leisure Questionnaire, NSDA is the Nottingham Stroke Dressing Assessment, GHQ is the General Health Questionnaire, and NEADL is the Nottingham Extended Activities of Daily Living

Table 6.1 Published results for the effect of community occupational therapy on stroke patients

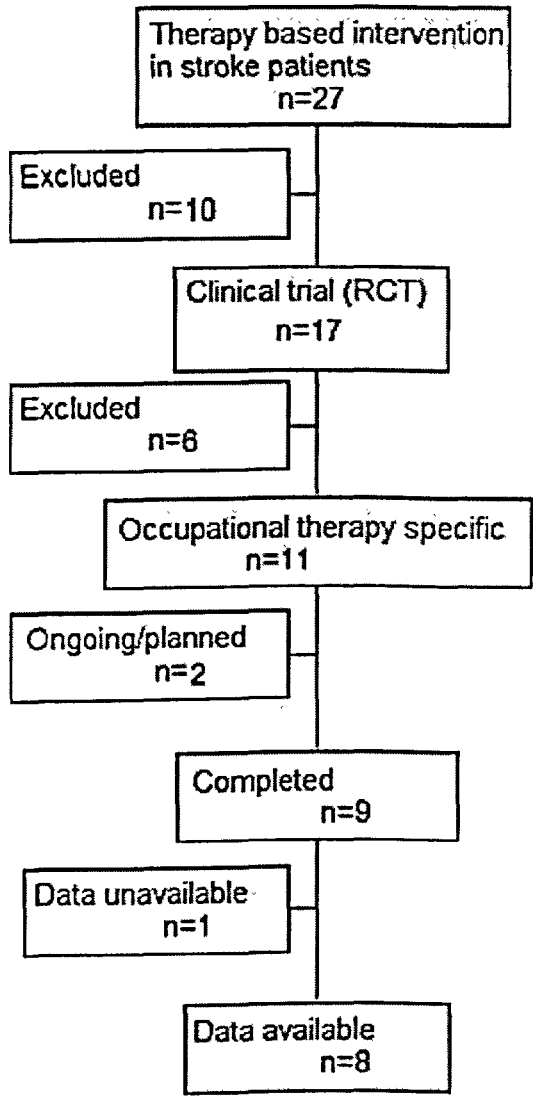

Figure 6.1 Flow chart for trial identification and selection

These disparate findings need to be statistically combined to identify firstly, whether community occupational therapy improves functioning, and if so which components of occupational therapy are beneficial, and in which patients they provide the most benefit.

6.4 Data collection and management

The trialists were contacted and asked if they would share their data with the collaboration. Data received from the trialists were either sent

via email or were collected by the collaboration in the form of questionnaires. All of the trialists contacted were willing to share the data; however, one data set had been discarded a few years previously (Turton *et al.* 1990).

The data from the questionnaires were entered into SAS version 8.02 (SAS Institute Inc) and double-checked to ensure minimal errors. Data from computer disks or emails were formatted into a SAS 8.02 format. The analyses used in the publications were repeated to ensure consistency of results. Where inconsistencies were found the trialists were contacted and the issues were resolved. The discrepancies were mostly attributed to the typographical errors in the publications.

Before the data were merged into a single data set, re-coding of variables, including outcome measures, was performed to achieve consistency across trials. In the case of the Nottingham Extended Activities of Daily Living (EADL) scale, the assessment may be based on the 66-point version where a scoring of (0,1,2,3) is used or on the 22-point version where a scoring of (0,0,1,1) is used. Since the (0,1,2,3) scale can be converted easily into the (0,0,1,1) scale, this latter format was used across all trials.

Similarly, with the General Health Questionnaire the scoring may be performed in a similar manner. A common scoring of (0,0,1,1) was used across the trials for comparability. However, another complication of this questionnaire is that various versions exist. Two versions of the general health questionnaire were identified in this

meta-analysis; the 12-point scale and the 28-point scale. *A priori* cut offs were used to make these two scales comparable. For the 12-point scale a cut of >2 indicates a case, which is comparable to a cut off of >4 on the 28-point scale (Wade 1992).

For the Nottingham Leisure Questionnaire, a 3-point (0,1,2) or a 5-point (0,1,2,3,4) scoring method may be used. The five-point scoring method was collapsed to the three-point scoring method by combining the 0 and 1 scores, and the scores 3 and 4.

The self care section of the Rivermead Activities of Daily Living scale was used in two of the trials to assess functional performance (Drummond *et al.* 1995; Walker *et al.* 1996). The Barthel Index was used to assess functional performance in the majority of the other trials and therefore a comparable functional performance measure needed to be specified to ensure that these two measures represented similar scores. An *a priori* cut off of ≤10 was used for the Rivermead scale which has been shown empirically to be comparable to the *a priori* cut off of ≤16 for the Barthel Index (≤80 on the 100 point-scale indicates dependency).

6.5 *Assessment of publication bias and quality*

Begg's funnel plot was used to assess visually if there was evidence of publication bias within the meta-analysis. The plot was symmetrical in appearance and the point estimates were all located within the pseudo 95% confidence interval lines; hence the test concluded there was no substantial evidence of publication bias (Figure 6.2).

However, due to the subjective nature of interpreting this figure, formal tests for publication bias were also performed.

The Begg and Mazumdar rank correlation test (p=0.624), Egger's asymmetry test (p=0.728) and the trim and fill method did not find evidence of publication bias.

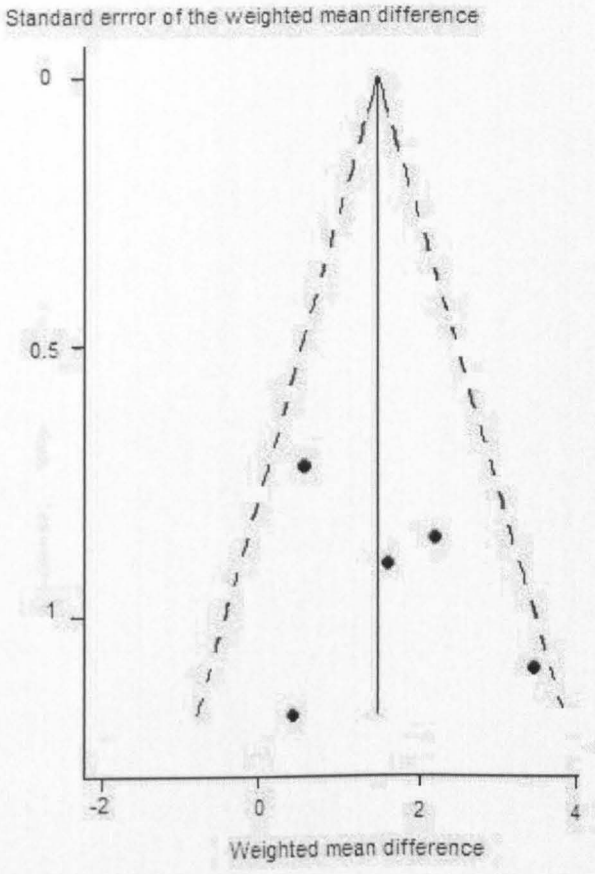

Figure 6.2 Begg's funnel plot for EADL at the end of the intervention phase

Seven of the eight trials with individual patient data were deemed to be of high methodological quality, with evidence of blinded randomisation procedures, concealment of allocation, and masked outcome assessments (Table 6.2) (The Cochrane Collaboration 2003). The trialists involved in the Jonglboed trial were asked to provide more information about the design of the trial; unfortunately

they could not clarify the method of randomisation used in the trial. The ninth trial which had discarded the raw data, also had a lower level of quality due to the randomisation process being based on date of birth of the patients, and doubts over whether the outcome assessor was blinded (Table 6.2).

Trial, year of publication	Type of study	Blinding	Method of randomisation	Blinded outcome assessment/ assessor
Turton, 1990	RCT	B	Quasi, block randomisation	Unclear
Jongbloed, 1991	RCT	B	Unclear	Yes
Corr, 1995	RCT	A	Opaque, sealed envelopes	Yes
Drummond, 1995	RCT	A	Numbered opaque, sealed envelopes	Yes
Walker, 1996	Cross-over	A	Numbered opaque, sealed envelopes	Yes
Logan, 1997	RCT	A	Numbered opaque, sealed envelopes	Yes
Walker, 1999	RCT	A	Numbered opaque, sealed envelopes	Yes
Gilbertson, 2000	RCT	A	Numbered opaque, sealed envelopes	Yes
Parker, 2001	RCT	A	Central randomisation by telephone	Yes

A= Low risk of bias, B= Moderate risk of bias, C= High risk of bias

Table 6.2 Assessment of quality for the nine identified trials

6.6 *Trial level demographics*

Trials were assessed to determine whether patients were similar between the trials (Table 6.3). High percentages of dependent stroke patients were found in the Drummond and Walker trials (95.4% and 73.3%, respectively). Both of these trials utilised the Rivermead self care section scale as the assessment of dependency. The average ages were relatively similar across the trials. The sex of the patients

in each trial were mostly balanced except for slight imbalances in two of the trials (Jongbloed *et al.* 1991; Corr *et al.* 1995).

Trial, year of publication	Number of patients	Gender, male (%)	Dependent patients (%)	Age, mean (SD)
Jongbloed, 1991	40	27 (71.1)	5 (13.5)	68.8 (10.6)
Corr, 1995	110	41 (37.3)	51 (52.6)	75.5 (9.0)
Drummond, 1995	65	37 (56.9)	62 (95.4)	66.0 (11.2)
Walker, 1996	30	16 (53.3)	22 (73.3)	68.1 (9.4)
Logan, 1997	111	56 (50.5)	-	72.4 (11.0)
Walker, 1999	185	94 (50.8)	51 (27.6)	68.1 (9.4)
Gilbertson, 2000	138	62 (44.9)	36 (26.1)	69.0 (12.0)
TOTAL, 2001	466	269 (57.7)	105 (22.6)	71.0 (10.3)

Table 6.3 Patient demographics at baseline by trial

6.7 Outcome assessments

The Extended Activities of Daily Living (EADL) scale was used at the end of intervention in five of the eight trials (Drummond *et al.* 1995; Logan *et al.* 1997; Walker *et al.* 1999; Gilbertson *et al.* 2000; Parker *et al.* 2001). In addition to the above trials, the trial by Corr and Bayer (Corr *et al.* 1995) used the assessment at the end of trial.

The Nottingham Leisure Questionnaire was used in four trials (Drummond *et al.* 1995; Walker *et al.* 1999; Gilbertson *et al.* 2000; Parker *et al.* 2001) however, one of the trials only used the assessment at the end of the intervention phase (Walker *et al.* 1999) and another trial only used it at the end of the trial phase (Gilbertson *et al.* 2000).

End of intervention and end of trial dependency scores were assessed using either the Barthel Index (Corr *et al.* 1995; Logan *et al.* 1997; Walker *et al.* 1999; Gilbertson *et al.* 2000; Parker *et al.* 2001) or the Rivermead self care section scale (Walker *et al.* 1996). The trial by Jongbloed and colleagues did not assess dependency at either follow-up time (Jongbloed *et al.* 1991).

Data from the General Health Questionnaire was available for patients from three trials (Logan *et al.* 1997; Walker *et al.* 1999; Parker *et al.* 2001) and for carers from two trials (Walker *et al.* 1999; Parker *et al.* 2001). All of the trials had data on death at the end of trial.

6.7.1 Mode of assessments

Outcome assessments were either postal or completed with the aid of an assessor which visited the patients in their home. Three trials used a postal end of treatment follow-up procedure and the other trials used an assessor. At the end of the trial assessments, four used a postal follow-up questionnaire and the remaining four used an assessor (Table 6.4). The trial by Gilbertson and colleagues used an independent assessor at the end of intervention and a postal questionnaire at the end of trial.

Trial, year of publication	Method of follow-up End of intervention	Method of follow-up End of trial
Jongbloed, 1991	Independent assessor	Independent assessor
Corr, 1995	Postal questionnaire	Postal questionnaire
Drummond, 1995	Independent assessor	Independent assessor
Walker, 1996	Independent assessor	Independent assessor
Logan, 1997	Postal questionnaire	Postal questionnaire
Walker, 1999	Independent assessor	Independent assessor
Gilbertson, 2000	Independent assessor	Postal questionnaire
TOTAL, 2001	Postal questionnaire	Postal questionnaire

Table 6.4 Method of follow-up used at the end of intervention and end of trial

6.7.2 Type of intervention

As mentioned earlier in the introduction, occupational therapy is a package of treatment which can be aimed at promoting independence through leisure pursuits and activity, or through activities of daily living. One trial solely used therapy aimed leisure pursuits (Jongbloed *et al.* 1991), 4 trials used therapies aimed at activities of daily living, one trial assess the usefulness of a component of activities of daily living; dressing practice, and the remaining two trials assessed activities of daily living and leisure therapy in a parallel design (Table 6.5).

6.7.3 Timing of assessments

All of the trials used different end of intervention and end of trial assessment times. Therefore it was decided that the first assessment after the final session of the intervention would be known as the 'end

of intervention' phase. The last follow-up assessment would be known as the 'end of trial' phase.

The end of intervention assessment ranged in times from as short as 5 weeks (Jongbloed *et al.* 1991; Gilbertson *et al.* 2000) to 6 months (Walker *et al.* 1999; Parker *et al.* 2001), with the most frequently used follow-up period was 3 months (Table 6.5). The end of trial assessments ranged from 18 weeks (Jongbloed *et al.* 1991) to one year (Corr *et al.* 1995; Walker *et al.* 1999; Parker *et al.* 2001; Walker *et al.* 2001), with the most commonly used time point of six months.

Trial, year of publication	Number of patients	Intervention	Intervention length	End of intervention assessment (months)	End of trial assessment (months)
Jongbloed, 1991	40	Leisure	5 sessions	1.25	4.5
Corr, 1995	110	ADL	Up to 6 months	-	12
Drummond, 1995	65	ADL or Leisure	Min 10 sessions	3	6
Walker, 1996	30	ADL	12 weeks	3	6
Logan, 1997	111	ADL	6 weeks	3	6
Walker, 1999	185	ADL	Up to 5 months	6	12
Gilbertson, 2000	138	ADL	6 weeks	2	6
TOTAL, 2001	466	ADL or Leisure	Min 10 sessions	6	12

Table 6.5 Timings for intervention and assessments by trial

6.8 *Analysis of the merged data set using summary data methods*

The data from the eight individual trials were merged into a single data set. Data from the ninth trial were not available for inclusion. The combined data set contained patient information on 1,143 stroke patients. Table 6.6 shows the demographics of the stroke patients by intervention group (dressing practice, conventional ADL or leisure therapy) and the control group (usual or no intervention).

The overall demographic characteristics were typical of an ageing stroke population, mean age 71 years (range 28-96), 53% male and 33% were dependent prior to stroke. The data set was balanced between the intervention groups for age, gender, baseline dependency, whether the patient lived alone (Table 6.6), and for side of stroke (left 48.0%, right 50.7%, bilateral 1.3%).

Variable	ADL based therapy	Leisure based therapy	Control / Usual care	Overall results
N	481	174	488	1143
Age yrs, mean (SD)	71.1 (10.4)	69.5 (11.5)	72.4 (10.2)	71.4 (10.5)
Gender male, n (%)	253 (53)	102 (59)	247 (51)	602 (53)
Dependent, n (%)	140 (33)	60 (34)	132 (31)	332 (33)
Lives alone, n (%)	182 (41)	44 (29)	189 (42)	415 (40)

ADL Activities of Daily Living. SD Standard Deviation.

Table 6.6 Demographics for the combined data set by intervention group

6.8.1 Data analysis for EADL at the end of the intervention phase

The mean score on the EADL scale (22 point scale) for each intervention group within each trial was used as the primary outcome for the summary measures analysis. The interventions were coded by the type of therapy assigned (control/usual therapy, and a global occupational therapy group which included leisure based therapy and activities of daily living based therapy) in each model. Leisure based therapy and activities of daily living based therapy were combined for the main analyses since the primary hypothesis under investigation related to the efficacy of all types of occupational therapy. Comparisons were made to compare efficacy between the intervention groups.

The difference between the means and pooled variances for each trial were calculated as detailed in section 2.3.2.1. All of the trials used the Nottingham EADL scale, therefore the data were analysed using the weighted mean difference approach with the weights proportional to the reciprocal of the pooled variance for each study (see section 2.4).

Trial, year	Active group		Control group		Difference in means	Pooled variance
	N_1	$Mean_1$ (SD_1)	N_2	$Mean_2$ (SD_2)		
Drummond 1995	41	8.6 (4.8)	21	8.2 (4.2)	0.42	1.539
Logan 1997	43	8.7 (5.4)	43	5.1 (4.8)	3.53	1.207
Walker 1999	84	13.5 (5.0)	79	11.2 (5.9)	2.26	0.725
Gilbertson 2000	64	9.5 (5.1)	69	7.9 (5.4)	1.64	0.830
TOTAL 2001	233	10.8 (6.4)	119	10.3 (6.4)	0.53	0.521

Table 6.7 Summary measures for the five trials that measured the EADL scale at the end of the intervention phase

6.8.2 Results for the EADL at the end of the intervention phase

In a conventional fixed effect model, as described in section 2.4.1; patients receiving community occupational therapy had, on average, an EADL score 1.55 points higher (on a 22-point scale) at the end of intervention as compared to those randomised to receive usual care (Weighted mean difference, WMD 1.55, 95% confidence intervals, CI 0.75, 2.34).

6.8.3 Results for secondary outcomes

Using a conventional fixed effect model as described in section 2.4.1, patients receiving community occupational therapy had an EADL score which was higher by 0.94 points (on a 22 point scale) at the end-of-trial, as compared with those randomised to receive usual care (95% CI 0.10, 1.78).

Subjects receiving community occupational therapy also had a higher Nottingham Leisure Questionnaire score, by 1.21 points (on a 37 item/74 point scale) (95% CI -0.17, 2.41) at end of intervention and 1.45 points at end-of-trial (95% CI 0.05, 2.85), as compared with usual care.

A significant odds reduction of 29% in disability (assessed using the Barthel Index or Rivermead Scale) was present at end of intervention (odds ratio, OR 0.71, 95% CI 0.52, 0.99) (Table 6.8). No effects were detected on common mental disorders (GHQ) in either patients or carers at end of intervention or end of trial, or on death by end of trial (Table 6.8).

Outcome	Number of trials	Odds Ratio (95% CI)
End of intervention phase		
Activities of Daily Living	5	0.71 (0.52, 0.99)
Patient GHQ	3	0.76 (0.54, 1.07)
Carer GHQ	3	0.74 (0.49, 1.13)
End of trial phase		
Activities of Daily Living	5	0.75 (0.55, 1.03)
Patient GHQ	2	1.09 (0.69, 1.61)
Carer GHQ	2	1.11 (0.70, 1.76)
Death	8	1.07 (0.69, 1.65)

Odds ratio is the odds of the particular outcome on occupational therapy relative to control/usual care

Table 6.8 Relationship between outcome measures and community occupational therapy at the end of intervention and end of trial, based on summary data

6.9 Assessment of heterogeneity using summary data methods

The above analyses have assumed that a common underlying estimate for the intervention effect exists for all of the trials. An assessment of heterogeneity needs to be performed to evaluate whether this assumption is valid. Visual and statistical assessments of heterogeneity were investigated for the primary outcome measure, EADL at the end of the intervention phase.

6.9.1 Graphical assessment of heterogeneity

Figure 6.3 shows the forest plot for the EADL at the end of the intervention phase. The forest plot indicates that there are some differences in the estimates for the intervention effect between the trials and hence formal tests for heterogeneity were performed.

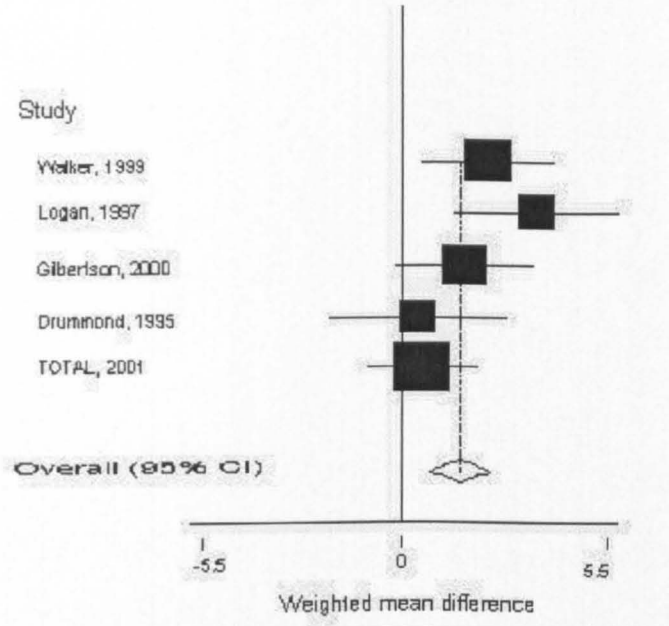

Figure 6.3 Forest plot for EADL at the end of the intervention phase

6.9.2 Statistical assessment of heterogeneity using classical and Bayesian methods

To formally assess whether there was evidence of heterogeneity between the trials estimates Cochran's homogeneity *Q* test, Higgins and Thompson's I^2 were performed. Also, other analyses were also used to quantify the heterogeneity between the trial estimates using weighted and un-weighted MM, ML, REML, EB and FB (Table 6.9).

Cochran's homogeneity test did not find evidence of heterogeneity between the trials estimates (p=0.145). However, I^2 indicated that 41.5% of the total variability within the meta-analysis may be attributed to heterogeneity between the trial estimates.

From Table 6.9, the estimates for heterogeneity from the un-weighted and weighted MM, ML, REML and EB were 0.703, 0.607, 0.269, 0.559 and 0.624 respectively. The magnitude of the heterogeneity from these models seemed to have little impact on the estimates and standard

errors for the intervention effect. The FB model found the most heterogeneity, 1.599, and hence the standard error and confidence intervals were slightly wider from this model as compared to the others, although the intervention effect remained highly significant.

Heterogeneity Test	χ^2 value	P value	I^2
Q test	6.830	0.145	
I^2			0.410
Estimation analyses	Intervention (SE)	95% CI	$\hat{\tau}^2$
MM, un-weighted	1.631 (0.564)	0.526, 2.736	0.703
MM, weighted	1.616 (0.544)	0.549, 2.682	0.607
ML	1.589 (0.474)	0.660, 2.518	0.269
REML	1.613 (0.535)	0.564, 2.661	0.559
EB	1.617 (0.550)	0.539, 2.696	0.624
FB	1.674 (0.803)	0.126, 3.221‡	1.599

Intervention relates to the estimate of EADL on occupational therapy relative to control, SE Standard Error, ‡ 95% credibility intervals

Table 6.9 Heterogeneity test and analysis results for EADL at the end of the intervention phase based on summary data

6.10 *Exploring heterogeneity*

There appears to be heterogeneity between the trials, this may be related to differences within the trials either at patients' level or trial level. Subgroup analyses and meta-regression methods were performed to assess whether trial level covariates such as method of follow up, or type of intervention or patient predictors such as age, gender or baseline dependency, may influence the results from each trial and hence explain some of this residual heterogeneity.

6.10.1 Subgroup analyses

Subgroup analyses were performed to assess the relationship between EADL at the end of intervention and the mode of follow-up used to record the scores, and the type of intervention received.

6.10.1.1 Mode of follow-up

In the five trials which assessed EADL at the end of the intervention phase, two used postal questionnaires, the remaining three used an independent assessor which visited the patients' home and aided with their completion of the assessment. Using an independent assessor yielded an overall estimate of 1.63 (95% CI 0.55, 2.71), whereas using the postal questionnaire yielded an overall estimate of 1.44 (95% CI 0.25, 2.63) (Figure 6.4). The p value for the change in heterogeneity was 0.823. Therefore, the method of follow-up does not appear to influence the EADL scores at the end of the intervention phase.

6.10.1.2 Type of intervention

Two main types of intervention exist for occupational therapy. The first concentrates on improving activities of daily living, and involves promoting independence in areas such as mobility. The second type of therapy concentrates on leisure activities and encourages patients to become more involved and independent in participation. The types of interventions assessed in the trials will vary from trial to trial since the intervention is therapy based and is tailor made to each patient. Therefore, the frequency and intensity of intervention will also vary from patient to patient.

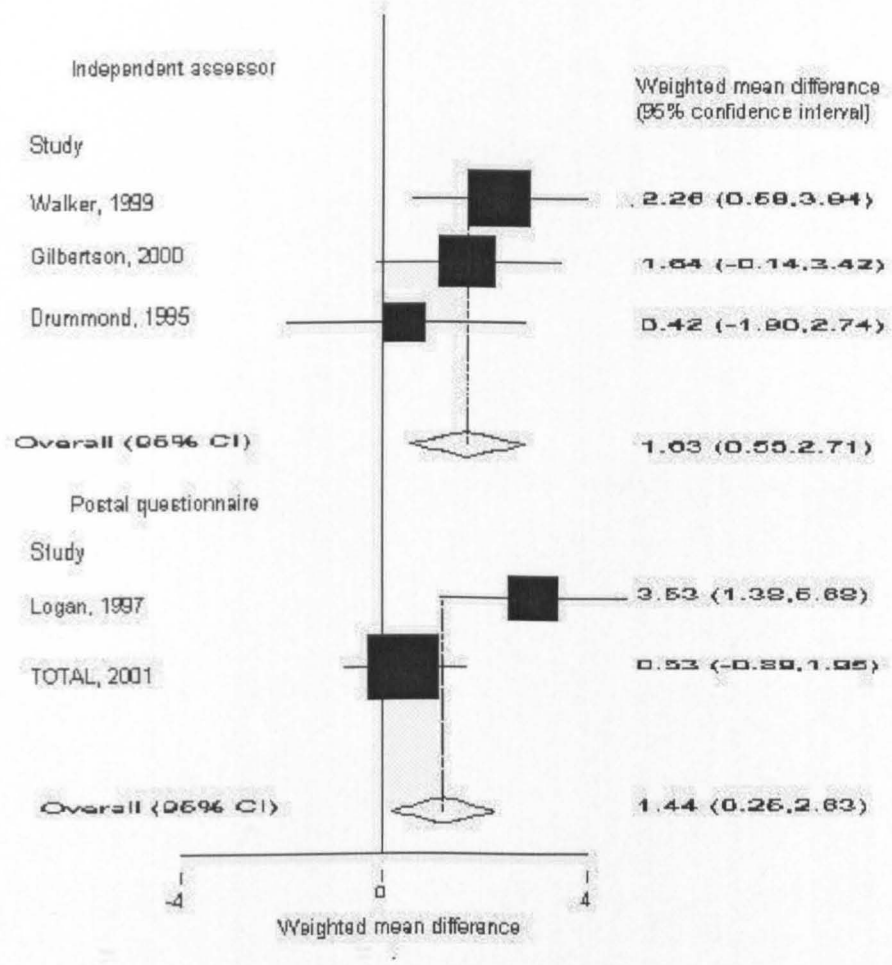

Figure 6.4 Subgroup analysis plot for EADL at the end of the intervention phase by the mode of follow up

Two types of active intervention were assessed in the trials, activities of daily living (ADL) therapy and leisure therapy. The data for the trials were split by the type of intervention received (ADL, leisure, control/usual therapy). Overall, 3 trials had used ADL based therapy, and 1 used leisure based therapy and two trials assessed both in a parallel design.

Using ADL based therapy was associated with a significant increase in the EADL scale of 1.56 points (95% CI 0.72, 2.39), whereas using leisure based therapy resulted in a non-significant increase of 0.61 points on the EADL scale (95% CI -0.81, 2.03) (Figure 6.5). Assuming 7

independent trials are evaluated, the p value for the change in heterogeneity was 0.262. Although this value is not statistically significant at the 5% level, the type of intervention used may to have some influence on the estimates for the intervention effect, where ADL based therapy may be more effective on impacting on extended activities of daily living (Figure 6.5).

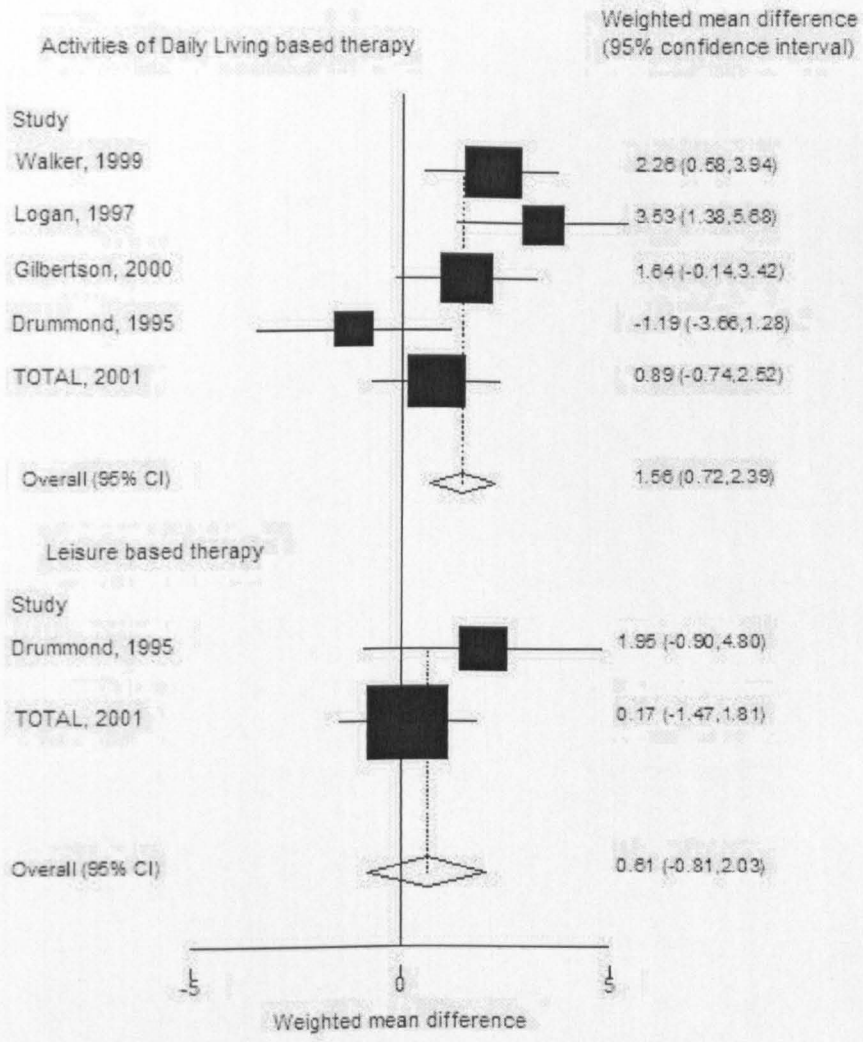

Figure 6.5 Subgroup analysis plot for EADL at the end of the intervention phase by the type of intervention used

6.10.2 Meta-regression analyses

Meta-regression methods were used to assess the relationship between the primary outcome and several predictors. These included differences in the average age of the patients between the two groups, the difference in the percentage of male patients between the two groups, and the difference in the percentage of dependent patients between the two groups. These covariates were created using a summary measure for each intervention group, in each of the trials.

All five trials with EADL at the end of intervention had age and gender data, however one of these trials did not record data on baseline dependency (Logan *et al.* 1997)

Previously, evidence of heterogeneity was found between the trial estimates, therefore a series of random effect models were used to explore the relationships between EADL and the covariates (see Section 3.4). The methods used to estimate heterogeneity were the weighted and approximate MM, ML, and REML, EB and FB. Assessments of convergence were adequate when a burn-in of 10,000 iterations and a sample chain of 10,000 iterations were used.

For each of the covariates, similar estimates for the effects were seen from all of the models. Slight differences between the magnitudes of the effects from the models were related to the differences in the estimates for heterogeneity. In each of the covariate adjusted models, none of the models found there was a relationship between any of the predictors and the estimates of the relative mean difference for the treatment effect (Table 6.10).

Method	Intervention[§] (SE)	95% CI	Age[Ω] (SE)	95% CI	$\hat{\tau}^2$
MM, approx	1.412 (1.265)	-1.068, 3.892	0.123 (0.553)	-0.960, 1.206	1.607
MM, weighted	1.361 (1.064)	-0.725, 3.447	0.146 (0.476)	-0.788, 1.079	0.927
ML	1.207 (0.754)	-0.271, 2.685	0.214 (0.363)	-0.498, 0.926	0.118
REML	1.364 (1.072)	-0.737, 3.464	0.145 (0.479)	-0.794, 1.083	0.950
EB	1.379 (1.126)	-0.828, 3.587	0.137 (0.500)	-0.842, 1.117	1.124
FB	1.532 (1.973)	-2.244, 5.336‡	0.073 (0.801)	-1.491, 1.617‡	2.003
Method	Intervention[§] (SE)	95% CI	Gender[Ω] (SE)	95% CI	$\hat{\tau}^2$
MM, approx	1.745 (0.586)	0.596, 2.894	0.086 (0.076)	-0.064, 0.235	0.780
MM, weighted	1.745 (0.527)	0.711, 2.779	0.085 (0.069)	-0.050, 0.220	0.461
ML	1.733 (0.427)	0.896, 2.570	0.083 (0.057)	-0.026, 0.194	0
REML	1.745 (0.528)	0.710, 2.780	0.085 (0.069)	-0.050, 0.220	0.465
EB	1.745 (0.527)	0.713, 2.777	0.085 (0.068)	-0.050, 0.219	0.457
FB	1.735 (0.834)	0.082, 3.377‡	0.088 (0.105)	-0.123, 0.302‡	1.570
Method	Intervention[§] (SE)	95% CI	Depend[Ω] (SE)	95% CI	$\hat{\tau}^2$
MM, approx	0.818 (0.734)	-0.621, 2.258	0.150 (0.191)	-0.224, 0.524	0.221
MM, weighted	0.814 (0.686)	-0.531, 2.159	0.151 (0.179)	-0.201, 0.502	0.099
ML	0.808 (0.643)	-0.452, 2.069	0.152 (0.169)	-0.180, 0.483	0
REML	0.817 (0.709)	-0.573, 2.206	0.150 (0.185)	-0.212, 0.512	0.155
EB	0.814 (0.682)	-0.522, 2.150	0.151 (0.178)	-0.199, 0.500	0.089
FB	0.805 (1.612)	-1.974, 3.504‡	0.149 (0.402)	-0.549, 0.849‡	1.534

§ EADL estimate on OT relative to control when the difference in the percentage of the covariate between the OT and control groups equals zero. SE Standard error, Ω Multiplicative increase in EADL estimate when the covariate is increased by 1%, Gender relates to male patients, Depend relates to dependent patients. ‡ credibility intervals.

Table 6.10 EADL at the end of the intervention phase using summary data meta-regression methods

6.11 *Individual patient data methods*

In a classical fixed effect IPD model where a common variance is assumed across the trials, patients randomised to occupational therapy were found to have significantly higher EADL scores at the end of intervention phase by 1.4 points as compared to those randomised to usual care (Table 6.11).

	Intervention (SE)	95% CI	$\hat{\tau}^2$
Fixed effect model			
Classical approach	1.437 (0.421)	0.611, 2.263	-
Random effect model			
Classical approach	1.576 (0.582)	0.435, 2.717	0.522
Bayesian approach	1.567 (0.584)	0.486, 2.731‡	0.826

Intervention is the EADL estimate for occupational therapy relative to control, SE standard error, ‡ credibility intervals.

Table 6.11 EADL at the end of intervention phase using individual patient data methodologies

6.12 *Assessments of heterogeneity using individual patient data methods*

An interaction term between trial and intervention was included in the above model to test whether there was heterogeneity between the trials; allowing for this test having low power when a small number of trials are combined, there may be some difference between the trial estimates in this meta-analysis ($p=0.192$). Therefore, heterogeneity was estimated using both classical and Bayesian methodologies.

Classical and Bayesian random effect IPD models were used where a common variance was assumed across the trials. Additionally, assessments of convergence for the FB model were made using

methods described in section 2.11. The assessments were satisfactory when a burn-in of 10,000 iterations and a sample chain of 40,000 iterations were used.

Similar results were seen from the classical and Bayesian IPD random effect models for the intervention effect which implied that occupational therapy was significantly associated with a higher EADL score of 1.6 points at the end of intervention as compared to usual care (Table 6.11). Slightly more heterogeneity was estimated from using the Bayesian model; however, this did not appear to impact greatly on the standard error for the intervention effect when compared to the classical model.

6.13 Exploring heterogeneity using individual patient data methods

Both classical and Bayesian IPD methods found that quantifiable estimates of heterogeneity existed between the trials. Therefore trial level covariates, such as method of follow up, and patient level covariates, such as age, gender, baseline dependency and type of intervention, were included in the random effect models to assess their influence on the EADL score at the end of intervention.

6.13.1 Trial level covariates using individual patient data

Method of follow-up was modelled using a classical random effect IPD meta-regression model as described in Section 3.5.1.3 to assess its impact on EADL at the end of intervention. No differences in the

pooled intervention effects for the EADL scores were seen between using a postal questionnaire or an independent assessor (p =0.993)

6.13.2 Patient level covariates using individual patient data

Classical and Bayesian models were performed to assess the impact of age, gender, baseline dependency and type of intervention on EADL at the end of intervention using random effect IPD models. For the FB models, assessments of convergence were performed as described in section 2.11 and found to be adequate using a burn-in of 10,000 iterations and a sample chain of 40,000 iterations.

Initially, an interaction term between the covariate and intervention terms was included in the models to ascertain whether the intervention effect varied across the range of the covariates (except for the type of intervention covariate). All of the models found no evidence of an interaction and therefore the interaction was subsequently dropped from the classical and Bayesian models (age interaction p =0.433, gender interaction p =0.765, and dependency interaction p =0.727).

In the covariate adjusted models presented in Table 6.12, the results for the intervention effects and covariate effects are very similar between the Bayesian and classical IPD models.

In the age adjusted models, patients randomised to occupational therapy had higher EADL scores of 1.4 points at the end of intervention phase as compared to patients receiving usual care.

However, the ages of the patients did not appear to be a significant predictor of EADL at the end of the intervention phase (Table 6.12).

In the gender adjusted models, patients randomised to occupational therapy had significantly higher EADL score of 1.6 points at the end of intervention phase as compared to patients that received usual care. Additionally, gender was found to be a significant predictor of EADL which suggested that male stroke patients scored approximately 1.3 points higher on the EADL as compared to females (Table 6.12).

In the models adjusted for baseline dependency, significant differences between the groups were seen which suggested that patients randomised to occupational therapy scored 1.4 points higher than patients randomised to usual care. Also, baseline dependency was found to be an important predictor of EADL; where dependent patients were found to score approximately 6 points lower on the EADL scale as compared to independent stroke patients.

Significantly higher EADL scores at the end of intervention were found in patients which received ADL based occupational therapy as compared to usual care (estimate 1.602, 95% CI 0.718, 2.487). However, no significant differences were seen in the EADL scores between the patients which received leisure based occupational therapy and those which received usual care (estimate 0.939, 95% CI -0.317, 2.195).

IPD Models	Estimate (SE)	95% CI	Estimate (SE)	95% CI	$\hat{\tau}^2$
	Intervention§		Age$^{\Omega}$ (years)		
REML	1.422 (0.543)	0.358, 2.486	-0.087 (0.019)	-0.125, -0.049	0.161
FB	1.447 (0.613)	0.347, 2.686‡	-0.026 (0.020)	-0.064, 0.012‡	0.866
	Intervention§		Gender$^{\Psi}$ (male)		
REML	1.562 (0.609)	0.368, 2.756	1.257 (0.411)	0.450, 2.065	0.337
FB	1.476 (0.655)	0.320, 2.828‡	1.248 (0.411)	0.437, 2.255‡	1.248
	Intervention§		Dependency$^{\Psi}$ (dependent)		
REML	1.454 (0.518)	0.439, 2.469	-6.265 (0.482)	-7.212, -5.319	0.111
FB	1.369 (0.675)	0.128, 2.598‡	-6.144 (0.484)	-5.195, -7.087‡	1.096

§ Intervention is the EADL estimate on occupational therapy relative to control, Ω Estimate for age is the multiplicative increase in the estimate of EADL for each year increase in age, irrespective of intervention assignment, Ψ the estimates for gender or dependency is the increase in the estimate for EADL for males or dependent patients as compared to female or independent patients, respectively, irrespective of treatment. SE Standard error, ‡ 95% Credibility Intervals.

Table 6.12 EADL at the end of the intervention phase using individual patient data adjusted for patient level covariates

6.14 *Discussion of findings*

The discussion of the findings from this systematic review is presented in six sections which relate to a summary of the overall findings, limitations of the data used in the review, comparison of the statistical methods, discussion of the results, practical implications of the results and suggestions for future research.

6.14.1 Summary of the overall findings

Nine randomised controlled trials of community occupational therapy in stroke patients were identified from a comprehensive search strategy which was performed up to November 2003. Individual patient data were available from eight of these trials; the data from the ninth trial had been discarded previously. The remaining eight trials recruited 1,143 stroke patients, and the patients from these trials were randomised to either occupational therapy or usual care. Two types of occupational therapy were identified from the trials; activities of daily living and leisure therapy. Baseline characteristics of the patients appeared to be balanced between the intervention groups from the individual trials.

At the end of intervention, patients randomised to occupational therapy had significantly higher scores for EADL and ADL measures, and a non-significant higher score for NLQ. Additionally, intervention effects for EADL and NLQ were significantly maintained to the end of trial. No

effects were seen between the intervention groups for death and minor psychiatric disorders as measured in the patients or their carers.

6.14.2 Limitations of the data used

6.14.2.1 Number of studies included in the meta-analysis

The search strategy used in this systematic review identified nine randomised controlled trials of occupational therapy given to stroke patients in the community setting. However, only five of these trials had recorded EADL at the end of the intervention phase and so were used in the primary analyses. Although the primary outcomes from each of the individual trials could have been combined using standardised mean difference methodology to allow for data from all of the nine trials to be used, it was decided *a priori* that the results from such an analysis would be meaningless to clinicians. An attempt was also made to combined the outcomes of personal and extended ADL however; this would have only contributed data from the same five trials as well.

6.14.2.2 Timings of the outcomes

It was decided *a priori* to use the end of intervention and end of trial phases as the two assessment times. These endpoints were decided to allow for the different lengths of intervention treatments across the trials; and were thought preferable to using a particular timing, such as six months, since there was a lack of a consistent timing across the trials. Also, it was thought that defining the *a priori* assessment timings

would be less biased and meant that the reviewers did not decide the timings based on the results presented in the publications. Also, the fact that the effects seen for the outcomes appeared to be maintained at the end of intervention and end of trial suggest that the timings used were appropriate.

6.14.2.3 Extent of the shared individual patient data

Although the original trials generally collected considerable amounts of data during the baseline and follow-up assessment periods, it was frequently seen that the information collected was not consistent across the trials. For example, although the trial by Logan and colleagues assessed EADL at the end of intervention, it did not collect baseline dependency data (Logan *et al.* 1997) hence the trials had to be excluded form the analyses when this variable was considered. Also, the trial by Corr and colleagues only assessed EADL at the end of trial, and no assessments were made at the end of the intervention phase for any of the outcomes (Corr *et al.* 1995).

Two scoring systems may be used for the EADL scale, within this meta-analysis a consistent scoring method had to be used to ensure the data from the trials could be combined. Therefore, the (0,1,2,3) scoring system was collapsed into the (0,0,1,1) system. A previous study only found significant intervention effects for occupational therapy when the (0,1,2,3) system was used as compared to the (0,0,1,1) scale (Walker

et al. 1999); this was thought due to the (0,0,1,1) scale being less sensitive to a change over a period of intervention. Therefore, it is suggested that the magnitude of the intervention effect seen from this meta-analysis may be an under-estimate of the true intervention effect that would be seen if the (0,1,2,3) system had been used since the magnitude seen in this meta-analysis would be related to an increase in the EADL scale from either (0) to (2) or (1) to (3) if the (0,1,2,3) scale had been implemented.

In addition to the outcome presented in this chapter, we also wanted to assess the effect of community occupational therapy in stroke patients on reducing the levels of handicap and caregiver strain since a previous study had found significant improvements on these two outcomes (Walker *et al.* 1999); however a lack of data from other trials prevented these outcomes being assessed in this systematic review.

6.14.2.4 Outcome assessments

The primary and secondary outcome measures used in this review were selected *a priori* before the trials were identified and following discussions with the principal collaborators of the group. These scales were thought to be pertinent to the aims that occupational therapy attempts to influence. It was acknowledged that the scales have ceiling and flooring effect, and therefore it was decided that the primary outcome should account for this. Therefore, to overcome the potential

ceiling effects seen with the ADL scales, EADL was chosen as the primary outcome to assess whether occupational therapy improves EADL tasks in stroke patients.

6.14.2.5 Experience of the therapists

Within the majority of the trials, the senior occupational therapist which provided the interventions for the patients were also studying for a PhD. Therefore, it could be argued that these therapists had a specific interest in the results of the trials, hence were more motivated to provide an intervention which was found to be significantly better than the usual care received by the patient randomised to the control groups. Also, these therapists were researcher occupational therapists and perhaps had more experience in targeting the areas that would be picked up on the outcome assessments as compared to the therapists which provided the usual care to the control groups. Another possibility is that the researcher occupational therapists may be more familiar with the literature on targeted rehabilitation being more effective and therefore may have modified their practice.

However, the EADL outcome assessments used within the trials were not biased since either a questionnaire was completed by the patient or an independent assessor was used who was blinded to the intervention allocation. Also, only two of the nine trials identified in this review were deemed to be of a slightly lower quality due to unclear reporting of

randomisation techniques, and since these two trials did not contribute data to the primary outcome analysis, it can be suggested that the results seen from this review are from high quality randomised controlled trials and are likely to be generalisable and unbiased.

6.14.3 Comparison of the statistical methods

6.14.3.1 Assessment of heterogeneity

The impact of heterogeneity on the results was assessed using various statistical models for EADL at the end of intervention. Initially, statistical assessments of heterogeneity were performed which produced results which were contrary to each other; where relatively high levels of heterogeneity appeared to be present in the results between the trials from using I^2 where as Cochran's homogeneity test yielded a non-significant p value. This difference is not surprising since the power associated with Cochran's homogeneity test is reduced when the meta-analysis contains data from a small number of trials (Fleiss 1986; Whitehead *et al.* 1991; Thompson 1994). No evidence of heterogeneity was also seen when assessed in an individual patient data model by including a fixed effect interaction term between intervention and trial. Again, the lack of significance is related to a lack of power from assessing a small number of trials included in this meta-analysis (Brown *et al.* 1994).

6.14.3.2 Impact of quantifying heterogeneity

A range of estimation methods were then used in a random effect model to assess whether there were differences between the results from the summary data models. All of these models estimated quantifiable amounts of heterogeneity between the trials; however, the magnitude of the heterogeneity had little impact on the significant results for the intervention effect, hence all performed adequately in this meta-analysis.

Fixed and random effect individual patient data models were then used to assess the impact of occupational therapy on EADL at the end of intervention. In a fixed effect model, patients randomised to occupational therapy had significantly higher EADL scores than those receiving usual care, this significant effect was maintained when a random intervention by trial effect was included in the model.

6.14.3.3 Impact of exploring heterogeneity

Various estimation methods were then used in meta-regression models to assess whether there were any intervention effect modifiers and whether there were any differences between the results from the estimation methods. All of the models yielded similar results for the intervention and covariate effects which implied that age, sex and baseline dependency did not modify the intervention effects. The lack of significance for the covariate in the models was somewhat expected

due to the small number of trials being assessed. The Cochrane Collaboration has suggested that meta-regression models should not be used to assess the influence of covariates on treatment effect unless there are at least ten studies included in the meta-analysis (The Cochrane Collaboration 2003).

From the adjusted IPD models, although no evidence of effect modifiers were seen for the covariates under investigation; several predictors for EADL were found. These included gender and baseline dependency; however age did not appear to impact on the results.

Male patients were found to have higher EADL scores than compared to females; this finding may be related to differences in the goals that the patients aim to achieve at the end of the intervention phase. Bearing in mind the ages of the patients included in this review, it is thought that the main aim of the men in this group is to get their independence to be able to drive a car again which assumes independence in personal tasks; where as the women are more likely to want to gain independence in household tasks. Therefore, since driving a car (and the associated tasks that go with this factor) scores highly on the EADL the differences seen in gender may well be a true finding.

Dependent patients were found to score approximately 6 points lower on the EADL as compared to patients with independence. This is intuitive since the dependent patients have a lack of ability to be able to

perform the higher functional tasks that are required of the EADL assessment since they are still dependent in personal ADL tasks.

Two of the trials in the meta-analysis had multiple intervention groups (Drummond *et al.* 1995; Parker *et al.* 2001) and having data at the IPD level enabled contrasts to be specified for the type of intervention received which wouldn't have been possible using the conventional subgroup analysis methods associated with summary data. In this instance, subgroup analysis was an inefficient method and also assumes independence for the control group, which was used more than once in the subgroup analyses.

Therefore, it appears that there were little differences in the results from the models using summary data or IPD in this meta-analysis. However, the individual patient data models allowed for more in-depth analysis to be performed and appeared to be more efficient where multiple intervention groups exist.

6.14.4 Discussion of the results

This systematic review provides substantial evidence for the efficacy of occupational therapy given to stroke patients in the community setting. The aim of this systematic review was to estimate the extent that occupational therapy given in the community setting to patients following a stroke generally influences EADL.

6.14.4.1 Impact of occupational therapy on Extended Activities of Daily Living

The principal finding from this systematic review was that occupational therapy for stroke patients living in the community was associated with a higher EADL score at the end of intervention and end of trial. This indicated that stroke patients were able to carry out more activities of daily living, such as walking outdoors, household chores or travelling on public transport.

Independence in any one of these activities would enable the patient to participate in the more demanding activities of daily living, thereby adding to their quality of life. Although a definition of the amount of change for EADL that would constitute a clinically meaningful improvement was not specified *a priori*, it is felt that an increment of one point may be clinically important. This modest benefit is in keeping with previously published studies, including the findings of a stroke unit trial (Juby *et al.* 1996), and is not negated by the neutral findings of TOTAL (Parker *et al.* 2001).

6.14.4.2 Impact of occupational therapy on other outcomes

Additionally, community based occupational therapy appeared to extend across a range of outcomes measures including activities of daily living, and leisure participation.

Activities of daily living were assessed using either the Barthel Index or the Rivermead self care Assessment. This outcome measure is known to have ceiling and floor effects, therefore it was reassuring that significant intervention effects were seen between the occupational therapy and control groups, since it implies that occupational therapy dies not just improve higher functional tasks such as those measured by the EADL, but also improve more basic personal tasks such as bathing, brushing hair, and continence.

Leisure participation was measured using the Nottingham Leisure Questionnaire at the end of intervention and end of trial. Occupational therapy was significantly associated with higher leisure scores at the end of trial. This implies that occupational therapy can have long lasting effects on leisure participation and possibly life satisfaction (Allen *et al.* 1984).

6.14.4.3 Differences between TOTAL and the other trials

The trial by Parker and colleagues may have demonstrated little benefit from community occupational therapy for several reasons (Parker *et al.* 2001). The use of postal outcome may have made the findings of TOTAL less open to observer bias and so it is possible that the smaller effect seen in TOTAL is more genuine that that found in other trials where independent assessments were implemented. Another reason why the TOTAL study may have not found a large clinical benefit is that

the intervention in TOTAL was administered by clinicians, and not research occupational therapists, who may have been less motivated as their daily work was not contributing to a higher degree (see Section 6.14.2.5). Another possible reason is that the research protocol imposed some restrictions on the type of interventions making their effectiveness less than optimal.

6.14.4.4 Impact of the type of intervention

Subgroup analyses indicated that the benefits of occupational therapy were greatest when targeted where occupational therapy aimed at influencing ADL appeared to improve extended activities of daily living. Conversely, further analyses suggested that occupational therapy directed at influencing leisure pursuits improved leisure activities, but not EADL scores (Walker *et al.* 2004). This observation that the provision of one specific intervention does not generalise to other areas is contrary to the current view held by many clinicians (Walker *et al.* 2003). However the findings from this systematic review and meta-analysis are consistent with other recent stroke rehabilitation trials (Seitz *et al.* 1987; Dean *et al.* 1997).

Although the active intervention was dichotomised based on either activities of daily living or leisure therapy, this may still not adequately describe the intervention received since the components of the interventions used in the trials were tailor-made for each patient. The

frequency and intensity of the intervention could not be assessed in this meta-analysis since it was found that they were highly correlated. Additionally, unmeasured factors, such as the number of aids and adaptations or the seniority of the therapist, are thought to be of influence too however, the data set had very incomplete data for these measures and further investigations could not be performed.

6.14.4.5 Generalisability of the findings

It is only in the last decade that a research culture has existed within the occupational therapy profession and existing evidence is sparse. This analysis of community occupational therapy trials included only relatively recent published work and no old unpublished studies were identified. An extensive search strategy was used to identify eligible studies, and statistical testing for missing trials was non-significant. Therefore probably all of the available data had been identified. The unavailable data for one of the identified trials (Turton *et al.* 1990), accounted for only 2% of the total data. The inclusion of these data would not have significantly altered the findings. There are currently two ongoing trials that need to be included in any future analyses. However, the results from this review should be generalisable since they form the collective body of evidence from nine randomised controlled trials of community occupational therapy in stroke patients. Although the trials individually attempted to recruit patients which were representative of a

population of stroke patients, this may have not been fully achieved and therefore the results from this meta-analysis may be biased towards patients which are likely to use this service.

The findings from this meta-analysis provide a balanced interpretation of the available evidence and endorse an earlier systematic review of published data by Steuljens and colleagues (Steuljens *et al.* 2003). However, our research provides more information about the relationship between specific interventions and outcome and shows the first results between patient characteristics and outcomes.

6.14.5 Practical implications

This review has demonstrated that occupational therapy offered to stroke patients in the community can improve their rehabilitation needs quickly and with lasting effects. Therefore, this review suggests that all stroke patients within one year of stroke onset should be offered occupational therapy in the community setting. In depth analyses were performed in this review which highlighted that certain subgroups of patients may benefit the most from occupational therapy. No significant interactions were seen between treatment and several predictors. However, analyses did indicate that male stroke patients and patients which were not dependent at baseline had higher EADL scores at the end of intervention. Surprisingly, age of the patients was not found to be an important predictor of a patients EADL score. However, these factors

were identified through post hoc analyses and the findings need to be validated from other data of a randomised controlled trial to truly assess whether these patients benefit the most from occupational therapy.

The results from this systematic review has important implications to service providers who need to ensure that patients are offered specific interventions, such as community occupational therapy, to those who would benefit the most.

There may well be differences between services offered in these trials because of the differences in interventions and settings, and in view of this, work now is needed to characterise the necessary conditions for effective and efficient services. However the provision of occupational therapy remains justified on evidence-based grounds and it would appear that the rehabilitation needs of a substantial number of stroke patients in the community can be met feasibly by occupational therapy with measurable and lasting benefits.

The costs of such a service as specified in this review could not be estimated from the data, and if such a scheme was available to all stroke patients the costs of providing such a service needs to be minimised to make the scheme viable. An attempt to quantify the optimal frequencies and intensities of the occupational therapy intervention were performed; however, analysis of these factors appeared to be closely negatively correlated with the baseline

dependency of the patients and therefore could not be quantified. The optimal amount these factors to ensure that the costs of providing such a service to the community are minimised could not be assessed.

6.14.6 Future research

This study attempted to combine all of the available evidence and to estimate the benefit of an intervention of occupational therapy to stroke patients in the community. However, it has not been able to identify exactly what it is about occupational therapy that significantly improved the patients EADL. This was due to not only a lack of specific data relating to the intervention used within the trials, but also due to the complexities involved with tailor making these intervention packages to the individual patients.

CHAPTER 7

DIPYRIDAMOLE IN STROKE PATIENTS: A SYSTEMATIC REVIEW AND META-ANALYSIS OF RANDOMISED CONTROLLED TRIALS

7.1 Introduction

Aspirin is recommended for use in patients with prior stroke or transient ischaemic attack to reduce the risk of recurrence (Lees *et al.* 2000). In a meta-analysis of summary data, the Anti-Thrombotic Trialists (ATT) found that aspirin reduced the relative odds of further vascular events by 22% (Antithrombotic Trialists Collaboration 2002). However, the side effects of aspirin (principally gastrointestinal disturbance and bleeding) and its modest efficacy mean that alternative or additional antiplatelet agents might be useful clinically. A number of alternative antiplatelet agents exist for use after stroke, including dipyridamole and clopidogrel. These antiplatelet agents have been shown to have a similar efficacy as aspirin in reducing recurrence in stroke patients (CAPRIE Steering committee 1996; Diener *et al.* 1996). However, these drugs are more expensive than aspirin, so their use is more limited.

It is postulated that the combination of either one of the treatments, dipyridamole or clopidogrel, with aspirin may provide extra benefits than compared to a single treatment alone. The largest study to date, ESPS II, was a factorial designed study of aspirin, dipyridamole, a combination of the two agents, or placebo. This study found that there was added benefit in reducing stroke recurrence by using the combination of aspirin and dipyridamole as compared to either dipyridamole (relative risk reduction, RRR 24.7%) or aspirin (RRR 23.1%) alone (Diener *et al.* 1996). However, the routine use of dipyridamole in secondary

prevention after cerebrovascular events has been controversial. This key trial (Diener *et al.* 1996), has been criticised on a number of grounds based on the design and conduct of the study (Davis *et al.* 1998). First, the relatively high recurrence rate seen in the aspirin only group may be related to the low level of dose (50 mg daily; as compared with ESPS 330mg three time daily), whereas a higher dose may have significantly lowered this rate. Additionally, high levels of drop-outs were seen in the dipyridamole and dual treatment groups, which were related to adverse events of headache and gastrointestinal bleeding. Although these criticisms do not necessarily invalidate the results that combined aspirin and dipyridamole was superior to both aspirin alone and dipyridamole alone in preventing further stroke. Additionally, ATT did not find that the combination of aspirin and dipyridamole was superior to aspirin alone in reducing a composite vascular outcome (comprising non-fatal stroke, non-fatal myocardial infarction and vascular death) in patients with prior vascular disease (Antithrombotic Trialists Collaboration 2002).

In addition to the ESPS II trial, several other smaller studies have been performed which assessed the combination of dipyridamole and aspirin in the secondary prevention of stroke. Therefore to fully evaluate the efficacy of the combination treatment, the information from these smaller trials needs to be taken into consideration. The trials appear to be disparate with regards to the efficacy of the combined treatment on

stroke recurrence therefore it is essential for the review to investigate the variations by assessing the importance of internal factors at trial level such as the dose of aspirin, and at patient level such as the type of qualifying event, age and gender of the patients.

7.2 Design of study

A systematic review was performed involving a meta-analysis of individual patient data from randomised controlled trials of dipyridamole in patients with prior ischaemic stroke and/or transient ischaemic attack.

7.2.1 Measures of interest

The primary outcome measure was subsequent fatal or non-fatal stroke at the end of trial. Secondary outcome measures at the end of trial were non-fatal stroke, fatal and non-fatal myocardial infarction, vascular death, and a composite outcome consisting of non-fatal stroke, non-fatal myocardial infarction, or vascular death.

7.3 Study selection and search strategy

Trials were eligible if they fulfilled the requirements of randomisation, double blinding to allocation of treatment, recruitment of patients with previous stroke or transient ischaemic attack, and involved dipyridamole in at least one treatment arm.

A comprehensive literature search was performed to identify all eligible randomised controlled trials, whether published or unpublished, of

dipyridamole in cerebrovascular disease. Electronic searches of the Cochrane Library (Issue 4, 2002), MEDLINE (1966-2001 inclusive), EMBASE (1980-2002), and Web of Knowledge (1981-2002) were performed using the keywords 'dipyridamole', 'stroke', 'prevention', and 'cerebr*' in combination with the recommended search routine for identifying randomised controlled trials (Deeks *et al.* 1996). Reference lists from the identified publications and earlier reviews of dipyridamole in stroke (Sze *et al.* 1988; Lowenthal *et al.* 1994; Diener 1998; Tijssen 1998; Wilterdink *et al.* 1999; Antithrombotic Trialists Collaboration 2002) were also searched, and the trialists and manufacturer of dipyridamole (Boehringer Ingelheim) were contacted. No restrictions on language of were made. Non-randomised or confounded trials were excluded, as were those which involved non-stroke patients or did not include dipyridamole in one of the treatment arms.

Trial, year of publication	Subjects	Treatment groups	Primary outcomes	Result
Acheson, 1969	169	D / P	Stroke	Neutral
Guiraud-Chaumeil, 1982	440	AD / A / C †	Stroke	Neutral
AICLA, 1983	604	AD / A / P	Stroke	Positive
ACCSG, 1985	890	AD / A	Stroke, death	Neutral
Caneschi, 1985	50	AD / A / D	Stroke	Neutral
ESPS, 1990	2,500	AD / P	Stroke, death	Positive
ESPS II, 1996	6,602	AD / A / D / P	Stroke	Positive

A, aspirin; C, control; D, dipyridamole; P, placebo † All groups had dihydroergotamine.

Table 7.1 Trial characteristics for the identified trials

Data from two further ongoing studies were not available; the ESPRIT trial is comparing combined aspirin and dipyridamole with aspirin (De Schryver 2000), and PRoFESS is comparing two combinations, aspirin and clopidogrel versus aspirin and dipyridamole (Sacco *et al.* 2004).

Seven completed trials involving 11,255 patients were identified which assessed dipyridamole in the secondary prevention of stroke (Acheson *et al.* 1969; Guiraud-Chaumeil *et al.* 1982; Bousser *et al.* 1983; Caneschi *et al.* 1985; The American-Canadian Co-operative Study Group 1985; ESPS Group 1990; Diener *et al.* 1996) (Figure 7.1).

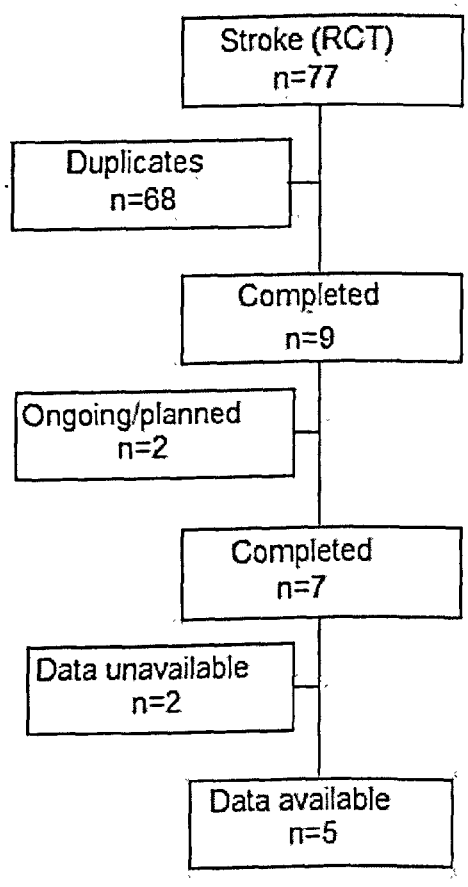

Figure 7.1 **Flow chart for trial identification and selection**

All of the seven completed randomised controlled trials had been published and used subsequent stroke as a principal outcome. From the publications, three of the trials found that recurrence was reduced with the combination treatment of dipyridamole and aspirin as opposed to other treatments (Bousser *et al.* 1983; ESPS Group 1990; Diener *et al.* 1996). The remaining four trials found no evidence individually that stroke recurrence was lowered when the combination treatment was used (Table 7.1).

7.4 Data collection and management

The principal investigator from each trial was contacted and asked if they would share their individual patient data with the collaboration. The data were exchanged electronically in all cases.

Data from two randomised controlled trials were unavailable, the first was published in 1969 and had been discarded previously (Acheson *et al.* 1969); the authors from the second trial could not be contacted (Caneschi *et al.* 1985). However, both of these trials were relatively small in size, 169 and 50 patients respectively, and contributed to only 1.9% of the total. Tabulated data from these two trials were extracted from the publications and used in the unadjusted analyses.

The shared individual patient data were checked against the results from the publications and any discrepancies were resolved through

contact with the relevant principal investigator. No major discrepancies were found in the data sets. Before the data were merged into a single data set in SAS version 8.02 (SAS Institute Inc), re-coding of the variables were performed to a uniform manner across all trials.

7.5 Assessment of publication bias and quality

Publication bias was assessed for subsequent stroke at the end of trial, and compared between the groups for the dual treatment against a combined group consisting of patients randomised to either aspirin alone, dipyridamole alone, or placebo. The trial by Acheson and colleagues was excluded from the assessment since the trial did not use the dual treatment of aspirin and dipyridamole.

Using Begg's funnel plot, the plot appeared to be relatively symmetrical in appearance and the point estimates were all located within the pseudo 95% confidence interval lines; hence there did not appear to be any evidence of publication bias from the seven trials (Figure 7.2). However, due to the small number of randomised controlled trials included in this meta-analysis further assessments were made.

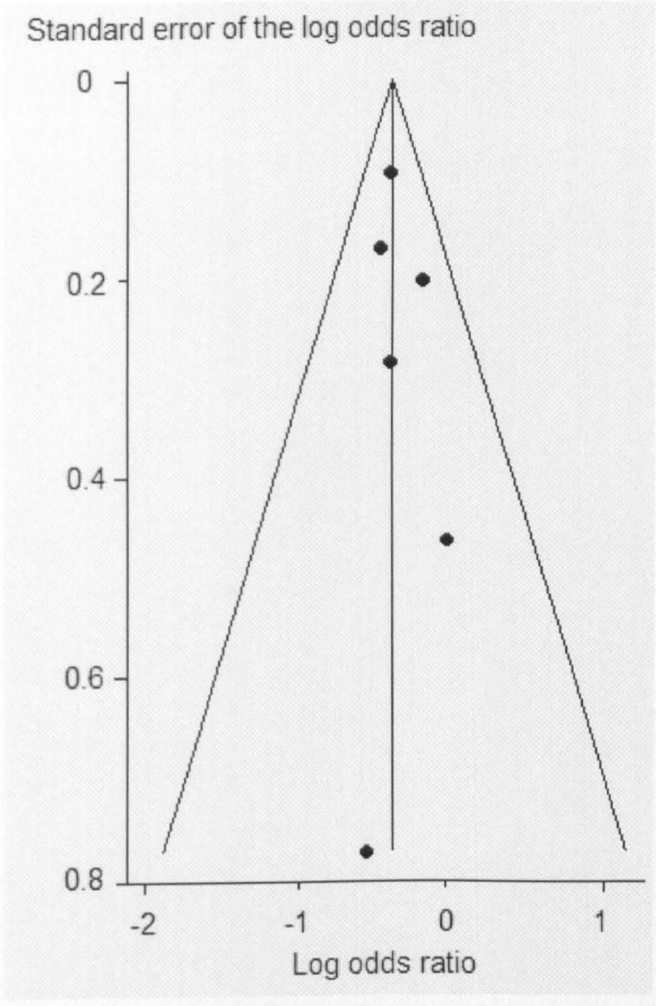

Figure 7.2 Begg's funnel plot for recurrence at the end of trial, AD vs. other groups

The Begg and Mazumdar rank correlation test found no evidence of publication bias (p=0.707, continuity corrected), also Egger's asymmetry test (p=0.482) and the trim and fill method did not find any evidence of publication bias.

Methodological quality for each trial was assessed on the basis of the method of randomisation employed, whether the allocation of treatment was concealed, the completeness of follow-up achieved, and on

whether the outcome assessment was blinded to the allocation of treatment using recognised criteria (The Cochrane Collaboration 2003). Six of the seven trials were deemed to have a high level of quality by satisfying at least three of the above criteria (Table 7.2). However, the remaining trial was found to have a lower quality due to the inadequate trial designs and reporting (Caneschi *et al.* 1985).

Trial, year of publication	Concealment to allocation	Randomisation	Blinded outcome assessor
Acheson, 1969	A	Central randomisation	Yes
Guiraud-Chaumeil, 1982	B	Pre-determined list	Yes
AICLA, 1983	A	Pre-determined schedule	Yes
ACCSG, 1985	A	Random allocation	Yes
Caneschi, 1985	B	Unclear	Unclear
ESPS, 1990	A	Not specified	Yes
ESPS II, 1996	A	Central randomisation	Yes

A= Low risk of bias, B= Moderate risk of bias, C= High risk of bias

Table 7.2 Assessment of quality for the identified trials

7.6 Trial level demographics

The data were assessed to determine whether patients were similar between the trials (Table 7.3). The demographics of the patients across the trials appeared to be relatively well balanced. All of the trials recruited slightly more male patients than female stroke patients. The Guiraud-Chaumeil trial had the highest percentage of recruited male patients. The percentage of patients whose qualifying event was stroke

varied considerably between the trials, ranging between 0% and 84%. The American-Canadian Co-operative study group trial only recruited patients whose qualifying event was a transient ischaemic attack.

Trial, publication year	Subjects	Age, mean (SD)	Male, (%)	Stroke (%)
Acheson, 1969	169	58.1 (-)	117 (69)	106 (63)
Guiraud-Chaumeil, 1982	440	62.4 (9.4)	372 (85)	260 (59)
AICLA, 1983	604	63.2 (10.3)	420 (70)	510 (84)
ACCSG, 1985	890	63.3 (10.2)	594 (67)	0 (0)
Caneschi, 1985	50	- (-)	36 (72)	40 (80)
ESPS, 1990	2500	63.5 (10.7)	1450 (58)	1302 (52)
ESPS II, 1996	6602	66.7 (11.1)	3828 (58)	5038 (76)

SD is the standard deviation. – Data not available

Table 7.3 Patient demographics at baseline by trial

7.7 Outcome assessments and measures

All of the seven randomised controlled trials recorded subsequent stroke, non-fatal stroke, and vascular death at the end of trial. Fatal or non-fatal myocardial infarction was only recorded in five of the trials (Guiraud-Chaumeil *et al.* 1982; Bousser *et al.* 1983; The American-Canadian Co-operative Study Group 1985; ESPS Group 1990; Diener *et al.* 1996); and the rates of myocardial infarction were low in all the comparison groups. A composite outcome was created for all trials and

was based on non-fatal events of stroke and myocardial infarction, and vascular death at the end of trial.

7.7.1 Type of intervention and formulation of dipyridamole

All except one study (Acheson *et al.* 1969) assessed the efficacy of the combination of dipyridamole and aspirin and compared this with aspirin, dipyridamole, or placebo. One trial had a control group instead of a placebo group (Guiraud-Chaumeil *et al.* 1982). Three studies had more than two groups of patients (Guiraud-Chaumeil *et al.* 1982; Bousser *et al.* 1983; Diener *et al.* 1996). Two formulations of dipyridamole were assessed in the trials; the majority of the trials used conventional formulation, and the remaining trial used modified release formulation (Diener *et al.* 1996). The doses for dipyridamole and aspirin varied between the trials (Table 7.4).

7.7.2 Timing of assessments and measures

The follow-up assessments varied between the trials, ranging between 3 and 72 months post-enrolment, and averaged at approximately 27 months (Table 7.4). Also, in four of the trials the length of follow-up varied within the trials (Acheson *et al.* 1969; Guiraud-Chaumeil *et al.* 1982; Caneschi *et al.* 1985; The American-Canadian Co-operative Study Group 1985).

Trial, year of publication	Dipyridamole dose	Aspirin dose	Follow up (months)
Acheson, 1969	100-200 mg qds	-	15-37
Guiraud-Chaumeil, 1982	50 mg tds	300 mg tds	36-72
AICLA, 1983	75 mg tds	330 mg tds	36
ACCSG, 1985	75 mg qds	325 mg qds	24-60
Caneschi, 1985	75 mg tds	300 mg od	22-34
ESPS, 1990	75 mg tds	330 mg tds	24
ESPS II, 1996	200 mg bd	25 mg bd	24

od once daily, bd twice daily, tds thrice daily, qds four times daily.

Table 7.4 Doses of dipyridamole and aspirin and length of follow-up at the end of trial

7.8 Analysis of the merged data set using summary data methods

The data from the seven trials were merged into a single data set and involved 11,255 patients. The overall demographic characteristics were typical of an ageing stroke population, where the average age was 65 years (standard deviation, SD 11.0), and 6,700 (60%) of patients were male. Time from stroke onset to recruitment into the trial averaged 33.8 days (SD 64.7).

7.8.1 Data analysis for subsequent stroke at the end of trial

The proportion of patients with subsequent stroke for each treatment group within each trial was used as the primary outcome measure. The treatments were coded into four groups; dual treatment (aspirin and

dipyridamole), dipyridamole alone, aspirin alone or control/placebo. Comparisons were made to compare the efficacy of the dual treatment as compared to the other treatments. The odds ratios and variances for each trial were calculated as detailed in section 2.3.1.1.

7.8.2 Results for subsequent stroke at the end of trial

In a conventional fixed effect analysis as described in section 2.4.1, patients randomised to the dual treatment had significantly reduced risks of subsequent stroke as compared to patients that received dipyridamole alone (OR 0.72, 95% CI 0.58, 0.89), aspirin alone (OR 0.78, 95% CI 0.65, 0.93), or control/placebo (OR 0.60, 95% CI 0.50, 0.72).

7.8.3 Results for secondary outcomes

From a conventional fixed effect analysis as described in section 2.4.1; patients randomised to aspirin and dipyridamole had significantly reduced risks of non-fatal stroke (25%) and an event defined from the composite outcome (28%) as compared to those which received dipyridamole; reduced risks of non-fatal stroke (29%) and an event defined from the composite outcome (16%) as compared to patients receiving aspirin; and reduced risks of non-fatal stroke (41%), fatal and non-fatal myocardial infarction (33%) and an event defined from the composite outcome (35%) as compared to control/placebo (Table 7.5). However, dual treatment of dipyridamole and aspirin was not

significantly associated with a reduction in the risk of vascular death at the end of trial as compared to dipyridamole, aspirin, or control/placebo (Table 7.5).

	AD vs. D	AD vs. A	AD vs. C
Outcome	OR (95% CI)	OR (95% CI)	OR (95% CI)
Non fatal stroke	0.75 (0.57, 0.98)	0.71 (0.57, 0.88)	0.59 (0.48, 0.72)
MI, all	0.73 (0.47, 1.13)	0.98 (0.67, 1.43)	0.67 (0.46, 0.92)
Vascular death	0.85 (0.62, 1.15)	1.09 (0.85, 1.38)	0.89 (0.72, 1.11)
Composite	0.72 (0.60, 0.88)	0.84 (0.71, 0.98)	0.65 (0.56, 0.75)

OR is the odds ratio for dual treatment relative to comparison group, MI myocardial infarction, A aspirin, D dipyridamole, C control/placebo

Table 7.5 Relationship between outcome measures and dual treatment with aspirin and dipyridamole at the end of trial, based on summary data

7.9 Assessment of heterogeneity using summary data

The conventional methods presented have assumed that the trials all have a common underlying estimate for the treatment effect for each pair wise comparison. An assessment of heterogeneity needs to be performed to evaluate whether this assumption is correct for each comparison. Visual and statistical assessments of heterogeneity were investigated for the primary outcome measure for each pair wise comparison.

7.9.1 Graphical assessment of heterogeneity

Figure 7.3 shows the forest plot for subsequent stroke at the end of trial split by the comparison group. The forest plot indicates that there is very little heterogeneity between any of the estimates for each pair wise comparison. However, to formally assess the presence of heterogeneity between the estimates, statistical testing of heterogeneity was performed.

7.9.2 Statistical assessment of heterogeneity using classical and Bayesian methods

To formally assess whether there was evidence of heterogeneity between the trials estimates the Cochran's homogeneity Q test and Higgins and Thompson I^2 were performed. Also, other analyses were used to quantify the heterogeneity between the trial estimates using un-weighted and weighted MM, ML, REML, EB and FB (Table 7.6).

Cochran's homogeneity test did not find evidence of heterogeneity between the trial estimates for the three pair wise comparisons (AD vs. D p=0.81; AD vs. A p=0.90; AD vs. C/P p=0.84). I^2 was estimated as zero for all of the comparisons (Table 7.6) indicating that none of the differences between the trial estimates could be attributed to heterogeneity.

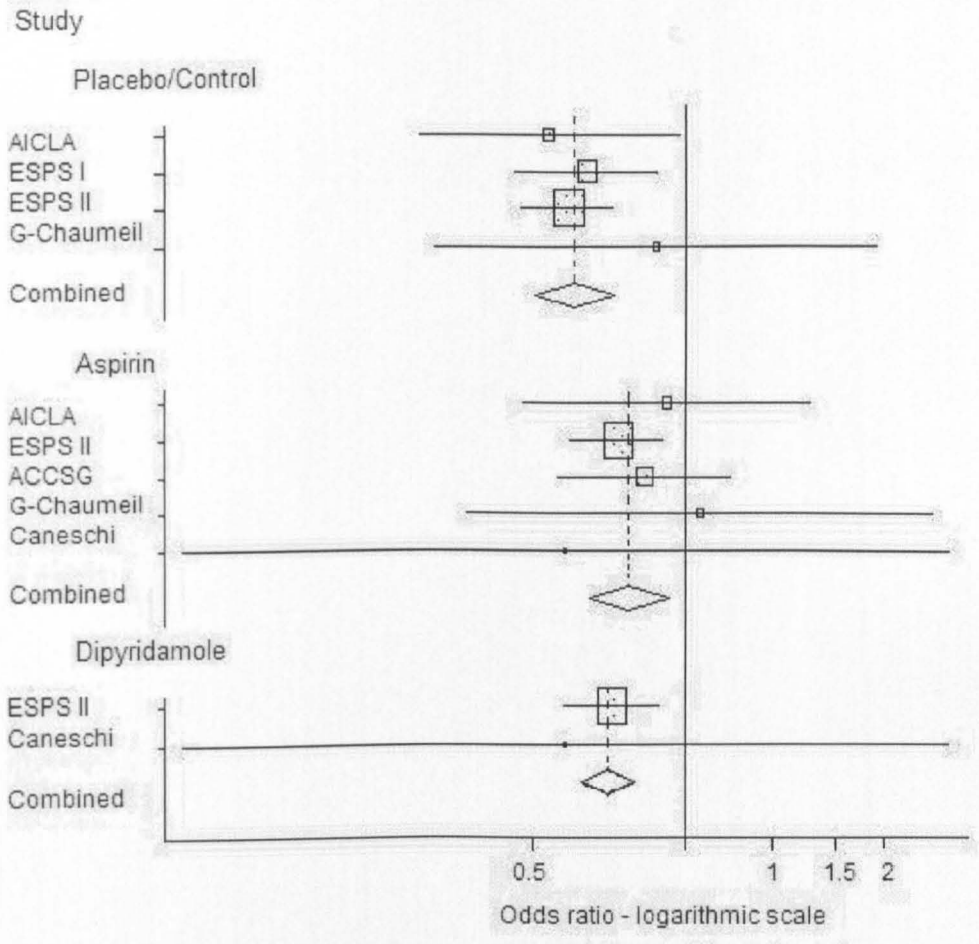

Figure 7.3 Forest plot for subsequent stroke at the end of trial split by type of comparison group

	Dual treatment vs. Dipyridamole			Dual treatment vs. Aspirin			Dual treatment vs. Control/Placebo		
Heterogeneity Test	χ^2	I^2		χ^2	I^2		χ^2	I^2	
Cochran's Q test	0.057			1.082			0.855		
I^2		0			0			0	
Estimation methods	Odds Ratio	95% CI	$\hat{\tau}^2$	Odds Ratio	95% CI	$\hat{\tau}^2$	Odds Ratio	95% CI	$\hat{\tau}^2$
MM, un-weighted	0.717	0.577, 0.891	0	0.776	0.647, 0.930	0	0.604	0.510, 0.716	0
MM, weighted	0.717	0.577, 0.891	0	0.776	0.647, 0.930	0	0.604	0.510, 0.716	0
ML	0.717	0.577, 0.891	0	0.776	0.647, 0.930	0	0.604	0.510, 0.716	0
REML	0.717	0.577, 0.891	0	0.776	0.647, 0.930	0	0.604	0.510, 0.716	0
EB	0.717	0.577, 0.891	0	0.776	0.647, 0.930	0	0.604	0.510, 0.716	0
FB	0.727	0.131, 4.108‡	0.844	0.801	0.606, 1.092‡	0.149	0.611	0.460, 0.608‡	0.141

Odds ratio is the odds of subsequent stroke on dual treatment relative to the comparison group, ‡ 95% credibility intervals

Table 7.6 Heterogeneity test and analysis results for subsequent stroke at the end of trial based on summary data

From Table 7.6, the estimates for heterogeneity from the classical and EB methods were zero for all of the pair wise comparisons, and hence the estimates for the treatment effect were identical as to those from the fixed effect models and indicated that the dual treatment was associated with significant reductions in subsequent stroke by 28%, 22%, and 40% as compared to dipyridamole alone, aspirin alone or control/placebo, respectively.

The FB approach found heterogeneity in all of the pair wise comparisons, and hence the standard errors for the treatment effect were larger as compared to the estimates from the classical or EB models. Although the magnitudes of the treatment effect were similar for the comparisons as compared to the classical and EB models, only the comparison of dual treatment vs. control/placebo remained statistically significant.

7.10 Exploring heterogeneity

It appears that the differences in the estimates for the treatment effects between the trials may be due to sampling errors only and not due to heterogeneity. Although these tests did not find evidence of heterogeneity, it may still be advantageous to determine whether any prognostic factors either relating to patient characteristics, such as age and gender; or trial factors, such as dose of treatments, can explain some of the residual variation between the trials. Subgroup analyses were used to assess the influence of trial factors, and meta-regression analyses were used to assess the influence of patient characteristics.

Data relating to patient characteristics were unavailable for two of the trials due to lack of individual patient data from the authors and could not be included in the following analyses (Acheson *et al.* 1969; Caneschi *et al.* 1985); however, data relating to gender of the patient could be extracted from the publication of the Caneschi trial.

7.10.1 Subgroup analyses

Subgroup analysis was used to assess whether the trial level factors for the type of formulation used (conventional or modified release), were important for two of the three pair wise comparisons. The comparison of dual treatment against dipyridamole alone was not performed since only the trial relating to ESPS II would be included.

7.10.1.1 Type of dipyridamole administration

Dipyridamole may be given in two oral formulations, either conventionally in a standard tablet or as a modified release preparation which maintains blood levels longer. In four of the trials, dipyridamole was given conventionally, however in ESPS II the modified release formulation was given (Figure 7.4). The conventional formulation yielded a non-significant reduction in the risk of stroke recurrence when the dual treatment was compared to aspirin alone (OR 0.865, 95% CI 0.628, 1.190); where as patients randomised to dual treatment had a significant 26% reduction in the risk of subsequent stroke as compared to patients which received aspirin alone when the modified release formulation was used (OR 0.736, 95% CI 0.591, 0.918). However, the p value for the change in heterogeneity was 0.418 indicating that there

does not appear to be evidence of a difference between the two estimates and their confidence intervals.

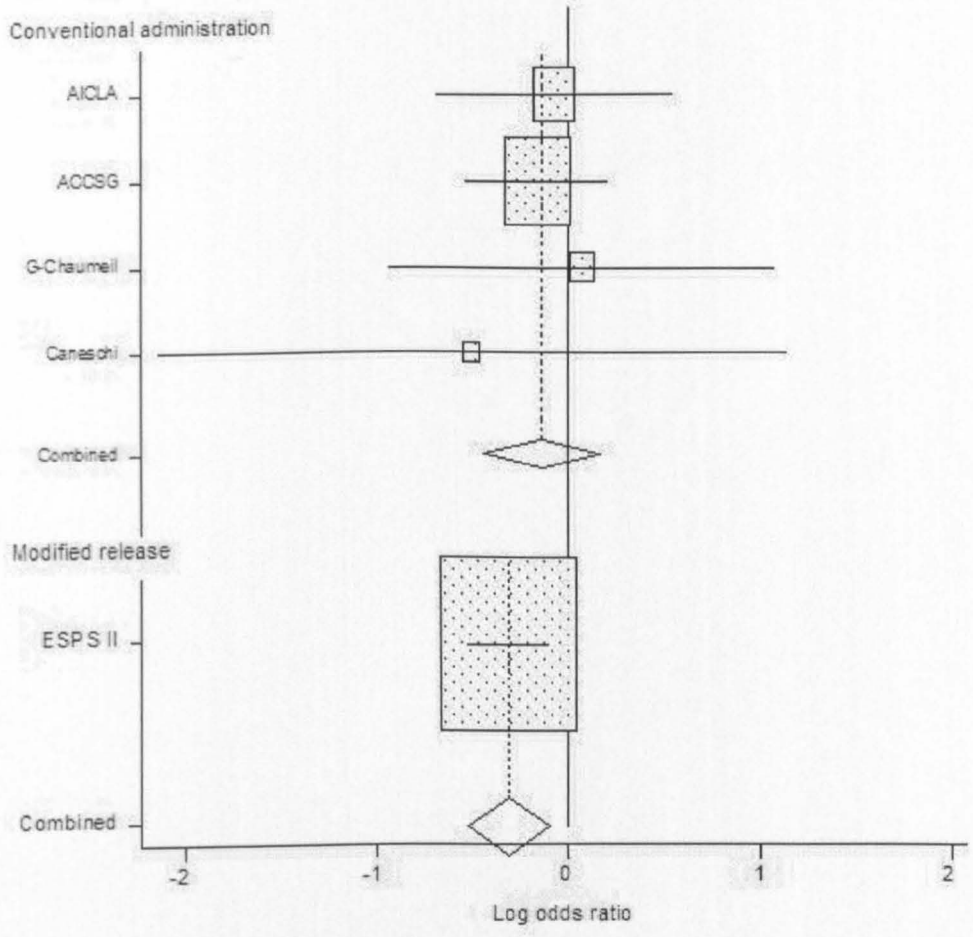

Figure 7.4 Subgroup analysis plot for recurrence at the end of trial by the formulation for dipyridamole for the comparison of dual treatment versus aspirin

For the comparison of dual treatment against control/placebo (Figure 7.5), similar reductions in the risk of subsequent stroke were seen for the two formulations, (conventional formulation OR 0.633, 95% CI 0.479, 0.836; modified release formulation OR 0.589, 95% CI 0.476, 0.728) (Figure 7.5). The p value for the change in heterogeneity was 0.685 indicating that there does not appear to be significant evidence of a difference between the two estimates and their confidence intervals.

Therefore the method of formulation of dipyridamole does not appear to significantly modulate the effectiveness of the combination treatment on the risk of recurrence at the end of trial. This subgroup analysis also allowed for the effect of excluding ESPS II to be determined in light of the criticisms presented earlier regarding this trial relating to the dose of aspirin used (see section 7.1).

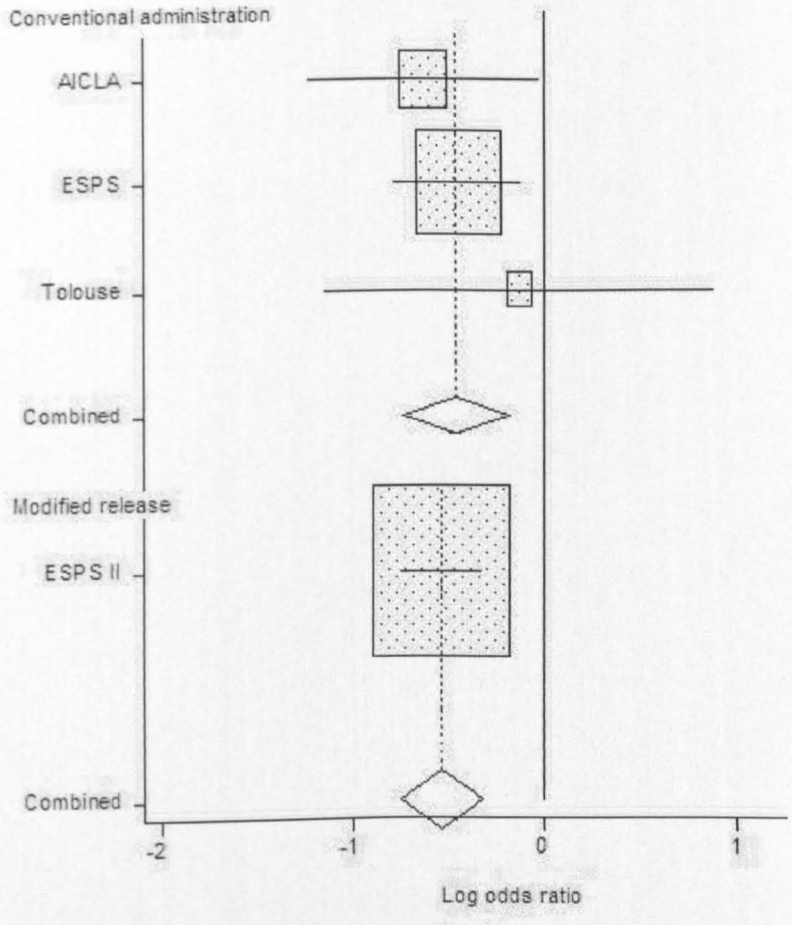

Figure 7.5 Subgroup analysis plot for recurrence at the end of trial by the formulation for dipyridamole for the comparison of dual treatment versus control/placebo

7.10.1.2 Dose of aspirin

With the exception of ESPS II and Caneschi trials, all of the other trials used relatively high level of dose for aspirin (range 900-1300 mg/day). When the studies with low aspirin doses are excluded from the analysis the magnitude for the treatment effects are slightly reduced (AD vs. A OR 0.88, 95% CI 0.63, 1.21; AD vs. C/P OR 0.63, 95% CI 0.48, 0.84) and only remain significant for the comparison of dual treatment against control/placebo. The comparison between dual treatment and dipyridamole alone was not performed since both of the eligible trials had been excluded.

7.10.2 Meta-regression analyses

The patient characteristics of interest were assessed using the difference between the percentage of the factor between the dual treatment group and the comparison groups of aspirin or control/placebo. The comparison against dipyridamole was not assessed using meta-regression since data from only two trials were available. The factors considered were age, gender, and the type of qualifying event.

Although heterogeneity was only found in the FB model, random effect models were used to explore the relationships between subsequent stroke and the covariates (see section 3.4). The methods used to estimate heterogeneity were the weighted and approximate MM, ML, REML, EB and FB. For the FB model assessments of convergence for the parameters were performed and found adequate when using a burn-in of 10,000 iterations and a sample chain of 40,000.

Method	Odds ratio[§]	95% CI	Age[Ω]	95% CI	$\hat{\tau}^2$
MM, approx	0.780	0.647, 0.940	0.963	0.438, 2.120	0
MM, weighted	0.780	0.647, 0.940	0.963	0.438, 2.120	0
ML	0.780	0.647, 0.940	0.963	0.438, 2.120	0
REML	0.780	0.647, 0.940	0.963	0.438, 2.120	0
EB	0.780	0.647, 0.940	0.963	0.438, 2.120	0
FB	0.790	0.545, 1.254‡	1.013	0.894, 1.139‡	0.207
Method	Odds ratio[§]	95% CI	Male[Ω]	95% CI	$\hat{\tau}^2$
MM, approx	0.761	0.627, 0.923	1.022	0.948, 1.103	0
MM, weighted	0.761	0.627, 0.923	1.022	0.948, 1.103	0
ML	0.761	0.627, 0.923	1.022	0.948, 1.103	0
REML	0.761	0.627, 0.923	1.022	0.948, 1.103	0
EB	0.761	0.627, 0.923	1.022	0.948, 1.103	0
FB	0.787	0.532, 1.222‡	1.018	0.920, 1.123‡	0.262
Method	Odds ratio[§]	95% CI	Stroke[Ω]	95% CI	$\hat{\tau}^2$
MM, approx	0.756	0.545, 1.049	0.926	0.665, 1.289	0.014
MM, weighted	0.745	0.567, 0.978	0.935	0.683, 1.279	0
ML	0.745	0.567, 0.978	0.935	0.683, 1.279	0
REML	0.745	0.567, 0.978	0.935	0.683, 1.279	0
EB	0.745	0.567, 0.978	0.935	0.683, 1.279	0
FB	0.737	0.450, 1.293‡	0.877	0.582, 1.329‡	0.216

§ Odds of subsequent stroke on dual treatment relative to aspirin when difference in percentage of covariate between the treatment groups equals zero, Ω Multiplicative increase in odds of subsequent stroke when the covariate is increased by 1%, ‡ credibility intervals.

Table 7.7 Subsequent stroke at the end of the trial using summary data meta-regression methods for the comparison of dual treatment and aspirin alone

Method	Odds ratio§	95% CI	Age$^{\Omega}$	95% CI	$\hat{\tau}^2$
MM, approx	0.622	0.497, 0.778	1.063	0.771, 1.466	0
MM, weighted	0.622	0.497, 0.778	1.063	0.771, 1.466	0
ML	0.622	0.497, 0.778	1.063	0.771, 1.466	0
REML	0.622	0.497, 0.778	1.063	0.771, 1.466	0
EB	0.622	0.497, 0.778	1.063	0.771, 1.466	0
FB	0.609	0.387, 1.007‡	0.987	0.611, 1.632‡	0.213
Method	Odds ratio§	95% CI	Male$^{\Omega}$	95% CI	$\hat{\tau}^2$
MM, approx	0.656	0.514, 0.838	0.930	0.797, 1.086	0
MM, weighted	0.656	0.514, 0.838	0.930	0.797, 1.086	0
ML	0.656	0.514, 0.838	0.930	0.797, 1.086	0
REML	0.656	0.514, 0.838	0.930	0.797, 1.086	0
EB	0.656	0.514, 0.838	0.930	0.797, 1.086	0
FB	0.626	0.451, 0.883‡	0.954	0.784, 1.160‡	0.179
Method	Odds ratio§	95% CI	Stroke$^{\Omega}$	95% CI	$\hat{\tau}^2$
MM, approx	0.629	0.472, 0.839	1.075	0.882, 1.310	0.034
MM, weighted	0.604	0.509, 0.716	1.007	0.604, 1.091	0
ML	0.604	0.509, 0.716	1.007	0.604, 1.091	0
REML	0.604	0.509, 0.716	1.007	0.604, 1.091	0
EB	0.604	0.509, 0.716	1.007	0.604, 1.091	0
FB	0.626	0.448, 0.863‡	1.074	0.875, 1.366‡	0.185

§ Odds of subsequent stroke on dual treatment relative to aspirin when difference in percentage of covariate between the treatment groups equals zero, Ω Multiplicative increase in odds of subsequent stroke when the covariate is increased by 1%, ‡ credibility intervals.

Table 7.8 Subsequent stroke at the end of the trial using summary data meta-regression methods for the comparison of dual treatment and control/placebo

Similar estimates for the treatment and covariate effects were seen from all of the classical and Bayesian models for each pair wise comparison (Tables 7.7 and 7.8). Slight differences between the magnitudes of the treatment and covariate effects from the FB models as compared to the classical and EB models is related to the estimates for the between trial heterogeneity. Also for the comparison of dual treatment against aspirin alone, the magnitude of heterogeneity found from the FB models impacted on the credibility intervals of the treatment effect so that they became non significant.

In the age adjusted models, significant reductions in the estimated relative odds ratio for the treatment effect was seen in patients randomised to dual treatment as compared to either aspirin alone, or control/placebo. However, age did not appear to be an important predictor of the estimated relative odds ratio for the treatment effect in either of the comparisons (Tables 7.7 and 7.8). Also, in the gender adjusted models, gender did not appear to be an important factor on the estimated relative odds ratio for the treatment effect in either of the comparisons (Tables 7.7 and 7.8).

When the models were adjusted for the type of qualifying event, the approximate MM method found small magnitudes of heterogeneity between the trials when the dual treatment was compared to aspirin alone and control/placebo, 0.014 and 0.034 respectively. The heterogeneity very slightly impacted on the odds ratio estimates by slightly reducing the importance of the dual treatment effect and emphasising the importance of the covariate effect; however, all of

the models found that the covariate did not significantly impact on the estimated relative odds ratio for the treatment effect (Tables 7.7 and 7.8).

7.11 *Individual patient data methods*

Using a classical fixed effect IPD model, patients randomised to dual treatment had significantly reduced risks of stroke at the end of trial by 26% (95% CI 7%, 31%) as compared to dipyridamole; by 22% (95% CI 4%, 36%) as compared to aspirin; and by 39% (95% CI 27%, 50%) as compared to control/placebo (Table 7.9).

Model method	Odds ratio (SE)	95% CI
Fixed effect modelling		
AD vs. D	0.744 (1.105)	0.593, 0.933
AD vs. A	0.782 (1.093)	0.640, 0.957
AD vs. C	0.606 (1.087)	0.502, 0.732
Random effect modelling		
Classical REML		
AD vs. D	0.744 (1.105)	0.593, 0.933
AD vs. A	0.782 (1.093)	0.640, 0.957
AD vs. C	0.606 (1.087)	0.502, 0.732
$\hat{\tau}^2$	0 (-)	-
Full Bayesian		
AD vs. D	0.712 (1.228)	0.476, 1.065‡
AD vs. A	0.736 (1.137)	0.572, 0.947‡
AD vs. C	0.576 (1.155)	0.434, 0.764‡
$\hat{\tau}^2$	0.118 (0.098)	-

SE Standard error, ‡ 95% credibility intervals, A aspirin, D dipyridamole, C control/placebo

Table 7.9 Subsequent stroke at the end of trial using individual patient data methodologies

7.12 *Assessment of heterogeneity using individual patient data methods*

An interaction term between trial and treatment status was included in the above model to test whether there was heterogeneity between the trials. Allowing for this test having low power when a small number of trials are combined in the meta-analysis, there was no evidence of heterogeneity between the trial estimates (p =0.988).

Classical and Bayesian random effect models were used where a common heterogeneity parameter was assumed. Additionally, for the FB models sample chains were run for a burn in of 5,000 iterations and monitored for 20,000 iterations. This length of chain was found to produce adequate assessments of convergence using techniques as described in section 2.5.3.

The classical IPD model found that the heterogeneity was estimated as zero and hence the results for the treatment efficacies where identical to those yielded from the fixed effect IPD models; where dual treatment was associated with significant reductions in subsequent stroke as compared to dipyridamole alone, aspirin alone, and control/placebo by 26% (95% CI 7%, 41%); 22% (95% CI 4%, 36%); and 39% (95% CI 27%, 50%), respectively (Table 7.9).

Although the treatment estimates for the comparisons were similar to those estimated form the classical model, the FB model estimated a small amount of heterogeneity between the trial estimates. The heterogeneity impacted on the standard errors for the treatment effect making them slightly larger; however, the treatment effects only

became non-significant for the comparison against dipyridamole alone (Table 7.9).

7.13 Exploring heterogeneity using individual patient data methods

Although only the Bayesian model estimated some heterogeneity between the trial estimates, it may be advantageous to assess the impact of covariates on the treatment efficacies. Therefore trial level covariates, such as the dose of aspirin; and patient level covariates, such as the age, gender and type of qualifying event of the patients, were included in the random effect models.

7.13.1 Trial level covariates using individual patient data

The dose of aspirin was modelled using a classical random effect IPD meta-regression model as described in Section 3.5.1.3 to assess its impact on subsequent stroke at the end of trial. The daily dose of aspirin did not significantly affect the risk of subsequent stroke (p =0.840).

7.13.2 Patient level covariates using individual patient data

Classical and Bayesian models were performed to assess the impact of age, gender and type of qualifying events on subsequent stroke at the end of trial using random effects IPD models. Assessments of convergence were found to be adequate using a burn-in of 10,000 iterations and sample chains of 40,000 iterations.

Initially, interaction terms between the covariate and treatment effects were introduced into the models to ascertain whether the treatment effects varied across the range of covariates. All of the models found no evidence of an interaction and hence the interaction terms were dropped from the models (age interaction p =0.99, gender interaction p =0.94, type of qualifying event interaction p =0.59).

Models	*Classical REML*		*Full Bayesian*	
	Odds ratio (SE)	95% CI	Odds ratio (SE)	95% CI
AD vs. D	0.740 (1.107)	0.607, 0.903	0.796 (1.344)	0.446, 1.420‡
AD vs. A	0.782 (1.094)	0.656, 0.933	0.847 (1.235)	0.560, 1.239‡
AD vs. C	0.606 (1.088)	0.513, 0.715	0.634 (1.223)	0.427, 0.941‡
Age	1.034 (1.003)	1.027, 1.040	1.034 (1.003)	1.027, 1.040‡
$\hat{\tau}^2$	0 (-)	-	0.105 (-)	-
AD vs. D	0.742 (1.106)	0.609, 0.903	0.808 (1.277)	0.500, 1.305‡
AD vs. A	0.782 (1.093)	0.657, 0.931	0.876 (1.220)	0.594, 1.294‡
AD vs. C	0.607 (1.087)	0.515, 0.715	0.646 (1.223)	0.435, 0.958‡
Gender	1.052 (1.066)	0.929, 1.191	1.063 (1.067)	0.936, 1.207‡
$\hat{\tau}^2$	0 (-)	-	0.232 (-)	-
AD vs. D	0.741 (1.085)	0.632, 0.870	0.771 (1.310)	0.454, 1.310‡
AD vs. A	0.783 (1.094)	0.656, 0.933	0.857 (1.228)	0.573, 1.281‡
AD vs. C	0.607 (1.088)	0.514, 0.715	0.635 (1.213)	0.435, 0.922‡
QE	1.571 (1.084)	1.339, 1.843	1.543 (1.093)	1.297, 1.836‡
$\hat{\tau}^2$	0 (-)	-	0.104 (-)	-

Age, in years; Gender, male vs. female; QE qualifying event, stroke vs. TIA; $\hat{\tau}^2$ is an estimate of heterogeneity; ‡ 95% credibility intervals

Table 7.10 Subsequent stroke at the end of trial using individual patient data adjusted for patient level covariates

7.14 Discussion of findings

The discussion of the findings from this systematic review is presented in six sections which relate to a summary of the overall findings, limitations of the data used in the review, comparison of the statistical methods, discussion of the results, practical implications of the results and suggestions for future research.

7.14.1 Summary of the overall findings

Seven randomised controlled trials of dipyridamole in stroke/TIA patients were identified from a comprehensive search strategy which was performed up to December 2002. Individual patient data were available from five of these trials; data from the other two trials had been either previously discarded or we were unable to contact the authors of the trial. However, tabulated data from these two trials was extracted from the publications and merged with the individual patient data from the other trials. The data base comprised of data from 11,255 stroke patients, and the patients from these trials were randomised to either a dual treatment of aspirin and dipyridamole, mono treatment with aspirin, mono treatment with dipyridamole or control/placebo. Baseline characteristics of the patients appeared to be balanced between the treatment groups from the individual trials.

At the end of trial, patients randomised to the dual treatment of aspirin and dipyridamole had significantly lower risks of subsequent fatal or non-fatal stroke, non-fatal stroke, and an event as defined by the composite outcome (non-fatal stroke, non fatal myocardial infarction,

or death from a vascular event) as compared to either mono treatment with aspirin, mono treatment with dipyridamole or control/placebo. Additionally, patients randomised to the dual treatment had significantly lower risks of fatal or non-fatal myocardial infarction than compared to the control/placebo group.

7.14.2 Limitations of the data used

7.14.2.1 Number of studies included in the meta-analysis

The search strategy used in this systematic review identified seven randomised controlled trials of dipyridamole given to stroke/TIA patients for secondary prevention. Outcome data for these trials were available from all of the trials either through sharing individual patient data or extracted from the publications. However, the focus of the meta-analysis was based on assessing the efficacy of the combined treatment of aspirin and dipyridamole against mono treatment or control/placebo. The oldest trial by Acheson and colleagues only assessed the efficacy of the mono treatment of dipyridamole compared with placebo (Acheson *et al.* 1969), and hence this trial wasn't included in any of the analyses.

7.14.2.2 Timings of the outcomes

It was decided *a priori* to use the end of trial as the primary outcome assessment time. This timing was used to allow for a consistent end point to be specified and was thought preferable to using a particular time point such as one year, since there was a lack of data at a consistent time point across the trials. However, within some of the

trial proportions of patients weren't follow-up to the end of trial (Acheson *et al.* 1969; Guiraud-Chaumeil *et al.* 1982; Caneschi *et al.* 1985; The American-Canadian Co-operative Study Group 1985). Generally, these patients were follow-up for a minimum length of time; however, it is acknowledged that this could have lead to biased estimates for these trials being estimated; where an over-estimate of the treatment effect would be yielded if the patients which dropped out early did so just before they suffered a subsequent event.

It would have been more efficient to have used survival analysis to analyse the data for subsequent stroke, where not only the timing of the events are taken into consideration, but also allow for the data from the early drop out patients to be included in the analysis by censoring their data. However, data for the time to subsequent stroke was not available from these four trials and hence the analysis could not be undertaken.

7.14.2.3 Extent of the shared individual patient data

Although the original trials generally collected considerable amounts of data during the baseline and end of trial follow up times, the data shared with the collaboration was not consistent across the trials. For the ESPS and ESPS II trials, all recorded measurements taken throughout the trials duration were shared with the collaboration; however, for the other three trials (Guiraud-Chaumeil *et al.* 1982; Bousser *et al.* 1983; The American-Canadian Co-operative Study Group 1985) the individual patient data was shared from the

Antithrombotic Trialists Collaboration (Antithrombotic Trialists Collaboration 2002) and not from the original trialists.

Bias may also result if data from some identified studies are unavailable, as occurred here with two studies (Acheson *et al.* 1969; Caneschi *et al.* 1985) although the missing data comprised <2% of the total data set. Tabulated data from the publications of these trials were included in the unadjusted analyses however, data for the exploration of heterogeneity were not available from the publications and hence these studies were not included in these analyses. Finally, we were not able to adjust for all potential prognostic factors, such as previous ischaemic heart disease and time from event to treatment; since these data were not available for most of the identified trials. Hence, some sources of heterogeneity will not have been explored in the analyses.

7.14.2.4 Outcome assessments

The primary and secondary outcome assessments used in this systematic review were selected *a priori* before the trials were identified and following discussion with the principal collaborators of the Dipyridamole in Stroke Collaboration (DISC).

A trial was deemed eligible for inclusion into the systematic review if it recruited either TIA and/or stroke patients. This was to allow for trials with the widest definition of a cerebrovascular event to be included. Therefore, subsequent stroke was chosen as the primary outcome measure because it is the most common vascular events in patients with recent cerebrovascular events. This outcome assessment

appears to be a justified assessment to use to assess the efficacy of the dual treatment of aspirin and dipyridamole since it a component of the primary outcome measures used, either alone or with other events, in each of the identified trials.

The secondary outcome measures specified included non-fatal stroke, combined fatal and non-fatal myocardial infarction, vascular death, and a composite outcome of non-fatal stroke, non-fatal myocardial infection, and vascular death to assess the impact of the dual treatment on all types of vascular outcomes.

7.14.3 Comparison of the statistical methods

7.14.3.1 Assessments of heterogeneity

The impact of heterogeneity on the results was assessed using various statistical models for subsequent stroke at the end of trial. Initially, statistical assessments of heterogeneity were performed based on summary data methods, which consistently found no heterogeneity between the results, from using Cochran's homogeneity test and I^2. No evidence of heterogeneity was also seen when using an individual patient data model by including a fixed effect interaction term between treatment and trial. Although, it is widely acknowledged that the power of both Cochran's homogeneity test and including an interaction term in an IPD model, are affected by having a small number of trials in the meta-analysis (Fleiss 1986; Whitehead *et al.* 1991; Brown *et al.* 1994; Thompson 1994), I^2 has found to not be affected (Higgins *et al.* 2002); and therefore it appears

that the differences between the estimates from the individual trials may be due to sampling variation only.

7.14.3.2 Impact of quantifying heterogeneity

A range of estimation methods were used in random effect models to assess whether there were differences between the results. All of the classical summary and IPD models and the empirical Bayes models found no heterogeneity between the results; therefore the results for the treatment effect were identical from these models as compared with the fixed effect models. Conversely, the full Bayesian summary and IPD models consistently quantified small amounts of heterogeneity. This is not surprising since the nature of the full Bayesian models are to assume that all of the parameters to be estimated in the random effect models are random variables with pre-specified distributions; hence it is inevitable that some heterogeneity will be estimated.

Individual patient data models were then used to assess the impact of the dual treatment of aspirin and dipyridamole on subsequent stroke at the end of trial. From the classical models, patient randomised to the dual treatment were found to have significantly lower risks of subsequent stroke than compared to patients receiving aspirin alone (22% reduced), dipyridamole alone (28% reduced), or placebo/control (29% reduced).

The significance of these findings were also seen in the full Bayesian models for two of the three pair wise comparisons, however, for the

comparison against dipyridamole the efficacy of the dual treatment was no longer significant. Also, the estimates for the treatment effects were slightly impact on by full Bayesian model estimating the heterogeneity making the magnitudes slightly larger by approximately 4%.

7.14.3.3 Impact of exploring heterogeneity

Various estimation methods were then used in meta-regression models based on summary data to assess whether there were any treatment effect modifiers and whether there were any differences between the results from the estimation methods. All of the models yielded similar results for the treatment and covariate effect which implied that age, gender, and type of qualifying event did not modify the treatment effects. The lack of significance for the covariate in the summary data models was somewhat expected due to the small number of trials being assessed. The Cochrane Collaboration has suggested that meta-regression models should only be used where there are more than ten studies being assessed in the meta-analysis (The Cochrane Collaboration 2003).

From the adjusted IPD models, although there no evidence of effect modifiers for any of the covariates assessed; several predictors of subsequent stroke were found. The included age and type of qualifying event; however gender did not appear to impact on the results.

Age suggested a small but significant effect were a 30% increase in the risk of subsequent stroke was seen for every 10 year increase in age. However, the single best predictor of subsequent stroke was based on the patients qualifying event. In this study, patients with stroke as their qualifying event were found to be at an increased risk of subsequent stroke by approximately 55% as compared to patients whose qualifying events was TIA. These findings are intuitive since stroke in known to be more common in the older population and in people who have already suffered a stroke (Johnston *et al.* 2003; Lee *et al.* 2004; Modrego *et al.* 2004). Although these predictors were also found in the full Bayesian models, the 95% credibility intervals suggested that the treatment effects were not significant in these models.

Four of the trials randomised patients to more than two treatment groups (Guiraud-Chaumeil *et al.* 1982; Bousser *et al.* 1983; Caneschi *et al.* 1985; Diener *et al.* 1996); and having data at the IPD level enables contrasts to be specified for the type of treatment received. In the summary data models, separate analyses were conducted where the dual treatment of aspirin and dipyridamole were used repeatedly for each analysis. In this systematic review, since it was meaningless to combined the comparisons into an overall comparison group, the summary data analysis methods used were inefficient and assumed independence of the dual treatment group, which was used three times for the pair wise comparisons.

7.14.4 Discussion of the results

This systematic review provides substantial evidence for the efficacy of dual treatment of aspirin and dipyridamole given to patients with recent cerebrovascular events. The aim of this systematic review was to estimate the magnitude of the efficacy of the dual treatment and also to identify which patients should be targeted for the dual treatment.

7.14.4.1 Impact of dual treatment of aspirin and dipyridamole on subsequent stroke

This meta-analysis of dipyridamole in patients with prior cerebrovascular events shows that a dual treatment of aspirin and dipyridamole is effective in reducing the risk of subsequent stroke. The risk of subsequent stroke was reduced with the combination of aspirin and dipyridamole as compared to aspirin alone (22%), or dipyridamole alone (28%). The combination of aspirin and dipyridamole gave twice the reduction (40%). These findings compare well with those of previous meta-analyses using aggregate data (Lowenthal *et al.* 1994; Diener 1998; Tijssen 1998).

7.14.4.2 Impact of dual treatment of aspirin and dipyridamole on other outcomes

A potential criticism of ESPS II and some other large trials such as Perindopril Protection Against Recurrent Stroke Study (PROGRESS Collaborative Group 2001) is their use of subsequent stroke, rather than the composite of non-fatal stroke, non fatal myocardial infarction,

and vascular death, as the primary outcome. Importantly, the dual treatment of aspirin and dipyridamole was found to reduce the risk of the composite outcome as compared with dipyridamole alone, as seen in the ATT (Antithrombotic Trialists Collaboration 2002); and as compared to aspirin alone or placebo/control.

Dipyridamole in combination with aspirin did not alter the rate of myocardial infarction in patients with previous cerebrovascular events when compared against aspirin alone. In contrast, the addition of aspirin reduced the risk of myocardial infarction non-significantly by 30% (AD vs. D), a findings that is compatible with the ATT findings for aspirin (Antithrombotic Trialists Collaboration 2002).

The dual treatment of aspirin and dipyridamole did not alter the risk of vascular death when compared to aspirin alone or dipyridamole alone. This finding is probably related to a lack of power due to small number of events being recorded in the individual trials, since no effect was also seen when the dual treatment was compared to placebo/control.

7.14.4.3 Differences between ESPS II and the other trials

The ESPS II trial may have found that the dual treatment of aspirin and dipyridamole significantly reduces the risk of vascular events for several reasons (Diener *et al.* 1996). ESPS II provided 57% of the total data in this review and was positive in its outcome; it is possible that this trial is the primary driver for the findings reported here. However, when the data for ESPS II was excluded from the analyses,

the assessment of the dual treatment of aspirin and dipyridamole versus control remained positive whilst the comparison against aspirin became non-significant although the point estimates for the dual treatment support the efficacy of the dual agents. Earlier trials individually failed to find a positive effect of dipyridamole in stroke is unsurprising since they were all much smaller with lower statistical *power (type II error).*

Another reason may be related to the dose of aspirin, since the additive effect of dipyridamole on aspirin would be relatively smaller for studies which had higher doses of aspirin than compared to the additive effect of dipyridamole on lower doses of aspirin, as used in the ESPS II trial.

7.14.4.4 Impact of trial and patient characteristics

Subgroup analyses indicated that trial characteristics such as the method of formulation for dipyridamole and the dose of aspirin appeared to have little effect on the overall conclusions. The dose of aspirin used in the ESPS II trial has been previously criticised where the high stroke rates seen in the aspirin only group were thought to be related to the low level of dose (Davis *et al.* 1998). However, the lack of heterogeneity between the trial estimates and the results from the subgroup analyses suggested that the efficacy of the dual treatment was unrelated to the dose of aspirin.

From individual patient data models, patients with stroke as their qualifying event were found to be more likely to suffer a subsequent stroke as compared to those with TIA; also, increasing age was found

to be an important predictor of subsequent stroke. However; there was no evidence that the dual treatment of aspirin and dipyridamole was not more beneficial in these patients. A lack of interaction may be related to suboptimal levels of power from the small number of trials included in the meta-analysis.

7.14.4.5 Generalisability of the findings

The findings contrast with the neutral results for dipyridamole in systematic reviews which included trials involving groups of patients other than just stroke, e.g. those with myocardial infarction (Antithrombotic Trialists Collaboration 2002; De Schryver *et al.* 2003), a situation which is unsurprising. First, the epidemiology of stroke and ischaemic heart disease are different; stroke patients are older and more likely to be female, also stroke has a stronger relationship with the risk factor hypertension, than seen in patients with myocardial infarction. Second, ischaemic stroke is of mixed cause (large artery disease, cardioembolic, and small vessel disease), where as myocardial infarction largely follows coronary artery plaque ruptures and thrombosis. Third, the main risk after stroke is of having a subsequent stroke, as seen in this analysis, whilst patients with a myocardial infarction are more likely to have a further cardiac event. Finally, trials of primary prevention have consistently shown a differential treatment effect so that reducing BP is more effective in reducing stroke than myocardial infarction: 40% versus 15% reduction for a 10/6 mm Hg reduction in BP.

Considerable discussion, largely based on the results of ESPS II, has focussed on whether dipyridamole has selective effects on stroke. These possible differential effects on vascular events have also been observed for antihypertensive agents, e.g. calcium channel blockers may reduce stroke more than myocardial infarction whilst angiotensin converting enzyme inhibitors appear to have the opposite effect (Blood Pressure Lowering Treatment Trialists' Collaboration 2000).

The neutral rather than negative results for oral dipyridamole on myocardial infarction, as reported specifically in ESPS II (Diener *et al.* 1996), is reassuring in the view of this perception based on the use of intravenous dipyridamole in cardiac stress testing (Pfisterer 1992), that it might cause myocardial infarction. However, if one assumes that any platelet agent should have an affect on platelet mediated diseases wherever in the body it occurs, as assumed for aspirin and for clopidogrel; then the neutral result for dipyridamole on myocardial infarction found in ESPS II and this review can be explained in two ways. First, if we assume the above is true then as suggested previously the results from these studies focussed on stroke and so there were few cardiac events and so the failure to see such a reduction may be an issue of power. Or alternatively, if there is the possibility that there is a true discrepancy, and somehow dipyridamole stops cerebral platelet mediated events but not cardiac ones, then perhaps dipyridamole has a cardiotoxic effect that cancels out its platelet effect (Gladman Personal communication). This cardiotoxic effect could explain why many people can not tolerate

dipyridamole since the side effects they have nothing to do with the antiplatelet effect and may have something to do with the cardiotoxic effect.

Systematic reviews are susceptible to missing unpublished trials or those which are published in non-English journals, so-called publication bias. Many such studies will be neutral or negative in outcome and reviews will then have positively biased results. We performed a comprehensive multilingual search strategy, utilised the publication lists of existing trials and reviews, and contacted the pharmaceutical company which manufactures dipyridamole to help identify relevant trials. Additionally, no statistical evidence of publication bias was seen for any of the methods, hence the trials identified in this review probably represent the totality of trial evidence relating to dipyridamole in patients with cerebrovascular disease, and it is unlikely that the results are biased by a failure to include relevant studies.

7.14.5 Practical implications

Dipyridamole in combination with aspirin reduces recurrence in patients with prior cerebrovascular events. The data are internally consistent between the trials and are broadly relevant to stroke patients external to the trials. Hence, this treatment has a place in secondary prevention after stroke or TIA, as recommended in current national and international guidelines (Bogousslavsky *et al.* 2000; The Intercollegiate Working Party for Stroke 2002).

In depth analyses were performed in this review, however it appears that all patients with recent cerebrovascular events have equal benefit of the dual treatment, and no specific subgroups of patients where identified as effect modifiers. However, analyses did indicate that age was an important predictor of subsequent stroke and patients which suffered stroke as their qualifying cerebrovascular event had significantly higher risks of subsequent stroke at the end of trial. Surprisingly, although it is widely acknowledged that females are more likely to suffer an initial stroke; it appears from these analyses that females are not at a higher risk of subsequent stroke than males; however, the reasons for this finding remain unclear. However, the factors assessed in this meta-analysis were done so as part of post hoc analyses and there are open to bias.

The results from this systematic review has important implications to service providers who need to ensure that patients which have recently suffered a cerebrovascular disease are offered the dual treatment of aspirin and dipyridamole. However, the administration of dipyridamole with aspirin will depend on patient-specific factors such as underlying risk, experience on existing antiplatelet drugs, and tolerance or allergies to each of the drugs. Since further analysis of the data revealed that patients were more likely to drop out of the individual trials or have significant headaches develop if they received dipyridamole (with or without aspirin) as compared to aspirin alone (p <0.001) or control/placebo (p <0.001). In contrast, the bleeding

rates were highest with aspirin treatment (with or without dipyridamole) (p <0.001) (Leonardi-Bee *et al.* 2005).

From a public health perspective, clinicians are likely to use this combination treatment is it is deemed not only effective but also cost-effective; however it is unsure at the present whether the cost of *providing such treatments as specified in this review could* be possibly met since although the cost of aspirin is minimal; the additional cost of dipyridamole may be difficult to meet; since minimal studies have been performed assessing this.

7.14.6 Future research

This review attempted to combine all of the available evidence and to estimate the benefit of a combined treatment of aspirin and dipyridamole in the secondary prevention of stroke/TIA. However, it has not been able to fully explore which patients benefit the most from the combination treatment; this was primarily due to a lack of specific data being consistently shared across the trials. Also, if any significant effect modifiers had been found from this meta-analysis, the findings should be replicated in a subsequent sufficiently powered randomised controlled trial.

Also, although this review has highlighted the combination of dipyridamole and aspirin is effective; the cost effectiveness of the combination treatment should be fully assessed. To do such a study would require health data and cost data, or data which generate costs, such as survival, stroke and disability. This type of analysis

could not be performed in this review due to a lack of data for specific cost generating outcomes for each of the individual trials.

CHAPTER 8
GENERAL DISCUSSION

8.1 Introduction

This thesis has described methods for conducting meta-analyses based on summary and individual patient data and has exemplified these methods using three meta-analyses of randomised controlled trials in stroke. The findings from these meta-analyses have been discussed in detail in the individual chapters.

This chapter will be presented in five sections which relates to a summary of the advantages of meta-analyses, summary of the main limitations of meta-analyses, practical implications and recommendations for researchers, suggestions for future research, and conclusions.

8.2 Summary of the advantages of meta-analyses

Clinicians are repeatedly faced with questions of how best to treat patients; answers to these questions should be based on evidence based practice however, this evidence can provide advice that is diverse or conflicting. This diversity has lead to methodology being devised which attempts to combine the findings and produce generalisable conclusions. Meta-analyses based on systematic reviews are the most common method that is used since they have the advantage that they have increased power to perform analyses which can yield estimates for pooled intervention effects which are more generalisable, and therefore help clinicians and researchers to make more informed choices.

Meta-analyses based on individual patient data have been described as the gold standard of systematic reviews and are quoted as *'a*

yardstick against which other forms of systematic reviews could be measured' (Chalmers *et al.* 1993). This is because they remove some of the problems associated with meta-analyses based solely on data extracted from the literature, by allowing for detailed data checking and having access to data which was not originally published. However, they require considerable time and effort in contacting original trialists, merging of individual patient data into a single database, and often involve complex analyses (Clarke *et al.* 1994). At present there appears to be little evidence that the gains from performing a meta-analysis based on individual patient data are worthwhile and justified (Sutton et al. 1999). Since the late eighties, much collaboration has been achieved between trialists and much data has been shared to try to answer pressing questions that could not be answered alone from using summary data.

So far very few individual patient data meta-analyses have been published in the research area of stroke and mostly centre in the medical research area of cancer and assess the relationship between treatment and time to event data. To date, within the research area of secondary prevention of stroke, apart from the meta-analyses conducted in this thesis only two other articles have been published which have performed a meta-analysis using individual patient data (Chen *et al.* 2000; Cornu *et al.* 2000). These meta-analyses were used to predict risks for outcomes and to allow for more in-depth subgroup analyses to be performed.

The findings from the meta-analyses considered in this thesis were very informative with regards to efficacy of specific interventions and highlighted which subgroups of patients should be targeted, and therefore allowed for a more balanced interpretation and wider endorsement of the available evidence in these three areas of controversy.

Contacting the collaborators from the original trials enabled a fuller exploration of heterogeneity between the trials to be performed using both summary and individual patient data methods. This wouldn't have been possible if the data had been solely extracted from the original publications, as a considerable amount of the data were not presented in ways in which it could have been extracted.

The findings from the meta-analyses considered in this thesis provide an insight into the gains that can be had through collaborations with other investigators and from the sharing of individual patient data. Therefore, these findings need to be widely disseminated to allow these revealing results to be available to clinicians, researchers and the wider stroke community so that it may inform them and if appropriate, change their clinical practice in line with these findings.

8.3 Summary of the main limitations of meta-analyses

Systematic reviews and meta-analyses have distinct disadvantages associated with them (see Chapter 1); a summary of the main limitations which were apparent in the three meta-analyses from within this thesis are discussed.

8.3.1 Publication bias

Evidence of publication bias is a common finding in meta-analyses of randomised controlled trials (Oxman *et al.* 1995), where a relatively large number of studies are combined; as seen in the Blood pressure in Acute Stroke Collaboration meta-analysis from this thesis. However publication bias is less likely to be found when a small number of studies are included in the meta-analysis due to a lack of power. Although an assessment of publication bias should always be made within a meta-analysis, evidence of publication bias may be seen due to other factors, such as methodological quality of the trials. Therefore, efforts should be taken to assess how publication bias could impact on the findings from the meta-analysis, either through performing a sensitivity analysis where poor quality studies are excluded, or by exploring reasons for publication bias through using subgroup or regression analyses.

8.3.2 Extent of shared individual patient data

All of the meta-analyses considered in this thesis suffered from problems due to the inability to get individual patient data for all of the trials. Therefore, data from the publications of the trials had to be sought and extracted. In the majority of cases the data not shared were from older studies which had been discarded previously. However, this limited the amount of exploratory analyses that could be performed.

Additionally, missing data at study level can have a high impact of the results of the meta-analysis by introducing significant biases in the

results, especially when exploring heterogeneity since these studies are then excluded from the analyses. Although the problems associated with data missing at study level are unique to meta-analyses (Sutton *et al.* 1998), little research has been performed to date.

Within the Blood pressure in Acute Stroke Collaboration meta-analysis, due to a large number of studies not having data at the individual level, where possible the covariates in the adjusted individual patient data models for these studies were included using the study level value for each patient. Other researchers have suggested that where data is missing one can assume that the covariate is balanced between the two treatment groups, hence the difference between groups is set at zero (Higgins *et al.* 2001). However, this assumption is probably invalid in most cases, and would be hard to verify. Alternatively, a data index variable could be created where the covariate takes a value of 1 if it is observed and a zero if it is missing (Pigott *et al.* 1994). This would allow the researcher to assess whether this index variable is correlated with other covariates where there is no missing data. If a correlation is observed, then it appears that there some evidence that the reasons for the missing data could be dependent on the other covariates with no missing data (Pigott *et al.* 1994). These two methods thought could only be considered where there is a minority of studies with missing data.

8.3.3 Misleading meta-analyses

Although each of the individual patient data meta-analyses considered in this thesis were conducted with a high methodological rigour, there is still the potential for the results from these meta-analyses to be misleading and not replicated in a mega-sized trial. Misleading results from meta-analyses have been systematically studied and found to only have fair agreement with large randomised controlled trials (Ioannidis *et al.* 1998) this was also observed in this thesis in both the occupational therapy and dipyridamole meta-analyses. A point worth mentioning is that meta-analyses should not be used as a solution instead of performing a large well designed randomised controlled trial since the results of a meta-analysis do not show how to treat individuals (Lau *et al.* 1998). Therefore, it is essential that large randomised controlled trials investigating any subsequent specific medical areas of debate identified from meta-analyses are conducted and explored.

8.3.4 Fixed effect or random effect model?

At present, there does not appear to be a clear and simple answer to whether a fixed effect or random effect model should be used to combine the results from the individual trials. It has been argued that it is not the significance of the pooled p value that is important, but the estimates and their standard errors for the intervention effect (Sutton *et al.* 1998). It is clearly important that whichever method is used, a thorough analysis is employed to assess the impact of any heterogeneity between the studies, and to attempt to explore any

heterogeneity through the use of subgroup analysis or regression analyses. In the meta-analyses considered in this thesis, little differences were seen in the treatment estimates between the fixed effect and random effect models, which is probably due to low levels of heterogeneity being observed in these examples.

Since meta-analysis is a tool that is used to combine studies which are thought to be comparable; it is only sensible to combine them if they are not too heterogeneous. A random effect model will allow for certain degrees of heterogeneity, however, there will come a point where the heterogeneity is so large that the random effect model will become inadequate, and the meta-analysis needs to be abandoned. However, it is unclear at present what the minimum level of comparability is before a judgement is made that the studies are too heterogeneous and therefore inappropriate to be combined in a meta-analysis (Sutton *et al.* 1998).

8.4 Practical implications and recommendations for researchers

8.4.1 Involving the original trialists

Researchers embarking on a systematic review and meta-analysis should ensure they perform a comprehensive search strategy to minimise the risk of not identifying all eligible studies. We contacted all of the original contact investigators from the trials to ask them to share their individual patient data with the specific collaboration. We found the investigators were more willing to share their data with the collaboration when they received reassurance of authorship status for

their contributions in providing data and editing of the draft manuscript of the systematic review.

8.4.2 Summary or individual patient data methodologies?

Stewart and Clarke have looked at the differences between meta-analysis using literature data and individual patient data (Stewart et al. 1995). They conclude that differences in the effect estimates of the intervention may be seen when comparing the two methods. The combination of unpublished trials, excluded patients, short follow-ups, and fixed time-point analysis are thought to contribute to an over estimation of the effects in a meta-analysis of literature data. Therefore the authors concluded that a meta-analysis based on individual patient data gave the least biased result (Stewart et al. 1995).

Both summary and individual patient data methodologies were considered in this thesis and taking into account the findings of Stewart and Clarke, the data used from each of the trials was primarily based on individual patient data, and summarised where appropriate. No clinically important differences for treatment effects were seen between the two methods; therefore it is recommended that researchers which are less confident with using more complex analysis packages, such as SAS for Windows or MLWin, should use random effect summary data methods to perform their analyses.

8.4.3 Quantifying and exploring heterogeneity

Within the thesis, methods for the testing of heterogeneity and a range of the most common estimation methods of estimating heterogeneity were considered. All of the estimation methods appeared to perform well when used in these three data sets; however, testing for heterogeneity seemed to be less important and the question of how much heterogeneity exists within the meta-analysis appeared more important (Higgins *et al.* 2002).

It also seems essential to explore the reasons behind heterogeneity using appropriate analyses as described in this thesis; however, it is recommended that a random effect model is used to do this, especially where heterogeneity is present. This is to ensure that the importance of the explored covariate is not overtly emphasised as would be seen from using a fixed effect model (Thompson *et al.* 1999).

A lower power has been associated with using meta-regression techniques are compared to using an individual patient data analysis (Lambert *et al.* 2002). They also found that the estimates for the patient level covariates rarely agreed between the methods as confirmed in this thesis. Therefore, it is recommended that a meta-analysis based on individual patient data is needed when the meta-analysis attempts to see how patient level covariates are related to the intervention. Lambert and colleagues stress that meta-regression is not a biased method and if an effect is detected then it is probably a large and important one. However, caution is needed since within

the Blood pressure in Acute Stroke Collaboration meta-analysis, systolic BP at baseline was found to be an important predictor of the relative estimate of the odds ratio for the treatment effect using meta-regression methods, but not found to be an important effect modifier when an individual patient data model was considered.

8.5 Suggestions for future research

Although we were generally successful in obtaining the individual patient data from the majority of the studies included within two of the meta-analyses, problems were encountered with the Blood pressure in Acute Stroke Collaboration meta-analysis where a third of the studies did not have covariate data at the individual level and so the average trial level value was used for those particular trials in the individual patient data models. More research is needed in developing methodologies for combining summary and individual patient data covariates to allow for this lack of individual patient data from original trials.

Within this thesis, some evidence of publication bias was seen in one of the meta-analyses; as yet, there has been little research into methods which adjust for publication bias during the analysis stage (Thompson *et al.* 1999). Also, it has been suggested that since publication bias can lead to heterogeneity between the trials within a meta-analysis (Sutton *et al.* 1998), then further research is needed into assessing the impact of publication bias on quantifying and exploring heterogeneity.

This thesis has focussed on whether there are clinically important differences between the heterogeneity assessment methods; however even though no clinically important differences were seen in these meta-analyses, there may be statistically important differences which could be evaluated using a series of simulation studies.

8.6 Conclusions

To conclude, collaborations within the area of stroke medicine can be successful and much data can be shared. The findings from meta-analyses, if conducted with high methodological rigour, can be informative about the effectiveness of particular treatments and about which patients should be targeted for treatment. The findings from a well conducted review can also help steer the direction of future trials.

Summary data meta-analyses are practically easier and can be very rewarding; however, assessments and explorations of heterogeneity should always be made. Meta-analyses based on individual patient data may be needed to allow for more in depth investigations of heterogeneity, especially of patient characteristics. However, they themselves are not the panacea to all difficulties since they are subject to particular problems, mainly related to obtaining individual patient data to enable in depth analyses.

REFERENCES

Acheson, J., Danta, G. and Hutchinson, E. C. (1969). "Controlled trial of dipyridamole in cerebral vascular disease." British Medical Journal 1: 614-615.

Aho, K., Harmsen, S. and Marrquardsen, J. (1980). "Cerebrovascular disease in the community: results of a WHO collaborative study." Bulletin of the World Health Organisation 58: 11-30.

Allen, L. R. and Beattie, R. J. (1984). "The role of leisure as an indicator of overall satisfaction with community life." Journal of Leisure Research 2: 99-109.

Antithrombotic Trialists Collaboration (2002). "Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients." BMJ 324(7329): 71-86.

Autret, A., Lehert, P. and Monsnier, M. (1992). "Study of Naftidrofuryl in Cerebral Ischaemic Accidents." Psychologie Medicale 24(1): 103-114.

Azcona, A. and Lataste, X. (1990). "Isradipine in Patients with Acute Ischaemic Cerebral Infarction: An Overview of the ASCLEPOIS Programme." Drugs 40(suppl. 2): 52-57.

Bamford, J., Dennis, M. S., Sandercock, P., Burn, J. and Warlow, C. (1990). "The frequency, cause and timing of death within 30 days of a first stroke: the Oxfordshire Community Stroke Project." Journal of Neurology Neurosurgery and Psychiatry 53: 824-9.

Bamford, J., Sandercock, P., Dennis, M., Burn, J. and Warlow, C. (1991). "Classification and natural history of clinically identifiable subtypes of cerebral infarction." Lancet 337: 1521-6.

Barer, D. H., Cruickshank, J. M., Ebrahim, S. B. and Mitchell, J. R. A. (1988). "Low Dose beta blockade in acute stroke ("BEST" trial): An evaluation." British Medical Journal 296: 737-741.

Barer, D. H., Cruickshank, J. M., Ebrahim, S. B. and Mitchell, J. R. A. (Unpublished). "Low dose beta blockade in acute stroke ("BEST" trial): An Evaluation (Pilot study)."

Bath, F. J. and Bath, P. M. W. (1997). "What is the correct management of blood pressure in acute stroke? The Blood pressure in Acute Stroke Collaboration." Cerebrovascular Diseases 7: 205-213.

Bath, F. J., Owen, V. E. and Bath, P. M. W. (1998). "Quality of full and final publications reporting acute stroke trials." Stroke 29: 2203-2210.

Bath, P. M. (2001). "Efficacy of Nitric Oxide in Stroke (ENOS) trial." Major Ongoing Stroke Trials, Stroke 32(10): 2450-2451 (abstract).

Bath, P. M. W., Pathansali, R., Iddenden, R. and Bath, F. J. (2001). "The effect of transdermal glyceryl trinitrate, a nitric oxide

donor, on blood pressure and platelet function in acute stroke." Cerebrovascular Diseases 11(3): 265-272.
Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K. F., Simel, D. and Stroup, D. F. (1996). "Improving the quality of reporting of randomized controlled trials: the {CONSORT} statement." Journal of the American Medical Association 276(8): 637-639.
Begg, C. B., Cooper, H. and Hedges, L. V. (1994). Publication bias. New York, Russel Sage Foundation.
Begg, C. B. and Mazumdar, M. (1994). "Operating characteristics of a rank correlation test for publication bias." Biometrics 50(4): 1088-1101.
Berkey, C. S., Hoaglin, D. C., Mosteller, F. and Colditz, G. A. (1995). "A random effects regression model for meta-analysis." Statistics in Medicine 14(4): 395-411.
Biggerstaff, B. J. and Tweedie, R. L. (1997). "Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis." Statistics in Medicine 16: 753-768.
Birge, R. T. (1932). "The calculation of error by the method of least squares." Physical Review 40: 1-32.
Blood pressure in Acute Stroke Collaboration (2003). "Blood pressure in Acute Stroke Collaboration (BASC)." Major Ongoing Stroke Trials, Stroke 34: 196e-207.
Blood pressure in Acute Stroke Collaboration (BASC) (2001(b)). "Interventions for deliberately altering blood pressure in acute stroke (Cochrane Review)." The Cochrane Library Issue 4.
Blood Pressure Lowering Treatment Trialists' Collaboration (2000). "Effects of ACE inhibitors, calcium antagonists, and other blood-pressure-lowering drugs; results of prospectively designed overviews of randomised trials." The Lancet 356(December 9th): 1955-1964.
Bogousslavsky, J., Kaste, M., Olsen, T. S., Hacke, W. and Orgogozo, J.-M. (2000). "Risk factors and stroke prevention." Cerebrovascular Diseases 10 suppl 3: 12-21.
Bogousslavsky, J., Regli, F., Zumstein, V. and Kobberling, W. (1990). "Double-blind study of nimodipine in non-severe stroke." Eur Neuro 30: 23-26.
Bogousslavsky, J., Victor, S. J., Salinas, E. O., Pallay, A., Donnan, G. A., Fieschi, C., Kaste, M., Orgogozo, J. M., Chamorro, A., Desmet, A. and for the European-Australian Fiblast (Trafermin) in Acute Stroke Group (2002). "Fiblast (Trafermin) in Acute Stroke: Results of the European-Australian Phase II/III Safety and Efficacy Trial." Cerebrovascular Diseases 14: 239-251.
Bonita (1992). "Epidemiology of stroke." Lancet 339: 342-44.
Boreas, A. M. H. P., Lodder, J., Kessels, F., de Leeuw, P. W. and Troost, J. (2001). "Predictors of poststroke blood pressure level and course." Journal Stroke Cerebrovascular Diseases 10: 85-91.
Bousser, M. G., Eschwege, E. and Haguenau, M. (1983). ""AICLA" controlled trial of aspirin and dipyridamole in the secondary

prevention of arthero-thrombotic cerebral ischaemia." Stroke 14: 5-14.

Britton, M., Carlsson, A. and de Faire, U. (1986). "Blood pressure course in patients with acute stroke and matched controls." Stroke 17: 861-864.

Britton, M., Carlsson, A. and de Faire, U. (1986). "Blood pressure in patients with acute stroke and matched controls." Stroke 17: 861-864.

Brown, H. K. and Kempton, R. A. (1994). "The application of REML in clinical trials." Statistics in Medicine 16: 1601-1617.

Caneschi, S., Bonaventi, C. and Finzi, F. (1985). "La malattia cerebrovascolare ischemica: il suo trattamento con differenti farmaci antiaggreganti piastrinici. Follow-up clinico di 80 pazienti (22-34 mesi) [The ischaemic cerebrovascular disease: its treatment with different antiaggregate drugs. Clinical follow-up of 80 patients (22-34 months)]." Minerva Medica 76: 1933-1943.

Caneschi, S., Bonventi, C., Finzi, F. and Tito, P. (1985). "Il Ditazolo nella terapia delle vasculopatie cerebrali ischemiche. Risultati clinici ed esperienze di laboratorio (follow-up clinico 22-34 mesi)." Minerva Medica 76: 1883-1894.

CAPRIE Steering committee (1996). "A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE)." Lancet 16(348(9038)): 1329-39.

Carlberg, B., Asplund, K. and Hagg, E. (1991). "Factors influecing admission blood pressure levels in patients with acute stroke." Stroke 22: 527-530.

Carlberg, B., Asplund, K. and Hagg, E. (1991). "Factors influencing admission blood pressure levels in patients with acute stroke." Stroke 4: 527-530.

Carlberg, B., Asplund, K. and Hagg, E. (1993). "The prognostic value of admission blood pressure in patients with acute stroke." Stroke 24: 1372-1375.

Carlin, J. B. (1992). "Meta-analysis for 2x2 tables: a Bayesian approach." Statistics in Medicine 11: 141-158.

Chalmers, I., Sandercock, P. and Wennberg, J. (1993). "The Cochrane Collaboration: Preparing, maintaining, and disseminating systematic reviews of the effects of health care." Annals of the New York Academy of Sciences 703: 156-165.

Chalmers, T. C., Smith, J., J, Blackburn, B., Silverman, B., Schroeder, B., Reitman, D. and Ambroz, A. (1981). "A method for assessing the quality of a randomised control trial." Controlled Clinical Trials 2: 31-49.

Chen, Z. M., Sandercock, P., Pan, H. C., Counsell, C., Collins, R., Liu, L. S., Xie, J. X., Warlow, C. and Peto, R. (2000). "Indications for early aspirin use in acute ischemic stroke - A combined analysis of 40 000 randomized patients from the Chinese Acute Stroke Trial and the International Stroke Trial." Stroke 31(6): 1240-1249.

Clarke, M. J. and Stewart, L. A. (1994). "Systematic Reviews - Obtaining Data from Randomized Controlled Trials - How Much Do We Need for Reliable and Informative Metaanalyses." British Medical Journal 309(6960): 1007-1010.

Cochran, W. G. (1937). "Problems arising in the analysis of a series of similar experiments." Journal of the Royal Statistical Society 4: 102-118.

Cochran, W. G. (1954). "The combination of estimates from different experiments." Biometrics 10: 101-129.

Cornu, C., Boutitie, F., Candelise, L., Boissel, J. P., Donnan, G. A., Hommel, M., Jaillard, A. and Lees, K. R. (2000). "Streptokinase in acute ischemic stroke: An individual patient data meta-analysis - The Thrombolysis in Acute Stroke Pooling Project." Stroke 31(7): 1555-1560.

Corr, S. and Bayer, A. (1995). "Occupational therapy for stroke patients after hospital discharge - a randomized controlled trial." Clinical Rehabilitation 9: 291-296.

COSSACS (2003). "Continue Or Stop post-stroke Antihypertensives Collborative Study (COSSACS)." Major Ongoing Stroke Trials, Stroke 34: 196e-207.

Dandapani, B., Suzuki, S., Kelley, R. E., Reyes-Iglesias, Y. and Duncan, R. C. (1995). "Relation between blood pressure and outcome in intracerebral hemorrhage." Stroke 26: 21-24.

Davis, S. M. and Donnan, G. A. (1998). "Secondary prevention for stroke after CAPRIE and ESPS-2 - opinion 1." Cerebrovascular Diseases 8: 73-75.

De Deyn, P. P., De Reuck, J., Deberdt, W., Vlietinck, R. and Orgogozo, J. M. (1997). "Treatment of acute ischaemic stroke with piracetam." Stroke 28(12): 2347-2352.

De Schryver, E. L. (2000). "Design of ESPRIT: an international randomized trial for secondary prevention after non-disabling cerebral ischaemia of arterial origin. European/Australian Stroke Prevention in Reversible Ischaemia Trial (ESPRIT) group." Cerebrovasc Dis 10(2): 147-50.

De Schryver, E. L. L. M., Algra, A. and van Gijn, J. (2003). "Cochrane review: dipyridamole for preventing major vascular events in patients with vascular disease." Stroke 34: in press.

Dean, C. M. and Shepherd, R. B. (1997). "Task-related training improves performance of seated reading tasks after stroke. A randomised controlled trial." Stroke 28: 722-728.

Deeks, J., Glanville, J. and Sheldon, T. (1996). Undertaking systematic reviews of effectiveness: CRD guidelines for those carrying out or commissioning reviews. York, York Publishing Services Ltd.

Deeks, J. J., Altman, D. G. and Bradburn, M. J. (2001). Statistical methods for examining heterogeneity and combining results from several studies in meta-analyses. Systematic reviews in Health Care: Meta-analysis in Context. D. G. Altman. London, BMJ Publication Group.

Dennis, M. S. (1988) Transient ischaemic attacks in the community. London, University of London. Doctor of Medicine Thesis.
Department of Health (2001). National Service Framework for the Elderly, Department of Health.
Department of Health (2003). National Service framework for older people, Department of Health.
DerSimonian, R. and Laird, N. (1986). "Meta-analysis in clinical trials." Controlled Clinical Trials 7: 177-188.
Diener, H. C. (1998). "Dipyridamole trials in stroke prevention." Neurology Suppl 3: S17-S19.
Diener, H. C., Cunha, L., Forbes, C., Sivenius, J., Smets, P. and Lowenthal, A. (1996). "European Stroke Prevention Study 2. Dipyridamole and acetylsalicylic acid in the secondary prevention of stroke." J Neurological Sciences 143: 1-13.
Diener, H. C., Cunha, L., Sivenius, J., Smets, P. and Lowenthal, A. (1996). "European Stroke Prevention Study 2. Dipyridamole and acetylsalicylic acid in the secondary prevention of stroke." Journal of Neurological Sciences 143: 1-13.
Drummond, A. (1990). "Leisure activity after stroke." International Disability Studies 12: 157-160.
Drummond, A. and Walker, M. (1995). "A randomised controlled trial of leisure rehabilitation after stroke." Clinical Rehabilitation 9: 283-290.
Drummond, A. E. R., Parker, C. J., Gladman, J. R. F. and Logan, P. A. (2001). "Development and validation of the Nottingham Leisure Questionnaire (NLQ)." Clinical Rehabilitation 15: 647-656.
Drummond, A. E. R. and Walker, M. F. (1994). "The Nottingham Leisure Questionnaire for Stroke Patients." British Journal of Occupational Therapy 57(11): 414-418.
Duncan, P. W., Jorgensen, H. S. and Wade, D. T. (2000). "Outcome measures in acute stroke trials." Stroke 31(6): 1429-1438.
Duval, S. and Tweedie, R. (2000). "A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis." Journal of the American Statistical Association 95(449): 89-99.
Dyker, A. G., Grosset, D. G. and Lees, K. R. (1997). "Perindopril Reduced Blood Pressure but Not Cerebral Blood Flow in Patients with Recent Cerebral Ischemic Stroke." Stroke 28: 580-583.
Easterbrook, P. J., Berlin, J. A., Gopalan, R. and Matthews, D. R. (1991). "Publication bias in clinical research." Lancet 337(8746): 867-872.
Ebrahim, S. (1985) Physical disability, health, depressed mood, and use of services after stroke. Nottingham, University of Nottingham. Doctor of Medicine Thesis.
Egger, M. (1998). "Under the meta-scope: possibilities and limits of meta-analyses." Schweizerische Medizinische Wochenschrift. Journal Suisse de Medeciné 128(48): 1893-1901.

Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997). "Bias in meta-analysis detected by a simple, graphical test." British Medical Journal 315: 629-634.

Emerson, J. D. (1994). "Combining estimates if the OR: the state of art (review)." Statistical Methods in Medical Research 3: 157-178.

Engels, E. A., Schmid, C. H., Terrin, N., Olkin, I. and Lau, J. (2000). "Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses." Statistics in Medicine 19(13): 1707-1728.

ESPS Group (1990). "European Stroke Prevention Study." Stroke 21: 1122-1130.

Feldmann, E. and Skolnick, B. E. (1999). "Cerebral haemodynamics, autoregulation, and blood pressure management." Journal of Stroke and Cerebrovascular Diseases 8(3): 176-182.

Fisher, L. D. (1996). "Comments on Bayesian and frequentist analysis and interpretation of clinical trials - comment." Controlled Clinical Trials(17): 423-434.

Fleiss, J. L. (1986). "Analysis of data from multiclinic trials." Controlled Clinical Trials 7: 267-275.

Fleiss, J. L. (1993). "The statistical basis of meta-analysis (review)." Statistical Methods in Medical Research 2: 121-45.

Fleiss, J. L., Cooper, H. and Hedges, L. V. (1994). Chapter 17: Measures of the Effect Size for Categorical Data. The handbook of research synthesis editors. L. V. Hedges. New York, Russell Sage Foundation: 245-260.

Gelber, D. R. and Goldhirsch, A. (1987). "The interpretation of results from sub-set analyses within overviews of randomised clinical trials." Statistics in Medicine 19: 1707-1728.

Gelman, A. and Rubin, D. B. (1992). "Inference from iterative simulation using multiple sequences." Statistical Sciences 7: 457-472.

Gelmers, H. J. (1988). "A controlled trial of nimodipine in acute ischemic stroke." New England Journal of Medicine 318: 203-207.

Geman, S. and Geman, D. (1984). "Stochastic relaxation, Gibb's distributions and the Bayesian restoration of images." IEEE Transactions on Pattern Analysis and Machine Intelligence 6: 721-741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. Bayesian Statistics 4. A. F. M. Smith. Oxford, UK, Clarendon Press.

Gilbertson, L., Langhorne, P., Walker, A., Allen, A. and Murray, G. D. (2000). "Domiciliary occupational therapy for patients with stroke discharged from hospital: randomised controlled trial." British Medical Journal 320(7235): 603-606.

Gladman, J. R. F. (Personal communication). Possible cardiotoxic effect of dipyrdiamole. Nottingham.

Gladman, J. R. F., Lincoln, N. B. and Adams, S. A. (1993). "Use of the extended ADL scale with stroke patients." Age and Ageing 22: 419-424.

Glass, G. V. (1976). "Primary, secondary and meta-analysis of research." Educational Research 5: 3-8.

Glasziou, P. P. and Sanders, S. L. (2002). "Investigating the causes of heterogeneity in systematic reviews." Statistics in Medicine 21: 1503-1511.

Goldberg, D. (1972). The detection of psychiatric illness by questionnaire. London, Oxford University Press.

Goldstein, H., Yang, M., Omar, R. Z., Turner, R. M. and Thompson, S. G. (2000). "Meta-analysis using multilevel models with an application to the study of class size effects." Applied Statistics-Journal of the Royal Statistical Society Series C 49(Part 3): 399-412.

Gompertz, P., Pound, P. and Ebrahim, S. (1993). "The reliability of stroke outcome measures." Clinical rehabilitation 7: 290-296.

Gompertz, P., Pound, P. and Ebrahim, S. (1994). "A postal version of Barthel Index." Clinical Rehabilitation 8: 233-239.

Gotzche, P. C. (1987). "Reference bias in reports of drug trials." British Medical Journal 295: 654-656.

Granger, C. V., Devis, L. S., Peters, M. C., Sherwood, C. C. and Barrett, J. E. (1979). "Stroke rehabilitation: analysis of repeated Bathel Index measures." Archives of Physical Medicine and Rehabilitation 60: 14-17.

Gray, C. S., French, J. M., Venebles, G. S., Cartlidge, N. E. F., James, O. F. W. and Bates, D. (1990). "A Randomized Double-Blind Controlled Trial of Naftidrofuryl in Acute Stroke." Age and Ageing 19: 356-363.

Greenland, S. (1987). "Quantitative methods in the review of the epidemiologic literature." Epidemiologic Reviews 9: 1-30.

Guiraud-Chaumeil, B., Rascol, A., David, J., Boneu, B., Clanet, M. and Bierme, R. (1982). "Prevention des recidives des acidents vasculaires cerebraux ischaemiques par les anti-agregants plaquettaires [Prevention of recurrence of cerebro-vascular accidents by anti-platelet agents]." Rev Neurol (Paris) 138: 367-385.

Hardy, R. J. and Thompson, S. G. (1996). "A likelihood approach to meta-analysis with random effects." Statistics in Medicine 15: 619-629.

Hardy, R. J. and Thompson, S. G. (1998). "Detecting and describing heterogeneity in meta-analysis." Statistics in Medicine 17(8): 841-856.

Harper, G., Castleden, C. M. and Potter, J. F. (1994). "Factors affecting changes in blood pressure after acute stroke." Stroke 25: 1726-1729.

Hasselblad, V. I. C. and McCrory, D. C. (1995). "Meta-analytic tools for medical decision making: a practical guide." Medical Decision Making 23: 1063-85.

Hatona, S. (1976). "Experiences from a multicenter stroke register: a preliminary report." Bull WHO 54: 541-53.
Hedges, L. V. and Olkin, I. (1985). Statistical methods for meta-analysis. London, Academic press.
Herrschaft, H. (1988). "Die wirksamkeit von piracetam bei akuten zerebralen ischamie des menschen [Effectiveness of piracetam in acute cerebral ischaemia in humans]." Med Klin 83: 667-677.
Hier, D. B. and Edelstein, G. (1991). "Deriving clinical prediction rules from stroke outcome research." Stroke 22: 1431-6.
Higgins, J. P. T. and Thompson, S. G. (2002). "Quantifying heterogeneity in a meta-analysis." Statistics in Medicine 21: 1539-1558.
Higgins, J. P. T. and Thompson, S. G. (2004). "Controlling the risk of suprious findings from meta-regression." Statistics in Medicine 23: 1663-1682.
Higgins, J. P. T., Thompson, S. G., Deeks, J. J. and Altman, D. G. (2003). "Measuring inconsistencies in meta-analysis." British Medical Journal 327: 557-560.
Higgins, J. P. T., Whitehead, A., Turner, R. M., Omar, R. Z. and Thompson, S. G. (2001). "Meta-analysis of continuous outcome data from individual patients." Statistics in Medicine 20: 2219-2241.
Hillen, T., Coshall, C., Tilling, K., Rudd, A. G., McGovern, R. and Wolfe, C. D. A. (2003). "Cause of stroke recurrence is multifactorial: patterns, risk factors, and outcomes of stroke recurrence in the South London Stroke Register." Stroke 34(6): 1457-1463.
Horn, J., Limburg, L. and Orgogozo, J. M. (2001). "Calcium antagonists for acute ischaemic stroke (Cochrane Review)." In: The Cochrane Library, Update Software: Oxford.
Huczynski, J., Krostka-Trabka, E., Sotowska, W., Bieron, K., Grodzinska, L., Dembinska-Kiec, A., Pykosz-Mazur, E., Peczak, E. and Gryglewski, R. J. (1985). "Double-Blind Controlled Trial of the Therapeutic Effects of Prostacyclin in Patients with Completed Ischaemic Stroke." Stroke 16(5): 810-814.
Huque, M. F. (1988). "Experiences with Meta-Analysis in NDA Submissions." Proceedings of the Biopharmaceutical Section of the American Statistical Association 2: 28-33.
International Society of Hypertension Writing Group (2003). "International Society of Hypertension (ISH): statement on the management of blood pressure in acute stroke." Journal of Hypertension 21: 665-672.
Ioannidis, J. P. A., Cappelleri, J. C. and Lau, J. (1998). "Issues in comparisons between meta-analyses and large trials." Jama-Journal of the American Medical Association 279(14): 1089-1093.
ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group (1995). "ISIS-4: a randomised factorial trial assessing

early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58050 patients with suspected acute myocardial infarction." Lancet 345: 669-685.

Jadad, A. R., Moore, R. A. and Carroll, D. (1996). "Assessing the quality of reports of randomized clinical trials: Is blinding necessary?" Controlled Clinical Trials 17: 1-12.

Johnston, S. C., Sidney, S., Bernstein, A. L. and Gress, D. R. (2003). "A comparison of risk factors for recurrent TIA and stroke in patients diagnosed with TIA." Neurology 60(2): 280-285.

Jongbloed, L. and Morgan, D. (1991). "An investigation of involvement in leisure activities after a stroke." The American Journal of Occupational Therapy 45(5): 420-427.

Juby, L. C., Lincoln, N. B. and Berman, N. (1996). "The effect of a stroke rehabilitation unit on functional and psychological outcome: a randomised controlled trial." Cerebrovascular Diseases 6: 106-110.

Kaste, M., Folgelholm, R., Erila, T., Palomaki, H., Murros, K., Rissanen, A. and Sarna, S. (1994). "A Randomized, Double-Blind, Placebo-Controlled Trial of Nimodipine in Acute Ischemic Hemispheric Stroke." Stroke 25: 1348-1353.

King's Fund Consensus Statement (1988). "Treatment of Stroke." British Medical Journal 297: 126-8.

Kjaergard, L. L. and Gluvd, C. (2001). "Reported methodological quality and discrepancies between large and small randomised trials in meta-analyses." Annals of Internal Medicine 135: 982-989.

Kramer, G., Tettenborn, B., Schmutzhard, E., Aichner, F., Schwartz, A., Busse, O., Hornig, C. and Ladurner, G. (1994). "Nimodipine in acute ischaemic stroke." Cerebrovascular Diseases 4(182-188).

Lambert, P. C., Sutton, A. J., Abrams, K. R. and Jones, D. R. (2002). "A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis." Journal of Clinical Epidemiology 55(1): 86-94.

Lamsudin, R., Misbach, J., Andradi, Petru Cahyadi, A., Hadirioto, S. and Sumaryanto, F. (1997). "A controlled trial of nimodipine in acute ischemic stroke. A multicentre Indonesian Neurological Association Study Group." Indonesian Journal of Clinical Epidemiology and Biostatistics 2(1).

Larholt, K. M., Tsiatis, A. A. and Gelber, R. D. (Unpublished). "Variability of coverage probabilities when applying a random effects methodology for meta-analysis."

Lau, J., Antman, E. M. and Jimenez-Silva, J. (1992). "Cumulative meta-analysis of therapeutic trials for myocardial infarction." New England Journal of Medicine 327: 248-254.

Lau, J., Ioannidis, J. P. A. and Schmid, C. H. (1998). "Summing up evidence: one answer is not always enough." Lancet 351: 123-127.

Lavin, P. (1986). "Management of hypertension in patients with acute stroke." Archives of Internal Medicine 146: 66-68.

Lawes, C. M. M., Bennett, D. A., Feigin, V. L. and Rodgers, A. (2004). "Blood pressure and Stroke. An overview of published reviews." Stroke 35: 776-785.

Lee, A. H., Somerford, P. J. and Yau, K. K. (2004). "Risk factors for ischaemic stroke recurrence after hospitalisation." Medical Journal of Australia 181(5): 244-246.

Lees, K. and Muir, K. W. *(Personal communication). Images Pilot.*

Lees, K. R., Bath, P. M. W. and Naylor, A. R. (2000). "ABC of arterial and venous disease - Secondary prevention of transient ischaemic attack and stroke." BMJ 320: 991-994.

Lees, K. R. and Muir, K. W. (2001). "Intravenous Magnesium Therapy Efficacy in Stroke Trial (IMAGES)." Major Ongoing Stroke Trials, Stroke 32: 1453 (abstract).

Leonardi-Bee, J. (2000) Meta-analysis of cholesterol lowering trials. Leicester, University of Leicester. MSc.

Leonardi-Bee, J., Bath, P. M. W., Bousser, M. G., Davalos, A., Diener, H. C., Guiraud-Chaumeil, B., Sivenius, J., Yatsu, F., Dewey, M. E. and on behalf of the Dipyridamole in Stroke Collaboration (DISC) (2005). "Dipyridamole for preventing recurrent ischaemic stroke and other vascular events. A meta-analysis of individual patient data from randomized controlled trials." Stroke 36(1): 162-168.

Leonardi-Bee, J., Bath, P. M. W., Phillips, S. J., Sandercock, P. A. G. and for the IST collaborative group (2002). "Blood pressure and clinical outcomes in the International Stroke Trial." Stroke 33: 1315-1320.

Li, Y., Shi, L. and Roth, H. D. (1994). "The bias of the commonly-used estimate of variance in meta-analysis." Communications in Statistics: Theory and Methods 23(4): 1063-1085.

Light, R. J. and Pillemar, D. B. (1984). Summing up: the science of reviewing research. *Cambridge, MA, Harvard University Press.*

Limburg, M. and Hijdra, A. (1990). "Flunarizine in Acute Ischaemic Stroke: A Pilot Study." Eur Neuro 30: 121-122.

Logan, P. A., Ahern, J., Gladman, J. R. F. and Lincoln, N. B. (1997). "A randomized controlled trial of enhanced Social Service occupational therapy for stroke patients." Clinical Rehabilitation 11(2): 107-113.

Lowe, D. G. O. and Forbes, C. D. (Unpublished). "Nimodipine in acute cerebral hemispheric infarction." Unpublished report BAY e 9736/0449.

Lowenthal, A. and Buyse, M. (1994). "Secondary prevention of stroke: does dipyridamole add to aspirin?" Acta Neurol.Belg. 94: 24-34.

Mahoney, F. I. and Barthel, D. W. (1965). "Functional evaluation: the Barthel Index." Md Med J 14: 61-65.

Man-Son, H. M. and Wells, G. (1995). "Meta-analysis of efficacy of quinine for treatment of nocturnal leg cramps in elderly people." British Medical Journal 310: 13-17.

Mantel, N. and Haenszel, W. (1959). "Statistical aspects of the analysis of data from retrospective studies of disease." Journal of National Cancer Institiute 22: 719-748.

Martin, J. F., Hamdy, N., Nicholl, J., Lewtas, N., Bergvall, U., Owen, P., Syder, D. and Holroyd, M. (1985). "Double-blind controlled trial of prostacyclin in cerebral infarction." Stroke 16: 386-390.

Martinez-Villa, E., Guillen, F., Villanueva, J. A., Matias-Guiu, J., Bigorra, J., Gil, P., Carbonell, A. and Martinez-Lage, J. M. (1990). "Placebo-controlled trial of nimodipine in the treatment of acute ischemic cererbal infarction." Stroke 21: 1023-1028.

Meyer, J. S., Shimazu, K., Fukuuchi, Y., Ohuchi, T., Okamoto, S., Koto, A. and Ericsson, A. D. (1973). "Impaired neurogenic cerebrovascular control and dysautoregulation after stroke." Stroke 4: 169-186.

Modrego, P. J., Mainar, R. and Turull, L. (2004). "Recurrence and survival after first-ever stroke in the area of Bajo, Spain. A prospective cohort study." Journal of the Neurological Sciences 224: 49-55.

Moher, D., Fortin, P., Jadad, A. R., Jüni, P., Klassen, T., Le Lorier, J., Liberati, A., Linde, K. and Penna, A. (1996). "Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews." Lancet 347: 363-366.

Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P. and Walsh, S. (1995). "Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists." Controlled Clinical Trials 16: 62-73.

Mohr, J. P. and for the American Nimodipine Study Group (1992). "Clinical trial of nimodipine in acute ischemic stroke." Stroke 23: 33-38.

Morfis, L., Schwartz, R. S., Poulos, R. and Howes, L. G. (1997). "Blood pressure changes in acute cerebral infarction and hemorrhage." Stroke 28: 1401-1405.

Morris, C. N. (1983). "Parametric empirical Bayes inference: theory and applications." Journal of the American Statistical Association 78: 47-55.

Muir, K. W. and Lees, K. R. (1995). "A Randomized, Double-Blind, Placebo-Controlled Pilot Trial of Intravenous Magnesium Sulfate in Acute Stroke." Stroke 26: 1183-1188.

Muir, K. W. and Lees, K. R. (1998). "Dose optimisation of intravenous magnesium sulfate after acute stroke." Stroke 29: 918-923.

Murphy, J. and TRUST Study Group (1990). "Randomised, double-blind, placebo-controlled trial of nimodipine in acute stroke." Lancet 336: 1205-1209.

Naylor, C. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments - Discussion. Bayesian Statistics 4. A. F. M. Smith. Oxford, UK, Clarendon Press.

Neter, J. and Wasserman, W. (1977). Applied linear statistical models. Illinois, R. D. Irwin, Inc.

Nicolucci, A., Carinci, F., Graepel, J. G., Hohman, T. C., Ferris, F. and Lachin, J. M. (1996). "The efficacy of tolrestat in the treatment of diabetic peripheral neuropathy. A meta-analysis of individual patient data." Diabetes Care 19: 1091-1096.

Norris, J. W., LeBrun, L. H., Anderson, B. A. and The Canwin Study Group (1994). "Intravenous Nimodipine in Acute Ischaemic Stroke." Cerebrovascular Diseases 4: 194-196.

Nouri, F. M. and Lincoln, N. B. (1987). "An extended activities of daily living index for stroke patients." Clinical Rehabilitation 1: 301-305.

O'Connell, J. and Gray, C. (1994). "Treating hypertension after stroke." British Medical Journal 308: 1523-1524.

Oppenheimer, S. and Hachinski, V. (1992). "Complications of acute stroke." Lancet 339: 721-724.

Oxman, A. D., Clarke, M. J. and Stewart, L. A. (1995). "From science to practice; meta-analyses using individual patient data are needed." Journal of the American Medical Association 274(10): 845-846.

Paci, A., Ottaviano, P., Trenta, A., Iannone, G., De Stantis, L., Lancia, G., Moschini, E., Carosi, M., Amigoni, S. and Caresia, L. (1989). "Nimodipine in acute ischemic stroke: a double-blind controlled study." Acta Neurol Scandinavica 80: 282-286.

Parker, C. J., Gladman, J. R. F., Drummond, A. E. R., Dewey, M. E., Lincoln, N. B., Barer, D., Logan, P. A. and Radford, K. A. (2001). "A multicentre randomized controlled trial of leisure therapy and conventional occupational therapy after stroke." Clinical Rehabilitation 15(1): 42-52.

Petitti, D. B. (2001). "Approaches to heterogeneity in meta-analysis." Statistics in Medicine 20: 3625-3633.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R. and Howard, S. V. (1977). "Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II: Analysis and examples." British Journal of Cancer 35: 1-39.

Petticrew, M., Gilbody, S. and Sheldon, T. A. (1999). "Relation between hostility and coronary heart disease: evidence does not support link." British Medical Journal 319: 917.

Pfisterer, M. (1992). "Intravenous dipyrdiamole for stress thallium-201 myocardial perfusion scintography." Cardiovascular Imagery 4: 31-39.

Pigott, T. D., Cooper, H. and Hedges, L. V. (1994). Methods for handling missing data in research synthesis. The handbook of research synthesis. L. V. Hedges. New York, Russell Sage Foundation: 163-176.

Pocock, S. J., Cook, D. J. and Beresford, S. A. A. (1981). "Regression of area mortality rates on explanatory variables: what weighting is appropriate?" Applied Statistician 30(3): 286-295.

Pokrupa, R., Hakim, A., Villanueva, J., Francis, G. and Wolfe, L. (1986). "Clinical Study of Prostacyclin Infusion after Acute

Ischemic Stroke." Le Journal Canadien Des Sciences Neurologiques 13(2): 165 (Abstract).
PROGRESS Collaborative Group (2001). "Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack." Lancet 358: 1033-1041.
Raferty, A. L. and Lewis, S. (1992). How many iterations in the Gibbs sampler? Bayesian Statistics 4. A. F. M. Smith. Oxford, Oxford University Press: 763-774.
Raghunathan, T. E. and Yoichi, I. I. (1993). "Analysis of binary data from a multicentre clinical trial." Biometrika 80(1): 127-139.
Rashid, P., Leonardi-Bee, J. and Bath, P. (2003). "Blood pressure reduction and secondary prevention of stroke and other vascular events: A systematic review." Stroke 34: 2741-2748.
Rashid, P., Weaver, C., Leonardi-Bee, J., Fletcher, S., Bath, F. J. and Bath, P. M. W. (2003). "The effects of transdermal glyceryl trinitrate, a nitric oxide donor on blood pressure, cerebral and cardiac haemodynamics and plasma nitric oxide levels in acute stroke." Journal of Stroke and Cerebrovascular Diseases 12(3): 143-151.
Rashid, P. A. (2003) Blood pressure regulation and acute stroke. Nottingham, University of Nottingham. Doctor of Medicine.
Raudenbush, S. W. (1994). Chapter 20: Random effects models. The Handbook of research synthesis. L. V. Hedges. New York, Russel Sage Foundation.
Roberts, L. and Counsell, C. (1998). "Assessment of clinical outcomes in acute stroke trials." Stroke 29: 986-991.
Robins, J., Breslow, N. E. and Greenland, S. (1986b). "Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models." Biometrics 42: 311-323.
Robins, J., Greenland, S. and Breslow, N. E. (1986a). "A general estimator for the variance of the Mantel-Haenszel OR." American Journal of Epidemiology 124: 719-723.
Robinson, T., Waddington, A., Ward-Close, S., Taub, N. and Potter, J. (1997). "The predictive role of 24-hour compared to casual blood pressure levels on outcome following acute stroke." Cerebrovascular Diseases 7(5): 264-272.
Robinson, T. G., Potter, J. F., Ford, G., Jagger, C. and Bulpitt, C. (2003). Controlling hypertension and hypotension immediately post-stroke trial (CHHIPS). 28th International Stroke Conference, Phoenix, Arizona.
Rordorf, G., Crmaer, C., Efird, J. T., Schwamm, L. H., Buonanno, F. and Koroshetz, W. J. (1997). "Pharmacological elevation of blood pressure in acute stroke. Clinical effects and safety." Stroke 28: 2133-2138.
Rosenthal, R. (1994). Chapter 16: Parametric measures of effect size. The Handbook of Research Synthesis. L. V. Hedges. New York, Russel Sage Foundation: 231-244.

Rothwell, P. M. (2001). "The high cost of not funding stroke research: a comparison with heart disease and cancer." Lancet 357(9268): 1612-6.

Rubin, D. B. (1990). A new perspective. The future of meta-analysis. M. L. Straf. New York, Russel Sage Foundation. Chapter 14: 155-165.

Rudd, A., Irwin, P. and Penhale, B. (2000). Stroke at your fingertips. London, Class Publishing Ltd.

Sacco, R. L., Diener, H. C., Yusuf, S. and Steering Committee and Study Group (2004). Prevention and Regimen For Effectively avoiding Second Strokes (PRoFESS). 29th International Stroke Conference, San Diego.

Sandercock, P., Anderson, C., Bath, P., Bereczki, D., Candelise, L., Forster, A., Fraser, H., Hankey, G. J., Langhorne, P., Lewis, S., McInnes, A., Thomas, B. and Cochrane Stroke Group (2003). In: The Cochrane Library Issue 1.

Sankey, S. S., Weisfeld, L. A., Fine, M. J. and Kapoor, W. (1996). "An assessment of the use of the continuity correction for sparse data in metaanalysis." Communications in Statistics: Simulation and Computation 25: 1031-1056.

SAS Institute Inc. Cary, NC, USA.

Saxena, R. and Koudstaal, P. (1997). "Induced elevation of blood presure in the acute phase of ischemic stroke in man." Cerebrovascular Diseases(Suppl. 1): 1-128.

Saxena, R., Wijnhoud, A. D., Carton, H., Hacke, W., Kaste, M., Przybelski, R. J., Stern, K. N. and Koudstaal, P. J. (1999). "Controlled safety study of a hemoglobin-based oxygen carrier, DCLHb, in acute ischemic stroke." Stroke 30: 993-996.

Schulz, K. F., Chalmers, I., Hayes, R. J. and Altman, D. G. (1995). "Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials." Journal of the American Medical Association 273(5): 408-412.

Schulz, K. F., Chalmers, I., Hayes, R. J. and G., A. D. (1995). "Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials." Journal of the American Medical Association 273: 408-442.

Searle, S. R. (1971). Linear models. New York, Wiley.

Seitz, R. H. and Wilson, C. L. (1987). "Effect on gait of motor task learning acquired in a sitting position." Physical Therapy 67: 1089-1094.

Shadish, W. R. and Haddock, C. K. (1994). Chapter 18: Combining estimates of effect size. The Handbook of Research Synthesis. L. V. Hedges. New York, Russel Sage Foundations.

Smith, A. F. M. and Roberts, G. O. (1993). "Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods." Journal of the Royal Statistical Society (Series B) 55(1): 3-23.

Smith, T. C., Spiegelhalter, D. J. and Thomas, A. (1995). "Bayesian approaches to random-effects meta-analysis: A comparative study." Statistics in Medicine 14(24): 2685-2699.

Song, F., Easterbrook, A., Gilbody, S., Duley, L. and Sutton, A. J. (2000). "Publication and other selection biases in systematic reviews." Health Technology Assessment 4(10): 1-115.

Spence, J. D. and Del Maestro, R. F. (1985). "Hypertension in acute ischaemic strokes - Treat." Arch.Neurol. 42: 1000-1002.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). WinBUGS: Bayesian inference using Gibbs Sampling, MRC Biostatistics Unit.

Squire, I. B., Lees, K. R., Pyrse-Phillip, W., Kertesz, A., Bamford, J. and Lifarizine Study Group (1996). "The Effects of Lifarizine in Acute Cerebral Infarction: A Pilot Safety Study." Cerebrovascular Diseases 6: 156-160.

Steinberg, K. K., Smith, S. J., Stroup, D. F., Olkin, I., Lee, N. C., Williamson, G. D. and Thacker, S. B. (1997). "Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies." American Journal of Epidemiology 145(10): 917-925.

Steiner, T. J. and for the PRISTINE Group (1996). "Naftidrofuryl in the Treatment of Acute Cerebral Hemisphere Infarction (ACHI)." Stroke 27: Abstract No. 163.

Steiner, T. J. and Rose, F. C. (1986). "Randomized double-blind placebo-controlled clinical trial of naftidrofuryl in hemiparetic CT-proven acute cerebral hemisphere infarction." Royal Society Med International Congress Symposium Series no 99: 85-98.

Stenier, T. J. and for the PRISTINE Group (1996). "Naftidrofuryl in the Treatment of Acute Cerebral Hemisphere Infarction (ACHI)." Stroke 27: Abstract No. 163.

Steuljens, E. M. J., Dekker, J., Bouter, L. M., van de Nes, J. C. M., Cup, E. H. C. and van den Ende, C. H. M. (2003). "Occupational therapy for stroke patients. A systematic review." Stroke 34: 676-687.

Stewart, L. A. and Clarke, M. J. (1995). "Practical Methodology of Metaanalyses (Overviews) Using Updated Individual Patient Data." Statistics in Medicine 14(19): 2057-2079.

Stroke Care (1998). Stroke Care; reducing the burden of disease. London.

Sudlow, C. L. M., Warlow, C. P. and the International Stroke Incidence Collaboration (1997). "Comparable studies of the incidence of stroke and its pathological types: results from an international collaboration." Stroke 28: 491-99.

Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A. and Song, F. (1998). "Systematic reviews of trials and other studies." Health Technology Assessment 2(19).

Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R. and Jones, D. R. (2000). "Empirical assessment of effect of publication

bias on meta-analyses." British Medical Journal 320: 1574-1577.

Sutton, A. J., Jones, D. R., Abrams, K. R., Sheldon, T. A. and Song, F. (1999). "Systematic reviews and meta-analysis: a structured review of the methodological literature." Journal of Health Services Research Policy 4(1): 49-55.

Sze, P. C., Reitman, D., Pincus, M. M., Sacks, H. S. and Chalmes, T. C. (1988). "Antiplatelet agents in the secondary prevention of stroke." Stroke 19: 436-442.

The American-Canadian Co-operative Study Group (1985). "Persantine aspirin trial in cerebral ischaemia. Part II: endpoint results." Stroke 16: 406-415.

The Blood pressure in Acute Stroke Collaboration (BASC) (2001(a)). "Vasoactive drugs for acute stroke (Cochrane Review)." The Cochrane Library(Issue 1).

The Cochrane Collaboration (2003). Cochrane Reviewers Handbook 4.2.0 [updated March 2003]. The Cochrane Library. A. D. Oxman. Oxford, Update Software.

The Intercollegiate Working Party for Stroke (2002). National Clinical Guidelines for Stroke. Update 2002. London, Royal College of Physicians.

The Stroke Association (1996). "Stroke services and research: an overview, with recommendations for future research." The Stroke Association: 3.

Thompson, S. G. (1994). "Systematic Review: Why sources of heterogeneity in meta-analysis should be investigated." BMJ 309(6965): 1351-1355.

Thompson, S. G. (1994). "Why sources of heterogeneity in meta-analysis should be investigated." British Medical Journal 309: 1351-1355.

Thompson, S. G. and Sharp, S. J. (1999). "Explaining heterogeneity in meta-analysis: a comparison of methods." Statistics in Medicine 18: 2693-2708.

Tijssen, J. G. P. (1998). "Low-dose and high-dose acetylsalicylic acid, with and without dipyridamole." Neurology 51 Suppl 3: S15-S16.

Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H. and Thompson, S. G. (2000). "A multilevel model framework for meta-analysis of clinical trials with binary outcomes." Statistics in Medicine 19(24): 3417-3432.

Turton, A. and Fraser, C. (1990). "The use of home therapy programmes for improving recovery of the upper limb following stroke." British Journal of Occupational Therapy 53(11): 457-462.

Uzuner, N., Ozdemir, G. and Gucuyener, D. (1985). The interaction between nimodipine and systemic blood pressure and pulse rate. Pan - European Consensus Meeting on Stroke Management, Helsingborg, Sweden.

van Swieten, J. C., J., K. P., Visser, M. C., Schouten, H. J. A. and van Gijn, J. (1988). "Interobserver agreement for the assessment of handicap in stroke patients." Stroke 19: 604-607.
Wade, D. T. (1986). Assessing disability after acute stroke. London, Clifford Rose F ed.
Wade, D. T. (1992). Measurement in neurological rehabilitation. Oxford, Oxford University Press.
Wahlgren, N. G., MacMahon, D. G., De Keyser, J., Indredavik, B., Ryman, T. and for the INWEST Study Group (1994). "Intravenous Nimodipine West European Stroke Trial (INWEST) of Nimodipine in the Treatment of Acute Ischaemic stroke." Cerebrovascular Diseases 4: 204-210.
Walker, C. M., Walker, M. F. and Sunderland, A. (2003). "Dressing after a stroke: a survey of current occupational therapy practice." British Journal of Occupational Therapy 66(6): 263-268.
Walker, M., Drummond, A. E. R. and Lincoln, N. B. (1996). "Evaluation of dressing practice for stroke patients after discharge form hospital: a crossover design study." Clinical Rehabilitation 10: 23-31.
Walker, M. F., Gladman, J. R. F., Lincoln, N. B., Siemonsma, P. and Whiteley, T. (1999). "Occupational therapy for stroke patients not admitted to hospital: a randomised controlled trial." Lancet 354(9175): 278-280.
Walker, M. F., Hawkins, K., Gladman, J. R. F. and Lincoln, N. B. (2001). "Randomised controlled trial of occupational therapy at home: results at 1 year." Journal of Neurology Neurosurgery and Psychiatry 70(2): 267-267.
Walker, M. F., Leonardi-Bee, J., Bath, P., Langhorne, P., Dewey, M., Corr, S., Drummond, A., Gilbertson, L., Gladman, J. R. F., Jongbloed, L., Logan, P. and Parker, C. (2004). "An individual patient data meta-analysis of randomised controlled trial of community occupational therapy for stroke patients." Stroke 35(9): 2226-2232.
Wallace, J. D. and Levy, L. L. (1981). "Blood pressure after stroke." Journal of the American Medical Association 246: 2177-2180.
Warlow, C. P., Dennis, M. S., van Gijn, J., Hankey, G. J., Sandercock, P. A. G., Bamford, J. M. and Wardlaw, J. (1996). Stroke: A practical guide to management. United Kingdom, Blackwell Science.
Warn, D. E., Thompson, S. G. and Speigelhalter, D. J. (2002). "Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales." Statistics in Medicine 21: 1601-1623.
Wester, P. O., Asplund, K., Eriksson, S., Hagg, E., Lithner, F. and Strand, T. (1984). "A Double-Blind Randomized Pilot Trial of Magnesium Therapy in Acute Cerebral Infarction." Acta Neurol Scandinavica 70(2): 143.
Whitehead, A. (2002). Meta-analysis of controlled clinical trials. Chichester, John Wiley & Sons Ltd.

Whitehead, A., Omar, R. Z., Higgins, J. P. T., Savaluny, E., Turner, R. M. and Thompson, S. G. (2001). "Meta-analysis of ordinal outcomes using individual patient data." Statistics in Medicine 20: 2243-2260.

Whitehead, A. and Whitehead, J. (1991). "A general parametric approach to the meta-analysis of randomized clinical trials." Statistics in Medicine 10: 1665-1677.

Whiting, S. and Lincoln, N. B. (1980). "An ADL assessment for stroke patients." British Journal of Occupational Therapy 2: 44-46.

Williams, B., Poulter, N. R., Brown, M. J., Davis, M., McInnes, G. T., Potter, J. F., Sever, P. S. and McG Thom, S. (2004). "British Hypertension Society Guidelines: Guidelines for management of hypertension: Report of the fourth working party of the British Hypertension Society, 2004 - BHS IV." Journal of Human Hypertension 18: 139-185.

Willmot, M., Ghadami, A., Weaver, C. and Bath, P. M. W. (Unpublished). "Assessment of transdermal glyceryl trinitrate on cerebral blood flow in acute and subacute stroke using Xenon computerised tomography and transcranial doppler."

Willmot, M., Leonardi-Bee, J. and Bath, P. M. W. (2004). "High blood pressure in acute stroke and subsequent outcome. A systematic review." Hypertension 43: 18-24.

Willmot, M. R. and Bath, P. M. W. (2003). "The potential of nitric oxide therapeutics in stroke." Expert Opinion Investigational Drugs 12(3): 455-470.

Wilterdink, J. L. and Easton, J. D. (1999). "Dipyridamole plus aspirin in cerebrovascular disease." Archive Neurology 56(9): 1087-1092.

Wimalaratna, H. S. K. and Capildeo, R. (1994). "Nimodipine in acute ischaemic cerebral hemisphere infarction." Cerebrovascular Diseases 4: 179-181.

Wood, H. (1984). "Managing the hypertensive patients with cerebrovascular disease." Archives of Neurology 42: 1000-1002.

Woolf, B. (1955). "On estimating the relation between blood group and disease." Annals of Human Genetics 19: 251-53.

Yatsu, F. M. and Zivin, J. (1985). "Hypertension in acute ischaemic strokes - not to treat." Arch.Neurol. 42: 999-1000.

Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. (1985). "Beta blockade during and after myocardial infarction: an overview of the randomized trials." Progress in Cardiovascular Diseases 27: 335-371.

Yusuf, S., Teo, K. and Woods, K. (1993). "Intravenous magnesium in acute myocardial infarction: An effective, safe, simple, and inexpensive intervention." Circulation 87(6): 2043-2046.

PUBLICATIONS

JOURNAL ARTICLES

Leonardi-Bee J, Bath PMW, Bousser M-G, Davalos A, Diener H-C, Guiraud-Chaumeil B, Sivenius J, Yatsu F, Dewey ME, on behalf of the Dipyridamole in Stroke Collaboration (DISC) (2005) Dipyridamole for preventing recurrent ischemic stroke and other vascular events. A meta-analysis of individual patient data from randomized controlled trials. Stroke 36(1):162-8.

Walker MF, Leonardi-Bee J, Bath P, Langhorne P, Dewey M, Corr S, Drummond A, Gilbertson L, Gladman JRF, Jongbloed L, Logan P, Parker C (2004). An individual patient data meta-analysis of randomised controlled trial of community occupational therapy for stroke patients. Stroke 35(9):2226-32.

Outpatient Service Trialists.(2004) Rehabilitation therapy services for stroke patients living at home: systematic review of randomised trials. Lancet 363:352-356.

Leonardi-Bee J, Bath PMW, Phillips SJ, Sandercock PAG, for the IST Collaborative Group. (2002) Blood pressure and clinical outcomes in the International Stroke Trial. Stroke 33:1315-1320.

Willmot M, Leonardi-Bee J, Bath PMW. (2004) High blood pressure in acute stroke and subsequent outcome. A systematic review. Hypertension 43:18-24.

ABSTRACTS AND A SELECTION OF CONFERENCE PROCEEDINGS

Leonardi-Bee J, Bath FJ, Bath PMW on behalf of the BASC. Lowering blood pressure in acute stroke: Data from the 'Blood pressure in Acute Stroke Collaboration' (BASC). To be presented at: 13th European Stroke Conference 2004 (Mannheim-Heidelberg).

Leonardi-Bee J. (2003) Heterogeneity in meta-analyses. Presented at: 26th Research Students' Conference in Probability and Statistics (University of Surrey)

Leonardi-Bee J. (2002) An overview of individual patient data meta-analysis: results from the community occupational therapy trial. Presented at: Centre for Research and Rehabilitation (University of Nottingham).